ETH ZÜRICH

MASTERS THESIS

The Chromatic Symmetric Function on Random Graphs

Author:
Raúl Penaguião

Supervisors:
Valentin Féray
Angelika Steger

A thesis submitted in fulfilment of the requirements for the degree of Masters in Pure Mathematics

in the

ETH Zürich

September 9, 2016

“Mathematics is the most beautiful and most powerful creation of the human spirit."

Stefan Banach

ETH ZÜRICH

Abstract

D-MATH

Master in Science

The chromatic symmetric function on random graphs

by Raúl Penaguião

Considering the chromatic symmetric function of a random graph we get a random variable that takes values as homogeneous symmetric functions. The coefficients over several bases are random variables that describe some combinatorial properties of the random graphs.

In this thesis, we will focus on the $G(n,p)$ random graph model and find thresholds values for $X_\lambda = 0$, where X_λ is related to the coefficients of the monomial basis m_λ of the chromatic symmetric function of $G(n,p)$. For m_λ where the partition $\lambda = 1^s\mu$ has a fixed block μ and $s \to \infty$, the threshold problem is approached via the subgraph containment problem, whereas for varying μ the problem is much deeper.

Here we also compute bounds to the threshold value of $X_\lambda = 0$ for the partition $\lambda = 1^s2^r$ for different rates of $r \to \infty$. This subject will be dealt with after a discussion on finding big matchings in $G(n,p)$, which will serve as intuition for the results on the variables X_λ. We also discuss some results on the uniform tree model and the coefficients over the power-sum basis.

Acknowledgements

This work would have never been done without the help of many people. All my friends and family deserve a kind word for being what I needed the most, making me understand that as challenging as a Masters degree might be, I shouldn't carry it all upon my shoulders.

A very special thanks to Miguel Martins do Santos and Rita Neves for being huge role models, in their own different ways, for me. I kindly thank Frederico Toulson for the comradeship and Ricardo Moreira for the insightfulness that always brought with them. To Beatrice da Costa, that received me and helped without asking anything in return when I most needed, I warmly thank.

None of this and much more would be written without the much needed help of what was for me the best school, Grupo 16 de Carcavelos.

To the one that tirelessly helped me with editing this work, listening and checking my mathematical reasoning, debating the next steps of my work and in the most simple manner was a friend, Ana Broges, I'm forever thankful.

For Prof. Dr. Valentin Féray, that guided me and presented me to research in Combinatorics, inviting me to a world where all mathematicians find more beauty and coherence, challenge and dare, success and frustration but among all, fulfilment. It had its ups and downs but being ready for the next one was always half the work done and Prof. Dr. Valentin Féray was always ready to show that there was another way to keep going.

And above all, a very special thanks to my mother, to whom I everything owe and from whom all opportunities I was given.

Contents

Chapter 1

Introduction

In this thesis we will study an algebraic invariant of a graph, the *chromatic symmetric function*, introduced in [19], and the behaviour of this invariant in Random Graph Models.

A homogeneous symmetric function of degree n over infinitely many variables $x_1, x_2, \dots$ is a formal power sum of the form

$$f(x) = \sum_{\alpha, k \geq 0} a_{\alpha_1, \dots, \alpha_k} x_1^{\alpha_1} \cdots x_k^{\alpha_k},$$

where the sum runs over non-negative integers k and α_i. Additionally, for f to be symmetric we require that changing the role of two variables doesn't change the formal sum.

For fixed n, the set of homogeneous symmetric functions of degree n is a finite dimensional vector space. Two of the most notable bases are the *monomial basis*, and the *power-sum basis*.

For a graph G with $V(G) = \{v_1, v_2, \dots, v_n\}$, Richard Stanley defines the chromatic symmetric function as

$$\chi_G = \sum_k x_{k(v_1)} \cdots x_{k(v_n)},$$

where the sum runs over all proper colourings $k : V(G) \to \mathbb{N}$ of the graph G: colourings in which two neighbouring vertices always have different colours. This invariant is a symmetric function, so we can expand it in the monomial basis and power-sum basis.

The coefficients of χ_G over these bases are of combinatorial interest, namely, it's shown in [19] that the coefficients over the monomial basis are closely related with x_λ, the number of block partitions of $V(G)$ into independent blocks with a given partition type λ.

Let $\lambda(A)$ denote the partition type of the set partition given by the connected components of the edge set A. The coefficients a_λ of χ_G over the power-sum basis are related with the edge sets A with $\lambda(A)$ fixed, in a way that we will clarify latter. We denote by X_λ and A_λ the random variables that represent the coefficients x_λ and a_λ of χ_G when G is taken to be a random graph.

So, in the realm of random graphs, the coefficients of χ_G in the different basis sets are related to the independent sets and connected components of fixed types of the random model. A very common random graph model is $G(n,p)$, where each edge occurs independently with probability p

A natural question is, for a given graph property A, to study for which $p = p(n)$ we have

$$\mathbb{P}[G(n,p(n)) \text{ satisfies } A] \to 0\,,$$

or

$$\mathbb{P}[G(n,p(n)) \text{ satisfies } A] \to 1\,,$$

when n is big. Such is addressed as *the threshold problem.*

In this work we mainly ask for the threshold of the property $X_\lambda = 0$. The goal of this thesis is, not only to relate this problem with other well studied topics such as "number of big matchings" and "subgraph containment problem" but as well to find something new in the chromatic symmetric function context. Namely, a high concentration of the random variables X_λ and A_λ mean that there are a lot of graphs with the same chromatic symmetric function or where some coefficients are the same.

1.1 Structure of the Report

In the remaining of this chapter we will introduce the relevant background on the topic, namely fixing some notation on graphs, partitions and introducing the symmetric functions. We also introduce the main theorems on chromatic symmetric functions on graphs and add some historical context.

In chapter two we introduce a random graph model $G(n,p)$ as well as $G(n,M)$, where we state some thresholds of various properties, such as connectedness, existence of a perfect matching, etc.

In chapter three we investigate the threshold problem for the variables X_λ and A_λ.

In chapter four we summarize the discussion and introduce the further work. The appendices contain some notation that we will follow in the report, as well as some useful propositions well known in the literature.

1.2 Background

All definitions in this thesis should be clear to any last year bachelor student of mathematics. We use a dot as a decimal separator between the integer part and the fractional part of a number (American notation).

1.2.1 Graphs

A graph is a pair $G = (V, E)$ where each $v \in V$ is called *vertex* and $e = \{v_1, v_2\} \in E$ is an *edge* that has two indistinguishable endpoints in V.

The graph with all $\binom{n}{2}$ possible edges in n vertices is called the *complete graph* and we write it as K_n. For the graph with no edge among an n-vertex set, we call the *zero graph*, or 0_n. If we have two disjoint sets of vertices A, B with size m, n respectively, we refer to the graph with $V(G) = A \cup B$ and all possible edges with endpoints in both A and B as $K_{m,n}$. Equivalently, $K_{m,n} = (K_n \uplus K_m)^c$. This is called the *complete bipartite graph*.

Given a graph G, the complementary graph G^c is a graph in the same set of vertices but with the complementary set of edges, i.e. $V(G) = V(G^c)$ and $E(G) \uplus E(G^c) = E(K_n)$.

In this work we will always regard $V(G)$ as a finite set. We will also assume that there are no parallel edges, i.e. edges $e_1, e_2 \in E$ with the same endpoints, as well as no loops, i.e. an edge $e \in E$ which the endpoints are the same, unless stated otherwise, in which case we refer to *multigraphs*.

We write $v(G) := \#V(G)$, but often we address to a non-fixed parameter n as the number of vertices of a graph G, when such graph is clear by context. We also write $e(G) := \#E(G)$.

We call a *subgraph* of G any graph H such that $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$ and we write it as $H \subseteq G$. We say that H is an *induced subgraph* if any edge $e \in E(G)$ with both endpoints in $V(H)$ is also an edge in H, i.e. $e \in E(H)$. In graph theory, when it is ask for a minimal graph, it is meant with respect to the subgraph partial order.

A *connected graph* is a graph where for each pair of vertices v_1, v_2 there is a sequence of edges $e_1, \ldots, e_k$ such that, e_1 is incident to v_1, e_k is incident to v_2 and e_j and e_{j+1} share a vertex for any j. A maximal connected subgraph H of G is called a connected component, and we call $c_1(G)$ to the number of vertices of the biggest component.

Given a set of vertices $A \subseteq V(G)$, we denote by $N(A) = \{v \in V(G) | \exists v' \in A, \{v, v'\} \in E(G)\}$ the set of *neighbours* of the vertices of the set A.

A *colouring* is a map $k : V(G) \to \mathbb{N}$. A *proper colouring* is a colouring in which for each edge $e = \{v_1, v_2\} \in E$ we have $k(v_1) \neq k(v_2)$, i.e. we attribute to each vertex a colour in a way that each edge has both endpoints with different colours.

A *stable set*, or *independent set*, of vertices in a graph G is a set U of vertices $U \subseteq V(G)$ with no edge between any of them.

It is often the case that we talk about properties of graphs Q as sets of graphs, using the notation $G \in Q$. For instance, if Q is the property that a graph contains a triangle, then $K_5 \in Q$.

We say that a property Q is *increasing* if for any two graphs G, H in the same vertex set such that $H \subseteq G$ and $H \in Q$ implies $G \in Q$. We say that a property Q is *decreasing* if for any two graphs G, H in the same vertex set such that $H \subseteq G$ and $G \in Q$ implies $H \in Q$. We say that a property Q is *convex* if for any three graphs G, H_1, H_2 in the same vertex set such that $H_1 \subseteq G \subseteq H_2$ and $H_1, H_2 \in Q$ implies $G \in Q$. A property is called *monotone* if it's either increasing or decreasing.

1.2.2 Permutations and compositions

A *weak composition* α is an infinite ordered list of non-negative integers $(\lambda_1, \lambda_2, \ldots)$ where all but finitely many are zero. We say that α is a weak composition of n, or $\alpha \models n$, when $n = \sum_i \alpha_i$. We often disregard trailing zeroes.

A *partition* λ is an infinite multiset of integers $(\lambda_1, \lambda_2, \ldots)$ where all but finitely many are zero. Unlike weak compositions, there is no harm in disregarding the zero elements and writing a partition as a tuple. For instance we can write the partition $\lambda = (2, 2, 1, 1, 1)$. If $n = \sum_i \lambda_i \geq 0$, we say that λ is a partition of n, or $\lambda \vdash n$. We call *the parts* of λ to the different integers $\lambda_1, \ldots$ and, unless it's clearly otherwise, we assume $\lambda_1 \geq \lambda_2 \cdots$. By disregarding the order of a list we can map weak compositions α to partitions $\lambda(\alpha)$, which we call the *type* of α.

To count the number of weak compositions of n is an easy combinatorial problem. On the other hand, partitions are fundamentally different, and more complicated non-closed formulas, such as the ones mentioned in [10].

For a partition λ, the size $l(\lambda)$ is the number of non-zero parts. For any $i > 0$, we define $\lambda^{(i)}$ to be the number of parts of size i.

Take for instance the partition $\lambda = (3, 2, 2, 2, 1, 1, 1)$, then $\lambda^{(1)} = 3$, $\lambda^{(2)} = 3$ and $\lambda^{(3)} = 1$. The weak composition $\alpha = (2, 1, 0, 0, 1, 3, 2, 0, 0, 1, 2)$ has type λ.

Besides denoting a partition as an unordered list, we can as well describe it in a multiplicative notation, namely we write $\langle 1^3 2^2 \rangle$ for the partition $(2, 2, 1, 1, 1)$, and will be often the case that we drop the notational aid $\langle\rangle$. We also may write partitions in a concatenation-multiplicative notation, for instance if $\mu = 2^2$ then $\lambda = 1^3\mu$ represents the partition with 3 blocks of 1 part added to the partition μ, which is the same as $(2,2,1,1,1)$ again.

We define $\mathrm{aut}(\lambda) = \prod_i \lambda^{(i)}!$. We introduce as well $\lambda! = \prod_i \lambda_i!$, which is the number of automorphisms of a set that preserve the blocks of a set partition of type λ. We also define $\binom{n}{\lambda}$ whenever $\lambda \vdash n$ is a partition, as $\frac{n!}{\lambda_1!\cdots\lambda_{l(\lambda)}!} = \frac{n!}{\lambda!}$.

Given an integer partition $\lambda = (\lambda_1, \lambda_2, \cdots)$, we write $p(\lambda) = \prod_i \lambda_i$.

A *set partition* π of a ground set A is a family $\pi = \{\pi_1, \pi_2, \cdots\}$ of disjoint subsets of A, called *blocks*, such that $A = \biguplus_i \pi_i$. In an abuse of notation, we can write $\pi = \biguplus_i \pi_i$. In a graph G, a *stable partition* is a set partition π on the ground set of vertices $V(G)$ such that each block is stable.

1.2.3 Symmetric Functions and Chromatic Symmetric Functions

Definition 1.2.1. In this work, for a weak composition α, we denote by x^α the finite product

$$x^\alpha = \prod_i x_i^{\alpha_i}.$$

We will consider infinite homogeneous sums over infinitely many variables of degree n, which in the general form are

$$f = \sum_{\alpha \models n} a_\alpha x^\alpha,$$

where the sum runs over all weak compositions of n.

A *homogeneous symmetric function* of degree n is a sum f in which the coefficients a_α depend only on the type of α, i.e. if $\lambda(\alpha) = \lambda(\beta)$ then $a_\alpha = a_\beta$. The set of homogeneous symmetric functions of degree n forms a vector space written as Λ_n.

Definition 1.2.2. Let $m_\lambda = \sum_{\lambda(\alpha)=\lambda} x^\alpha$ and $p_\lambda = \prod_i m_{(\lambda_i)}$. Those are called *monomial symmetric functions* and *power-sum symmetric functions*, respectively.

Then, for instance, $m_{(0)} = 0$ and $m_{(1,1)} = \sum_{i<j} x_i x_j$. We also have

$$\begin{aligned}
p_{(2,1)} &= m_{(2)} m_{(1)} \\
&= \left(\sum_i x_i^2\right)\left(\sum_i x_i\right) \\
&= \sum_{i \neq j} x_i^2 x_j + \sum_i x_i^3 \\
&= m_{(2,1)} + m_{(3)}.
\end{aligned}$$

The next theorem is a classical one in the theory of symmetric functions, present in any introduction of the topic.

Theorem 1.2.3. Both sets $\{m_\lambda\}_{\lambda\vdash n}$ and $\{p_\lambda\}_{\lambda\vdash n}$ are a basis for Λ_n.

Proof. The proof that $\{m_\lambda\}_{\lambda\vdash n}$ is a basis is straightforward. To show that $\{p_\lambda\}_{\lambda\vdash n}$ is indeed a basis, we address to [20, Chapter 7]. □

Definition 1.2.4 (Chromatic Symmetric Function on Graphs)**.** For a graph G with $V(G) = \{v_1, v_2, \dots, v_n\}$, the chromatic symmetric function is

$$\chi_G = \sum_k x_{k(v_1)} \cdots x_{k(v_n)}\,,$$

where the sum runs over all proper colourings $k : V(G) \to \mathbb{N}$ of the graph G.

Example 1.2.5. Consider the complete graph K_n. A proper colouring of K_n is an injective function from $V(K_n)$ to $\mathbb{N}$. Hence, only monomials for weak compositions of type 1^n are counted in the sum. Now it's easy to see that the graph K_n has, for each α of type 1^n, $n!$ colourings of type α, so $\chi_{K_n} = n! m_{(1,\dots,1)}$.

This example easily generalizes to the following Theorem:

Theorem 1.2.6. Given a graph G, and for each partition $\lambda \vdash n$, let x_λ be the number of stable partitions of type λ.

Then

$$\chi_G = \sum_{\lambda\vdash n} \operatorname{aut}(\lambda) x_\lambda m_\lambda\,.$$

Proof. The proof can be found in [14]. □

For the power-sum expansion, the coefficients of χ_G also have some combinatorial meaning.

Theorem 1.2.7. Given a graph G.

Then

$$\chi_G = \sum_{A\subseteq E(G)} (-1)^{\#A} p_{\lambda(A)}\,.$$

Proof. The proof can be found in [14]. □

This theorem becomes much simpler when in the realm of trees.

Theorem 1.2.8. Given a tree T, and for each partition $\lambda \vdash n$, let θ_λ be the number of edge-sets A such that $\lambda(A) = \lambda$.

Then, the formula in Theorem 1.2.7 becomes:

$$\chi_T = \sum_{A \subseteq E(G)} (-1)^{\# A} p_{\lambda(A)} = \sum_{\lambda \vdash n} (-1)^{n-l(\lambda)} \theta_\lambda p_\lambda .$$

Proof. The proof can be found in [14]. □

1.2.4 Random Graphs

The most classical model is the following:

Definition 1.2.9 (Erdös-Rényi random graph model)**.** The random graph model $G(n,p)$ is a distribution over the set of graphs on a given n vertex set, where each edge occurs independently with probability p.

In this sense, a graph G with M edges over the given n vertex set satisfies, by independentness of the edges,

$$\mathbb{P}[G(n,p) = G] = p^M (1-p)^{\binom{n}{2} - M} .$$

We point out that there is also the random graph model $G(n,M)$ that takes uniformly at random a graph with M edges.

Definition 1.2.10 ($G(n,M$ graph model)**.** The random graph model $G(n,M)$ takes uniformly at random as value a graph with M edges and n vertices. This model, we will see, is closely related to $G(n,p)$ for $p = M\binom{n}{2}^{-1}$.

The random graph model that uniformly takes values over all n-vertex trees is denoted by T_n.

The random bipartite model will also be of use to us, which we introduce now:

Definition 1.2.11 (The $G(m,n,p)$ model)**.** The model $G(m,n,p)$, given two vertex sets A, B of size m, n respectively, is given by $G(n,m,p) = [G(n+m,p) | A, B \text{ is a bipartition }]$. So, each edge between vertices from A to B occurs independently with probability p, and no other edge occurs.

Some random variables will be useful that we refer to as *Indicator variables*: given an event E, the indicator variable I_E takes value 1 if E occurs, and value 0 otherwise.

In the Erdös-Rényi model we will often try to find thresholds: given a suitable graph property Q, the limiting probability $\lim_n \mathbb{P}[G(n,p(n)) \in Q]$ or $\lim_n \mathbb{P}[G(n,M(n)) \in Q]$ jumps rapidly from 0 to 1 by varying the function $p(n)$ or $M(n)$, and to find the "turning function", or *threshold*, leads to intriguing problems.

Definition 1.2.12 (Threshold)**.** For a non trivial increasing property Q, a sequence $\hat{p} = \hat{p}(n)$ is a threshold if

$$\mathbb{P}[G(n, p(n)) \in Q] \to \begin{cases} 1 & \text{if } p >> \hat{p}, \\ 0 & \text{if } p << \hat{p}. \end{cases}$$

For asymptotical notation check Appendix A.

It is, then, not natural to talk about a unique threshold function p, but nevertheless we talk about *the* threshold as the class of functions $\hat{p}' \asymp \hat{p}$. Note that the asymptotic notation is all summed up in the Appendix B.

We additionally define $p_Q(a; n)$ to be a value p such that $[G(n, p) \in Q] = a$: it is easily seen that, for non-trivial increasing properties Q, $p \mapsto [G(n, p) \in Q]$ is increasing continuous function that fixes 0 and 1, so such value exists and is unique. We write $p(a, n) := p_Q(a, n)$ when the property Q is clear from context.

For a decreasing property Q, the threshold is defined as the complement property. It will be often the case that our threshold function approaches 1 in the limit, and we want to be more precise regarding the rate at which the function approaches 1, whereas the threshold definition presented here and found in the literature is only rough enough to properly distinguish thresholds that approach 0. For that reason we will talk about thresholds of the form $1 - \Theta(f)$, when we mean that $\Theta(f)$ is a threshold for the complementary property, with no due warning.

For each fixed a, $p_Q(a, n)$ is a function on n. We will see now that such functions aren't actually much different to each other. We will see, as well, that this implies that these functions are actually the threshold of the property Q, proving that a threshold always exists for monotone properties. This demonstration follows the lines written in [12].

Proposition 1.2.13. For fixed $a > 0$ and a monotone graph property Q, we have

$$p(a; n) \asymp p(0.5; n).$$

Proof. Suppose wlog that Q is increasing and fix a such that $0 > a > 0.5$. Let m be integer such that $a \geq (1-a)^m$. We will show that $p(a; n) \leq p(0.5; n) \leq p(1-a; n) \leq mp(a; n)$ and since m does not depend on n, the asymptotic relation follows. Note that $p(a, n)$ is naturally an increasing function on a, hence the only non-trivial inequality is $p(1-a; n) \leq mp(a; n)$.

Let G be the graph such that $E(G) = \cup_i E(G_i(n, p(a; n)))$ where the different $G_i(n, p(a; n))$ are independent random graphs on the same vertex set. Then $G \sim G(n, p')$ where $p' = 1 - (1 - p(a; n))^m \leq mp(a; n)$ from Bernoulli Inequality (Proposition A.2.2), so

$$\begin{aligned}
\mathbb{P}[G(n, mp(a; n)) \notin Q] \leq \mathbb{P}[G(n, p') \notin Q] &= \mathbb{P}[G \notin Q] \\
&\leq \mathbb{P}[G_i(n, p(a; n)) \notin Q \, \forall \, i] \\
&= \mathbb{P}[G(n, p(a; n)) \notin Q]^m \\
&= (1 - a)^m \leq a.
\end{aligned}$$

So $\mathbb{P}[G(n, mp(a; n)) \in Q] \geq 1 - a$ and hence $mp(a; n) \geq p(1 - a; n)$. This also provides the proof for $0.5 \leq a < 1$. The proof for the case Q decreasing is analogous. □

Now it can be easily seen that $\hat{p}(n) = p(0.5; n)$ is a threshold: suppose otherwise, that it is not a threshold. Then there would be a sequence p such that either

- $p << \hat{p}$ and $\liminf \mathbb{P}[G(n, p(n)) \in Q] > 0$

  or

- $p >> \hat{p}$ and $\limsup \mathbb{P}[G(n, p(n)) \in Q] < 1$.

In the first case there is an $a > 0$ such that $\mathbb{P}[G(n, p(n)) \in Q] > a$ for every n, so $p(n) \geq p(a; n) \asymp \hat{p}$ by Proposition 1.2.13, so it is impossible to have $p << \hat{p}$. The other case is dealt with in a similar way, as well as the case of a decreasing property.

We conclude:

Theorem 1.2.14 (Existence of Thresholds)**.** Every monotone property has a threshold function given by $p_Q(0.5, -)$.

Normally, it's not easy to compute the function $p_Q(0.5; -)$, and it's normal to find thresholds through other methods. Namely, we will show, for a given function $\hat{p}$, that $p << \hat{p} \Rightarrow \mathbb{P}[G(n, p) \in Q] \to 0$ and that $p >> \hat{p} \Rightarrow \mathbb{P}[G(n, p) \in Q] \to 1$ directly. The first implication is usually addressed as the *0-statement* and the second as the *1-statement* for Q an increasing function.

Since thresholds always exist for monotone properties, we will now devise a refinement of the threshold notion that hold only for suitable monotone properties. Such is presented below:

Definition 1.2.15 (Sharp Thresholds)**.** A threshold $\hat{p}$ is called a *sharp threshold* if

$$\mathbb{P}[G(n,p(n)\in Q] \to \begin{cases} 1 & \text{if } p > (1+\eta)\hat{p}\,\forall n \\ 0 & \text{if } p < (1-\eta)\hat{p}\,\forall n. \end{cases}$$

Namely, the notion of sharp threshold asks what happens when $p \asymp \hat{p}$: if the passage from $\mathbb{P}[G(n,p)\in Q]\to 0$ to $\mathbb{P}[G(n,p)\in Q]\to 1$ occurs suddenly, we say that the property has a sharp threshold $\hat{p}$.

The properties Q that don't have a sharp threshold are said to have a *coarse threshold.*

We will see an example of coarse thresholds in this essay, for the threshold of subgraph containment. Namely, for the property $Q =$ "contains a triangle" we will see that the threshold is $\Theta(n^{-1})$ and for $p(n) = cn^{-1}$ the number of triangles follows a Poison distribution with parameter $c^3/6$, so the probability $\mathbb{P}[G(n,p)\notin Q] \to \exp(-c^3/6)$ varies smoothly with c and the threshold is not sharp.

On the other hand, the property that the graph is connected has a sharp threshold [9]. We will see in this work that the existence of isolated vertices also has a sharp threshold, in 2.1.7.

1.2.5 Important theorems and goals

A consequence from Theorem 1.2.6 is the next Theorem, that describes the expected value of the coefficients of χ_G in the monomial basis for $G(n,p)$.

Theorem 1.2.16. For the random graph model $G(n,p)$ we have

$$\mathbb{E}[\chi_{G_{n,p}}] = \sum_{\lambda\vdash n}\binom{n}{\lambda_1,\ldots,\lambda_{l(\lambda)}}(1-p)^{\sum_{j=1}^{l(\lambda)}\binom{\lambda}{2}} m_\lambda .$$

Proof. This is shown in [14]. □

In this project we will try to find properties of the coefficients in the chromatic symmetric function from different graph models and bases, namely the expected value and variance, threshold values and concentrations.

We will also prove a similar Theorem from Theorem 1.2.16 for the power-sum basis.

Historical Context

The interest of the community in random graphs is not recent, and the Erdös Rényi model dates back to 1959 in [6]. The subject of thresholds was introduced as well by Edrös and Rényi and has been flourishing ever since, generating the most variate results with common properties, for instance Friedgut has shown in 1995 that the property that a graph is connected has a sharp threshold [9]. Tied to these studies is also the study of big matchings on graphs, Erdös and Rényi found the threshold solution in 1966 in [7], and the hitting time version was only latter proved by Bollobás and Thomason in 1985, in [3].

Regarding the chromatic symmetric function, as it was already said it was introduced in [19] alongside with a conjecture that any two non-isomorphic trees have different chromatic symmetric function.

So far that remains as a conjecture but a lot of partial results have been done so far. For instance in 2003, Joshua Fougere in [8] showed that a small family of trees, called *forks*, can be classified using the chromatic symmetric function. On the other hand, there has been also some development in constructing algorithmically a tree from its chromatic symmetric function. For instance in 2013, Rosa Orellana and Geoffrey Scott in [13] introduced an algorithm that constructs a tree that has only one centroid given its chromatic symmetric function and some additional information. Latter that result was improved in 2015 by Isaac Smith et al. in [18] for any tree.

Despite there being some results about the chromatic symmetric function and many on random graphs, the study of both subjects together has been rather timid. There are some interesting results that relate chromatic invariants with random graphs. For instance it is a simple exercise to show that the chromatic number is highly concentrated around its mean value with Azuma's inequality. It is, then, natural to bring some questions regarding the chromatic symmetric function into the random graphs.

Chapter 2

Random graph models and thresholds

2.1 Probabilistic Tools

The simplest way to obtain some results will be through 1st and 2nd moment methods (Lemmas 2.1.3 and 2.1.6), as well as Markov and Chebychev inequalities (Lemmas 2.1.1 and 2.1.2), that, in favourable cases, reduce the threshold computation to evaluation of the expected value and variance of random variables.

These Lemmas are from [2] and compose the basis for the probabilistic method. The proofs of the following Lemmas are based on the lecture notes of a Prof. Dr. Angelika Steger's course[1].

Lemma 2.1.1 (Markov's Inequality)**.** If X is a random variable that takes non-negative values, and $a > 0$ a real value, then

$$\mathbb{P}[X \geq a] \leq \frac{\mathbb{E}[X]}{a}$$

Proof. Let Y be the indicator variable of the event "$X \geq a$". Then $X \geq aY$ and so

$$\mathbb{E}[X] \geq a\mathbb{E}[Y] = a\mathbb{P}[X \geq a],$$

as desired. □

Lemma 2.1.2 (Chebichev's Inequality)**.** Let X be a random variable with $\mathbb{E}[X] = \mu$, $\mathrm{Var}[X] = \sigma^2$ and let $a > 0$ be a real value, then

$$\mathbb{P}[|X-\mu| \geq a] \leq \left(\frac{\sigma}{a}\right)^2.$$

[1] Randomized Algorithms and Probabilistic Methods - `http://www.cadmo.ethz.ch/education/lectures/HS15/rand_alg.html`

Proof. Set $Y = (X-\mu)^2$. Note that $\mathbb{E}[Y] = \mathrm{Var}[X] = \sigma^2$ so applying Markov's Inequality to Y we get

$$\mathbb{P}[|X-\mu| \geq a] = \mathbb{P}[Y \geq a^2] \leq \frac{\mathbb{E}[Y]}{a^2} = \frac{\mathrm{Var}[X]}{a^2}. \quad \square$$

Lemma 2.1.3 (1st Moment Method)**.** If X_n is a sequence of non-negative integer valued random variables that satisfy $\mathbb{E}[X_n] = o(1)$, then

$$\mathbb{P}[X_n = 0] = 1 - o(1).$$

Proof. We have, by Markov's inequality, that

$$\mathbb{P}[X_n \neq 0] = \mathbb{P}[X_n \geq 1] \leq \frac{\mathbb{E}[X_n]}{1} = o(1).$$

So $1 \geq \mathbb{P}[X_n = 0] \geq 1 - \mathbb{E}[X_n] = 1 - o(1)$. □

We will now see some simple applications of the first moment method, that will be useful latter on.

Example 2.1.4. Let's count the number of isolated vertices in $G(n,p)$ for $np = \log n + \omega(1)$ (recall the asymptotic notation summarized in Appendix A).

We let $X = \sum_v I_v$ be the number of isolated vertices, where I_v is the indicator variable for the event "the vertex v is isolated".

Now note that $\mathbb{E}[I_v] = \mathbb{P}[I_v = 1] = (1-p)^{n-1} \leq \exp(-p(n-1))$ from Proposition A.2.2, so

$$\mathbb{E}[X] \leq n\exp(-p(n-1)) = \exp(\log n - np + p) = \exp(-\omega(1)) = o(1).$$

So from the 1st moment method (Lemma 2.1.3) we get that $\mathbb{P}[X = 0] \to 1$.

The following example is from [12, Chapter 4] as a Lemma to find big matchings.

Example 2.1.5. Take the random graph model $G(n,p)$. Let X_u be the number of unordered pairs (A,B) of disjoint subsets of vertices, each of size u, with no edges between A and B, and let's choose $u = \frac{n(\log\log n)^2}{\log n} \to \infty$. We will see that if $np = \Theta(\log n)$ then $\mathbb{P}[X_u = 0] \to 1$ by computing its expected value.

Incidentally, for disjoint sets A,B of size u, let $I_{A,B}$ be the indicator value of the event "there is no edge between A and B", so $X_u = \sum_{A,B} I_{A,B}$ and

$$\mathbb{E}[X_u] = \sum_{A,B} \mathbb{E}[I_{A,B}] = \binom{n}{2u}\binom{2u}{u}\frac{1}{2!}(1-p)^{u^2}$$

Using Proposition A.2.3 and Proposition A.2.2 we get:

$$\begin{aligned}\mathbb{E}[X_u] &\le \left(\frac{ne}{2u}\right)^{2u} \frac{4^u(1+o(1))}{2\sqrt{u\pi}} \exp(-u^2 n^{-1}\Theta(\log n)) \\ &= \frac{1}{2\sqrt{\pi}}\left(\frac{e\log n}{(\log\log n)^2}\right)^{2u} u^{-1/2} \exp(-u(\log\log n)^2\Theta(1))\end{aligned}$$

which amounts to

$$\begin{aligned}\log\mathbb{E}[X_u] &\le -\log(2\sqrt{\acute{L}pi}) + 2u(1+\log\log n - 2\log\log\log n) - \log u/2 - u(\log\log n)^2\Theta(1) \\ &= -\Theta(u(\log\log n)^2) \to -\infty .\end{aligned}$$

Concluding that $\mathbb{P}[X_u = 0] \to 1$.

Lemma 2.1.6 (2nd Moment Method)**.** Let X_n be a sequence of random variables that take non-negative values and satisfy $\mathbb{E}[X_n] = \omega(1)$. If additionally we have that $\mathrm{Var}[X_n] = o(\mathbb{E}[X_n]^2)$, then for any constant a we have that

$$P[X_n \le a] = o(1) .$$

Proof. Write $\mu_n = \mathbb{E}[X_n]$. Note that $\mathrm{Var}[X_n] = o(\mu_n^2)$.

Just note that there is an N such that $a < \mu_n/2$ for all $n \ge N$ so, by Chebichev's Inequality (Lemma 2.1.2), for all $n \ge N$,

$$\mathbb{P}[X_n \le a] \le \mathbb{P}[|X_n - \mu_n| \ge \mu_n/2] \le \frac{4\mathrm{Var[X_n]}}{\mu_n^2} = o(1) . \quad \square$$

We can develop now further the example of isolated vertices.

Example 2.1.7. We recover the notation from 2.1.4.

We have seen that if $np = \log n + \omega(1)$, then no isolated vertices occur. Now we will see that $\mathbb{P}[X = 0] \to 0$, through the 2nd moment method (Lemma 2.1.6), for $np = \log n - \omega(1)$.

First it is clear that $\mathbb{E}[X] = n(1-p)^{n-1} \to \infty$ as from $\log(1-x) = x + O(x^2)$ we have:

$$\begin{aligned}\log\mathbb{E}[X] &= \log n + (n-1)\log(1-p) \\ &= \log n - n(p + O(p^2)) + \log(1-p) \\ &= \omega(1) - O(np^2) + o(1) = \omega(1) .\end{aligned}$$

Now, to bound the variance of X, the trick here is to understand that the variables I_v are almost independent, so the variance is small. In particular, for vertices $v \neq w$, and $p = o(1)$.

$$\begin{aligned}
\mathrm{Var}[X] &= \sum_{v,w} \mathrm{Cov}[I_v, I_w] = n(n-1)(\mathbb{E}[I_v I_w] - \mathbb{E}[I_v]\mathbb{E}[I_w]) + n(\mathbb{E}[I_v^2] - \mathbb{E}[I_v]^2) \\
&= (n^2 - n)(1-p)^{2n-3}(1-(1-p)) + n((1-p)^{n-1} - (1-p)^{2n-2}) \\
&= n^2(1-p)^{2n-2}\frac{p}{1-p} + n(1-p)^{n-1}(-p(1-p)^{n-2} + 1 - (1-p)^{n-1}) \\
&= o(\mathbb{E}[X]^2) + \Theta(\mathbb{E}[X]).
\end{aligned}$$

So this amounts to $\mathrm{Var}[X] = o(\mathbb{E}[X]^2)$, which concludes the application of the 2nd moment method (Lemma 2.1.6) and that $\mathbb{P}[X=0] \to 0$. Hence, there are isolated vertices a.a.s.

Note that, incidentally, the threshold for the existence of isolated vertices is given by

$$p = n^{-1}\log n + \Theta(n^{-1}),$$

which is a sharp threshold, as, for any $\eta > 0$,

$$\mathbb{P}[G(n, p(n)) \text{ has no isolated vertices }] \to \begin{cases} 1 & \text{if } p > (1+\eta)n^{-1}\log n \ \forall n \\ 0 & \text{if } p < (1-\eta)n^{-1}\log n \ \forall n. \end{cases}$$

Lemma 2.1.8 (Chernoff Bounds)**.** Let $X \sim \mathrm{Bin}(n,p)$ and call $\mu = \mathbb{E}[X] = np$.

Then we have, for any $\lambda > 0$:

$$\mathbb{P}[X \geq \mu + \lambda] \leq \exp\left(-\frac{\lambda^2}{2\mu + 2\lambda/3}\right). \tag{2.1}$$

If $\lambda < \mu$ we have,

$$\mathbb{P}[X \geq \mu - \lambda] \leq \exp\left(-\frac{\lambda^2}{2\mu}\right). \tag{2.2}$$

Sketch of proof. Write $X = \sum_{i=1}^n B_i$ as a sum of independent Bernoulli variables. Then we have by independence and by Proposition A.2.2,

$$\mathbb{E}[e^{tX}] = \mathbb{E}[\prod_{i=1}^n e^{tB_i}] = \prod_{i=1}^n \mathbb{E}[e^{tB_i}] = ((1-p) + pe^t)^n \leq \exp p(e^t - 1).$$

So, by Markov inequality (Lemma 2.1.1),

$$\mathbb{P}[X \geq \mu + \lambda] = \mathbb{P}[e^{tX} \geq e^{t\mu + t\lambda}] \leq e^{-t\mu - t\lambda}\mathbb{E}[e^{tX}].$$

Maximizing for t we obtain that the best possible bound is for $e^t = \frac{\mu+\lambda}{\lambda} =: 1+\delta$ and then we obtain

$$\mathbb{P}[X \geq \mu + \lambda] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}$$

Now using a the well known inequality $1 + x \geq \exp \frac{x}{1+x/2}$ we have,

$$\begin{aligned} \mathbb{P}[X \geq \mu + \lambda] &\leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} \\ &\leq \exp\left(\mu\delta - \mu\frac{\delta(1+\delta)}{1+\delta/2}\right) \\ &= \exp\left(-\mu\frac{\delta^2}{2+\delta}\right). \end{aligned}$$

Equation (2.2) is obtained in a similar form. □

Example 2.1.9. We will use now Chernoff bounds (Lemma 2.1.8) in a descriptive example that will be useful latter on. It is indeed a Lemma from [12, Chapter 4] that will be used to find big matchings.

Take the $G(n, p)$ model with $np \leq 2\log n$.

Let X be the number of vertices with degree at least $8\log n$ and let I_v be the indicator variable for the event "v has degree at least $8\log n$". Then $X = \sum_v I_v$ and $\mathbb{E}[X] = \sum_v \mathbb{E}[I_v] = n\mathbb{P}[I_v = 1]$. We will show that a.a.s. $X = 0$.

According to 1st moment method (Lemma 2.1.3), in order to show that $X = 0$ a.a.s. we have only to show that $\mathbb{E}[X] = o(1)$. However, let $\deg(v) = B_v \sim \mathrm{Bin}(n-1, p)$, so that $\mathbb{P}[I_v = 1] = \mathbb{P}[B \geq 8\log n]$. Note that $\mathbb{E}[B] = (n-1)p \leq 2\log n$ so, from Chernoff bounds (Lemma 2.1.8),

$$\begin{aligned} \mathbb{P}[B \geq 8\log n] &\leq \mathbb{P}[B \geq \mathbb{E}[B] + 6\log n] \\ &\leq \exp\left(-\frac{(6\log n)^2}{2\mathbb{E}[B] + 2\cdot 6\log n/3}\right) \\ &\leq \exp\left(-\frac{(6\log n)^2}{8\log n}\right) \\ &= n^{-9/2}. \end{aligned}$$

So $\mathbb{E}[X] = n\mathbb{P}[I_v = 1] \leq n^{-7/2} = o(1)$ concludes that a.a.s. the maximum degree is smaller than $8\log n$.

This concludes the probabilistic tools that we want to use henceforth.

2.2 Results on the random graph models $G(n,p)$ and $G(n,M)$

We have already introduced the $G(n,p)$ model, or the Erdös-Rényi model, in Chapter 1. We recall now the $G(n,M)$ model.

Definition 2.2.1. The $G(n,M)$ model picks, uniformly at random, a graph with exactly M edges within a given n vertex set. It is usual to see this model as

$$G(n,M) = [G(n,p)|\#E(G(n,p)) = M],$$

for $p \neq 0,1$.

So, each graph with M edges and n vertices has probability $\binom{\binom{n}{2}}{M}^{-1}$ to be picked. Note that it is easy to construct $G(n,M+1)$ from $G(n,M)$: Let G_+ be a random graph obtained from $G(n,M)$ by adding a new edge, picked uniformly among the remaining edges. Then $G_+ \sim G(n,M+1)$.

We will also recall the bipartite random model.

Definition 2.2.2. The random graph $G(m,n,p)$, given two vertex sets A,B of size m,n respectively, has an edge between two vertices from A to B with probability p independent from the remaining edges.

2.2.1 Duality of the models - Asymptotic Equivalence

In this section we will discuss in detail how we can carry some results from the $G(N,M)$ model to $G(N,p)$ and back. Namely, asymptotic results on whether $G(N,p)$ satisfies an increasing property, will have dual results in $G(N,M)$. This is based on the explanation in [Chapter 1 12, Corollary 1.16], that regards a broader context of random sets, unlike here, where we will deal with random graphs.

Theorem 2.2.3. Let $N(n)$ be a sequence with $\lim N(n) = \infty$, and Q be an increasing property of graphs. Let $M = M(n) \in \{0,\cdots,\binom{N(n)}{2}\}$ such that both $M \to \infty$ and $\binom{N}{2} - M \to \infty$, call $p(n) = \frac{M}{\binom{N(n)}{2}}$ and let $\delta = \delta(n) > 0$ that satisfies $0 \leq (1\pm\delta)p(n) \leq 1$ for all n and $\delta\sqrt{M} = \omega(1)$, then:

- If $\mathbb{P}[G(N,p) \in Q] \to 1$ then $\mathbb{P}[G(N,M) \in Q] \to 1$.

- If $\mathbb{P}[G(N,p) \in Q] \to 0$ then $\mathbb{P}[G(N,M) \in Q] \to 0$.
- If $\mathbb{P}[G(N,M) \in Q] \to 1$ then $\mathbb{P}[G(N,(1+\delta)p) \in Q] \to 1$.
- If $\mathbb{P}[G(N,M) \in Q] \to 0$ then $\mathbb{P}[G(N,(1-\delta)p) \in Q] \to 0$.

This theorem represents a "rule of thumb" that we will apply in what follows from this work: whenever we have a result in the $G(N,p)$ realm we can readily translate to the $G(N,M)$ realm, and the reverse can be done without much cumbersome remarks.

A little side note regarding δ must be given, as to go from the claim $\mathbb{P}[G(N,M) \in Q] \to 1$ to the claim $\mathbb{P}[G(N,p) \in Q] \to 1$ is harder than the other way around (thing about Q = "graph has at least M edges"). The variable δ represents only how much should we over estimate p in order to obtain the duality, which can be thought of as a constant as small as one wants, or as a function on n, which should be small, of order $O\left(\frac{1}{\sqrt{pN^2}}\right)$.

Lemma 2.2.4. Let Q be an arbitrary property of graphs, $p(n) \in [0,1]$ a sequence of probabilities and $a \in [0,1]$. If, for every $M = M(n)$ such that $M(n) - p(n)\binom{N(n)}{2} = O\left(\sqrt{\binom{N}{2}p(1-p)}\right)$ it holds

$$\mathbb{P}[G(N,M) \in Q] \to a,$$

then it holds that

$$\mathbb{P}[G(N,p) \in Q] \to a.$$

Proof. For each n, let $\mathscr{M}_n(C) = \mathscr{M}(C) = \{M : |M - \binom{N(n)}{2}p(n)| \leq C\sqrt{\binom{N(n)}{2}p(n)(1-p(n))}\}$. Let $M_*(n) = M_*$ be the element in $\mathscr{M}_n(C)$ that minimizes $\mathbb{P}[G(N,M_*) \in Q]$. Note that M_* depends on C. Then, for a fixed n and C,

$$\begin{aligned}
\mathbb{P}[G(N,p) \in Q] &= \sum_{M=0}^{\binom{N}{2}} \mathbb{P}[G(N,p) \in Q | \#E(G(N,p)) = M]\mathbb{P}[\#E(G(N,p)) = M] \qquad (2.3)\\
&= \sum_{M=0}^{\binom{N}{2}} \mathbb{P}[G(N,M) \in Q]\mathbb{P}[\#E(G(N,p)) = M]\\
&\geq \sum_{M \in \mathscr{M}(C)} \mathbb{P}[G(N,M_*) \in Q]\mathbb{P}[\#E(G(N,p)) = M]\\
&= \mathbb{P}[G(N,M_*) \in Q] \times \mathbb{P}[\#E(G(N,p)) \in \mathscr{M}(C)].
\end{aligned}$$

Now for C a constant and $n \to \infty$, $M_* = \binom{N}{2}p + O\left(\sqrt{\binom{N}{2}p(1-p)}\right)$ so, $\mathbb{P}[G(N,M_*) \in Q] \to a$. To bound $\mathbb{P}[\#E(G(N,p)) \in \mathscr{M}(C)]$ we use Chebichev's inequality (Lemma 2.1.2) to show that $\#E(G(N,p))$, a binomial distributed variable, is concentrated around its mean.

$$\begin{aligned}\mathbb{P}[\#E(G(N,p)) \notin \mathcal{M}(C)] &= \mathbb{P}\left[|\#E(G(N,p)) - \binom{N}{2}p| > C\sqrt{\binom{N}{2}p(1-p)}\right] \\ &\le \frac{\binom{N}{2}p(1-p)}{C^2\binom{N}{2}p(1-p)} \\ &= C^{-2}.\end{aligned}$$

So $\mathbb{P}[\#E(G(N,p)) \in \mathcal{M}(C)] \ge 1 - C^{-2}$ and, from (2.3) we have that

$$\liminf_n \mathbb{P}[G(N,p) \in Q] \ge a(1-C^{-2}).$$

To estimate the upper bound, let M^* be the element in $\mathcal{M}_n(C)$ that maximizes $\mathbb{P}[G(N,M^*) \in Q]$ obtaining:

$$\begin{aligned}\mathbb{P}[G(N,p) \in Q] &= \sum_{M=0}^{\binom{N}{2}} \mathbb{P}[G(N,p) \in Q | \#E(G(N,p)) = M]\mathbb{P}[\#E(G(N,p)) = M] \\ &= \sum_{M=0}^{\binom{N}{2}} \mathbb{P}[G(N,M) \in Q]\mathbb{P}[\#E(G(N,p)) = M] \\ &\le \left(\sum_{M\in\mathcal{M}(C)} \mathbb{P}[G(N,M^*) \in Q] \times \mathbb{P}[\#E(G(N,p)) = M]\right) \\ &\quad + \sum_{M\notin\mathcal{M}_n(C)} 1 \times \mathbb{P}[\#E(G(N,p)) = M] \\ &= \mathbb{P}[G(N,M^*) \in Q]\mathbb{P}[\#E(G(N,p)) \in \mathcal{M}(C)] + \mathbb{P}[\#E(G(N,p)) \notin \mathcal{M}(C)].\end{aligned} \tag{2.4}$$

Since $M^* = \binom{N}{2}p + O\left(\sqrt{\binom{N}{2}p(1-p)}\right)$ for C fixed, we get that $\mathbb{P}[G(N,M^*) \in Q] \to a$

We also already know that $\mathbb{P}[\#E(G(N,p)) \notin \mathcal{M}(C)] \le C^{-2}$ so $\limsup_n \mathbb{P}[G(N,p) \in Q] \le a + C^{-2}$.

Then, since we have shown the previous for generic G, taking $C \to \infty$ gets us $\limsup_n \mathbb{P}[G(N,p) \in Q] = \liminf_n \mathbb{P}[G(N,p) \in Q] = a$. □

Lemma 2.2.5. Let Q be a monotone property of graphs, and let $M = M(n) \in \{0,\dots,\binom{N(n)}{2}\}$. Set $a \in [0,1]$ and suppose that for any $p = p(n) \in [0,1]$ such that $p - \frac{M}{\binom{N}{2}} = O\left(\sqrt{\frac{M(\binom{N}{2}-M)}{\binom{N}{2}^3}}\right)$, it holds

$$\mathbb{P}[G(N,p) \in Q] \to a,$$

then it holds

$$\mathbb{P}[G(N,M) \in Q] \to a\,.$$

Proof. We can assume that $M \neq 0, \binom{N}{2}$ from some order on, because if $\liminf M = 0$, for instance, then $M = 0$ for all n from some order on, and $G(n,M)$ is just the empty graph which makes the claim trivial. Similarly for $M = \binom{N}{2}$.

Thus, assume that $M \neq 0, \binom{N}{2}$ from some order on. We also assume wlog that Q is increasing. Call $p_0 = \frac{M}{\binom{N}{2}}$ and $q_0 = 1 - p_0$. Set a constant C. Let $p_+ = p_0 + C\sqrt{\frac{p_0 q_0}{\binom{N}{2}}}$ and $p_- = p_0 - C\sqrt{\frac{p_0 q_0}{\binom{N}{2}}}$, with a possible correction so $p_+, p_- \in [0,1]$.

So we have:

$$\begin{aligned}
\mathbb{P}[G(N,p_+) \in Q] &\geq \sum_{M' \geq M} \mathbb{P}[G(N,M') \in Q]\mathbb{P}[\#E(G(n,p_+)) = M'] \\
&\geq \mathbb{P}[G(N,M) \in Q]\mathbb{P}[\#E(G(n,p_+)) \geq M] \\
&= \mathbb{P}[G(N,M) \in Q](1 - \mathbb{P}[\#E(G(n,p_+) < M]) \\
&\geq \mathbb{P}[G(N,M) \in Q] - \mathbb{P}[\#E(G(n,p_+)) < M]\,.
\end{aligned} \tag{2.5}$$

And as well:

$$\begin{aligned}
\mathbb{P}[G(N,p_-) \in Q] &\leq \mathbb{P}[\#E(G(N,p_-)) > M] + \sum_{M' \leq M} \mathbb{P}[G(N,M') \in Q]\mathbb{P}[\#E(G(n,p_-)) = M'] \\
&\leq \mathbb{P}[\#E(G(N,p_-)) > M] + \mathbb{P}[G(N,M) \in Q]\mathbb{P}[\#E(G(n,p_+)) \leq M] \\
&\leq \mathbb{P}[\#E(G(N,p_-)) > M] + \mathbb{P}[G(N,M) \in Q]\,.
\end{aligned} \tag{2.6}$$

Now we estimate the variance of both $E_- = \#E(G(N,p_-))$ and $E_+ = \#E(G(N,p_+))$. So,

$$\begin{aligned}
\mathrm{Var}[E_-] &= Np_-(1-p_-) \\
&= \binom{N}{2}(p_0 - C\sqrt{p_0 q_0 / \binom{N}{2}})(1 - p_0 + C\sqrt{p_0 q_0 / \binom{N}{2}}) \\
&= \binom{N}{2} p_0 q_0 + C\sqrt{p_0 q_0 \binom{N}{2}}(p_0 - q_0) - C^2 p_0 q_0 \\
&\leq \binom{N}{2} p_0 q_0 + C\sqrt{p_0 q_0 \binom{N}{2}},
\end{aligned}$$

as well as

$$\begin{aligned}
\mathrm{Var}[E_+] &= Np_+(1-p_+) \\
&= \binom{N}{2}(p_0 + C\sqrt{p_0 q_0/\binom{N}{2}})(1 - p_0 - C\sqrt{p_0 q_0/\binom{N}{2}}) \\
&= \binom{N}{2} p_0 q_0 + C\sqrt{p_0 q_0 \binom{N}{2}}(q_0 - p_0) - C^2 p_0 q_0 \\
&\leq \binom{N}{2} p_0 q_0 + C\sqrt{p_0 q_0 \binom{N}{2}}.
\end{aligned}$$

So we get, by Chebichev Inequality (Lemma 2.1.2),

$$\begin{aligned}
\mathbb{P}[E_- > M] &= \mathbb{P}[E_- > \mathbb{E}[E_-] + \binom{N}{2}(p - p_-)] \\
&\leq \frac{\mathrm{Var}[E_-]}{\binom{N}{2}^2 (p_0 - p_-)^2} \leq \frac{\binom{N}{2} p_0 q_0 + C\sqrt{p_0 q_0 \binom{N}{2}}}{C^2 \binom{N}{2} p_0 q_0} \\
&= C^{-2} + C^{-1}\frac{1}{\sqrt{p_0 q_0 \binom{N}{2}}} \leq C^{-2} + \sqrt{2} C^{-1} =: \delta(C),
\end{aligned}$$

from the fact that $M \neq 0, \binom{N}{2}$. To obtain that $\mathbb{P}[E_+ < M] \leq \delta(C)$ we argue in the same way.

Now going back to equations (2.5) and (2.6) we get that $\mathbb{P}[G(N, p_+) \in Q] \geq \mathbb{P}[G(N, M) \in Q] - \delta(C)$ and $\mathbb{P}[G(N, p_-) \in Q] \leq \mathbb{P}[G(N, M) \in Q] + \delta(C)$.

So we have $\mathbb{P}[G(N, p_+) \in Q] + \delta(C) \geq \mathbb{P}[G(N, M) \in Q] \geq \mathbb{P}[G(N, p_-) \in Q] - \delta(C)$, and making $n \to \infty$ we get

$$a + \delta(C) \geq \lim \mathbb{P}[G(N, M) \in Q] \geq a - \delta(C).$$

Since C is arbitrary, we get the lemma. □

We will write here a simplification of Lemma 2.2.5, where we provide a convexity condition over property Q and set $a = 1$, obtaining:

Lemma 2.2.6. Let Q be a convex graph property and $M = M(n) \in \{0, \dots, \binom{N(n)}{2}\}$. Let $p(n) = \frac{M(n)}{\binom{N(n)}{2}}$. If it holds

$$\mathbb{P}[G(N, p) \in Q] \to 1,$$

then it holds

$$\mathbb{P}[G(N, M) \in Q] \to 1.$$

Proof. Let M_1 and M_2 be integers such that $M_1 \leq M \leq M_2$ and both maximize $\mathbb{P}[G(N, M_i) \in Q]$ subject to that.

Then we have

$$\begin{aligned}
\mathbb{P}[G(N,p) \in Q] &\leq \sum_{m=0}^{\binom{N}{2}} \mathbb{P}[G(N,m) \in Q]\mathbb{P}[\#E(G(N,p)) = m] \\
&\leq \mathbb{P}[G(N,M_1) \in Q]\mathbb{P}[\#E(G(N,p)) \leq M] + \mathbb{P}[\#E(G(N,p)) > M]\,.
\end{aligned}$$

For $n \to \infty$, since $\lim_n \mathbb{P}[\#E(G(N,p)) > M] = \lim_n \mathbb{P}[\#E(G(N,p)) \leq M] = \frac{1}{2}$ by the central limit theorem, this, with Lemma's hypothesis, yields

$$1 \leq \frac{1}{2} \liminf_n \mathbb{P}[G(N,M_1) \in Q] + \frac{1}{2}$$

Which implies that $\lim_n \mathbb{P}[G(N,M_1) \in Q] = 1$. From the similar inequality $\mathbb{P}[G(N,p) \in Q] \leq \mathbb{P}[G(N,M_2) \in Q]\mathbb{P}[\#E(G(N,p)) \geq M] + \mathbb{P}[\#E(G(N,p)) < M]$ we get that $\lim_n \mathbb{P}[G(N,M_2) \in Q] = 1$.

Now we use the fact that a convex property Q is the intersection of an increasing property Q' and a decreasing property Q''; specifically, Q' is the set of all graphs G such that there is a graph $G' \leq G$ that has the property Q, and Q'' is the set of all graphs G such that there is a graph $G'' \geq G$ such that $G'' \in Q$. Trivially $Q = Q' \cap Q''$.

Then

$$\begin{aligned}
\mathbb{P}[G(N,M) \in Q] &\geq \mathbb{P}[G(N,M) \in Q'] + \mathbb{P}[G(N,M) \in Q''] - 1 \\
&\geq \mathbb{P}[G(N,M_1) \in Q'] + \mathbb{P}[G(N,M_2) \in Q''] - 1 \\
&\geq \mathbb{P}[G(N,M_1) \in Q] + \mathbb{P}[G(N,M_2) \in Q] - 1\,.
\end{aligned}$$

And so $\lim_n \inf \mathbb{P}[G(N,M) \in Q] \geq \lim_n \inf \mathbb{P}[G(N,M_1) \in Q] + \mathbb{P}[G(N,M_2) \in Q] - 1 = 1$. □

Remark 2.2.7. The last part of this proof shows that a convex property and three numbers $M_1 \leq M \leq M_2$ satisfy

$$\mathbb{P}[G(N,M) \in Q] \geq \mathbb{P}[G(N,M_1) \in Q] + \mathbb{P}[G(N,M_2) \in Q] - 1\,.$$

The simple proof here presented was due to Johan Jonasson.

Proof of Theorem 2.2.3. Recall that there are four statements.

The first two were that if $\mathbb{P}[G(N,p) \in Q] \to 1,0$ then $\mathbb{P}[G(N,M) \in Q] \to 1,0$, respectively, for $M = \binom{N}{2}p$.

Both are a simple consequence from Lemma 2.2.6, where the 0 case we apply the Lemma to the complementary property.

Now we suppose that $\mathbb{P}[G(N,M) \in Q] \to 1$. Then if $M' \geq M$ for some order on, we have that $\mathbb{P}[G(N,M') \in Q] \geq \mathbb{P}[G(N,M) \in Q] \to 1$.

In particular, if $M' - (1+\delta)p\binom{N}{2} = O\left(\sqrt{\binom{N}{2}p(1-p)}\right)$, or equivalently,

$$\begin{aligned}
M' &= (1+\delta)p\binom{N}{2} + O\left(\sqrt{\binom{N}{2}p(1-p)}\right) \\
&= M + \delta p\binom{N}{2} + O\left(\sqrt{\binom{N}{2}p(1-p)}\right) \\
&= M + p\binom{N}{2}\left(\delta + O(M^{-1/2}\sqrt{1-p})\right),
\end{aligned}$$

If we recall that $\delta = \omega(M^{-1/2})$, then $M' \geq M$ from some order on and consequently $\mathbb{P}[G(N,M') \in Q] \to 1$, as $p\binom{N}{2} = M \to \infty$ and $\sqrt{1-p}$ is bounded.

Hence, Lemma 2.2.4 implies that $\mathbb{P}[G(n,(1+\delta)p) \in Q] \to 1$.

Similarly we get that, if $\mathbb{P}[G(n,M) \in Q] \to 0$ then $\mathbb{P}[G(n,(1-\delta)p) \in Q] \to 0$. □

2.2.2 Phase Transition

In this section we will give an overview on the properties of $G(n,M)$ for M varying from 0 to $\binom{n}{2}$. We will focus on the events "$G(n,M)$ has a component of size k" for varying k and "$G(n,M)$ has a perfect matching".

We will see in this thesis that $G(n,M)$ becomes connected at the same time that the last isolated vertex gets an incident edge, which is when a perfect matching appears if n is even. This, according to the duality theorems in the previous section, will translate naturally between the $G(n,M)$ model and the $G(n,p)$ model.

In what follows we will discuss some little remarks on what happens to $G(n,M)$ when M goes from 0 to $\binom{N}{2}$.

For M small some small trees occur. For fixed k, at $M = \omega(n^{\frac{k-2}{k-1}})$, a given tree with k vertices occurs as subgraphs a.a.s., where we mean subgraph in the sense described in the introduction, according to [12, Chapter 3] in Theorem 3.1.2.

When $M = o(n)$ there are no cycles, according to Theorem 3.1.2, when $M/n = c + o(1)$ for $c < \frac{1}{2}$ the biggest component has size $\Theta(\log n)$ and for $c > \frac{1}{2}$ it has size $\Theta(n)$, as we will show in Lemma 2.2.8. During this transition, any given graph G occurs as a subgraph a.a.s. only if G has only trees or unicycles. A little bit afterwards, when $M = \omega(n^{4/3})$, there are arbitrarily many 4-cliques. As we have seen, when $M = \frac{1}{2}n(\log n + \Theta(1))$, the last isolated vertex gets an incident edge and, at the same time, we will see that for n even, a perfect matching appears. Also the graph becomes connected at this moment.

At $M = \frac{n}{2}\log n + (k-1)\frac{n}{2}\log\log n + \omega(n)$, for k constant, the minimal degree is k, according to [12]. For $M = a\binom{n}{2}$ where $a \in (0,1)$, which corresponds to $p = \Theta(1-p) = \Theta(0.5)$ in the $G(n,p)$ model, any two vertices share $\Theta(n)$ neighbours, as it is easy to show. It is also easy to see that the biggest complete subgraph and the biggest independent set both have size $\Theta(\log n)$.

And now we can tell the tale backwards: When $M = \binom{n}{2} - o(n^{4/3})$ there is no independent set of size 4 and when $M = \binom{n}{2} - o(n\log n)$ and n is even, there is no stable set of type $\lambda = 2^{n/2}$ and some vertices become connected to all others.

We will see now that the biggest component grows from $\Theta(\log n)$ to $\Theta(n)$ in size when $p(n) = \Theta(n)$.

Lemma 2.2.8. Let $np(n) = \beta$, where β is a constant. If $\beta < 1$, then

$$\mathbb{P}\left[c_1(G(n,p)) \le \frac{3\log n}{(1-\beta)^2}\right] \to 1\,.$$

We first introduce the Branching process, a stochastic process that can be readily applied in the previous lemma.

Definition 2.2.9 (Markov chains and branching processes)**.** A *Markov chain* is a stochastic process $\{X_n\}$ that satisfies the Markov Chain Property:

$$\forall k,j\ \mathbb{P}[X_{n+1} = k | X_n = j] = \mathbb{P}[X_{n+1} = k | X_n = j,\ X_i = j_i,\ i = 1,\dots,n-1]\,.$$

Markov chains are a thoroughly studied subject and have many applications, in particular in random graphs. There are introductory works on the topic in [17] and some applications in [15].

Let T be a distribution over the non-negative integers, then a *branching process* with distribution T is a Markov Chain such that $X_0 = 1$ and $X_{n+1} = \sum_{i=1}^{X_n} T_{i,n}$ where $T_{i,n} \sim T$ are independent and identically distributed random variables.

Note that we require the random variables $T_{i,n}$ to be independent and identically distributed. In particular, $T_{i,n+1}$ is independent from the value of X_n and for that reason we should consider that we generate infinitely many variables $T_{i,n+1}$, for $i = 1, 2, \ldots$, in theory, regardless of how many we end up using.

Intuitively, a branching process models the reproduction of cells where, at each (discrete) moment any currently alive cell dies leaving offspring regulated by the distribution T, independently from the offspring of the remaining cells. The Markov chain counts the number of cells that are alive at each moment.

The application to connected components is clear, when components are small: starting at a vertex, its offspring is the set of new neighbour vertices. Problems arise when we revisit vertices which result in some over-counting - nevertheless, studying the behaviour and growth of a branching process gives us a bound to the size of the connected component.

Remark 2.2.10. A special case of branching process is the binomial case $\{X_i\}$, where the offspring distribution is $T \sim \mathrm{Bin}(n, p)$, where we regard that $n \to \infty$ and $\lim np(n) = \beta > 1$.

Consider the event E_n ="The branching process gets extinguished", equivalently

$$E_n :\Leftrightarrow \lim_n \mathbb{P}[X_n = 0] = 1\,.$$

To evaluate $\rho(\mathrm{Bin}(n, p)) := \mathbb{P}[E_n]$, we condition the probability to have

$$\mathbb{P}[E_n] = \sum_{k\geq 0} \mathbb{P}[E_n | X_1 = k]\mathbb{P}[X_1 = k] = \sum_{k\geq 0} \rho^k \binom{n}{k} p^k (1-p)^k = (1 - p + \rho p)^n\,.$$

Note that the probability of extinction of a branching process that has k cells alive is, by independence, $\mathbb{P}[E_n]^k$. So, for n big we can use the approximation $(1-p+\rho p)^n = \exp(-\beta(1-\rho)) + O(np^2(1-\rho)^2)$ from the exponential power series, which normally is a good enough approximation because $p = \Theta(n^{-1})$, and hence the error term becomes $np^2(1-\rho)^2 = \Theta(n^{-1}) = o(1)$.

Then we claim that the equation

$$\rho = \exp(-\beta(1-\rho))\,. \tag{2.7}$$

Gives the approximation to the extinction probability of the branching process, for n big, by Hurwitz's Theorem. Let $\alpha(\beta) = \rho \in (0, 1)$ be the real that solves the equation, Such $\alpha(\beta)$ exists for$\beta > 1$: The right hand side of equation (2.7) is a convex function on ρ, so a line intersects its graph at most twice. Since we have at $\rho = 1$ we have equality, there is at most one solution in $(0,1)$. The fact $\frac{\partial}{\partial\rho}\Big|_{\rho=1} \exp(-\beta(1-\rho)) = \beta$ guarantees that if $\beta > 1$, another solution exists in $\rho \in (0,1)$.

If $\beta \leq 1$, then the branching process will be extinguished almost surely, as the unique solution to the Equation (2.7) in the interval $[0,1]$ is $\rho = 1$.

We briefly observe that in the case $\beta > 1$ the solution that we want is indeed the one with $\rho \neq 1$, i.e. that for $\beta > 1$ there is a non-zero probability that the Markov chain runs forever: we can write $X_{i,n}$ for the offspring of different cells and set $T = \sum_{i,n \text{ until k-th cell}} X_{i,n} \sim \text{Bin}(nk, p)$, then Chernoff bounds (Lemma 2.1.8) tell us that $\mathbb{P}[T < k+1]$ is exponentially small and a union bound argument tells us that if we sum that over all cells we get a number smaller than one for the upper bound of the extinction probability ρ.

Some twists can be done to the branching process, preserving some properties. For instance, if we want to estimate the size of the branching process by deleting some cells before having offspring, we obtain a different random variable that is *stochastically smaller* - see Appendix B.

Proof of Lemma 2.2.8. Fix a vertex v and consider the process given by $S_0 = \{v\}$, and $S_n = N(S_{n-1}) \cup S_{n-1}$. Consider as well $X_n = \#(S_n \setminus S_{n-1})$. To obtain the Lemma, our goal is to show:

$$\mathbb{P}\left[\sum_k X_k > \frac{3\log n}{(1-\beta)^2}\right] = o(n^{-1}). \tag{2.8}$$

Note that equation (2.8) is enough to show the Lemma, from an union bound argument: denote E_v the event that the connected component of v has size bigger than $\frac{3\log n}{(1-\beta)^2}$, then, using the hypothesis (2.8) we get

$$\mathbb{P}[\cup_v E_v] \leq \sum_v \mathbb{P}[E_v] = n\, o(n^{-1}) = o(1),$$

concluding that all vertices are in small components a.a.s.

So in order to obtain (2.8), we consider now $\{Y_n\}$, a branching process with distribution $T \sim \text{Bin}(n-1, p)$. Our goal is to compare the random processes X_n and Y_n and once we established that the random variables X_n are statistically smaller than Y_n, to obtain (2.8) it is only necessary to show that

$$\mathbb{P}\left[\sum_k Y_k > \frac{3\log n}{(1-\beta)^2}\right] = o(n^{-1}). \tag{2.9}$$

This last equation is a consequence form Chernoff bound (Lemma 2.1.8) that we will prove now, and we postpone for latter the proof that the random variables X_n are statistically smaller than Y_n. Namely, we denote by $Y_{k,n} \sim \text{Bin}(n-1,p)$ the number of offspring of the k-th element of the n-th generation, and let's consider the total offspring of the k_t first cells

$$Y^{(k_t)} = \sum_{k_t \text{ first cells}} Y_{k,n} \sim \text{Bin}(k_t(n-1),p)\,.$$

Note that $Y^{(k_t)} \le k_t - 2 \Rightarrow$ the process stops with at most k_t cells.

So, if we prove that $Y^{(k_t)} \le k_t - 2$ with high probability, then the branching process doesn't yield more than k_t cells with high probability. Recall that $\mathbb{E}[Y^{(k_t)}] = k_t(n-1)p = k_t(\beta - p)$.

Chernoff bounds (Lemma 2.1.8) gives us the following, for n big enough, suitable small $\epsilon_i > 0$ and for $k_t = \frac{3\log n}{(1-\beta)^2}$.

$$\begin{aligned}
\mathbb{P}[Y^{(k_t)} \ge k_t - 1] &= \mathbb{P}[Y^{(k_t)} \ge k_t(\beta - p) + k_t(1 + p - \beta - k_t^{-1})] \\
&\le \exp\left(-\frac{k_t^2(1+p-\beta-k_t^2)^2}{2(k_t(\beta-p)+k_t(1+p-\beta-k_t^{-1})/3)}\right) \\
&= \exp\left(-\log n \frac{3(1+p-\beta-k_t^{-1})^2}{2(1-\beta)^2} \frac{1}{\frac{1-2p+2\beta}{3} - (3k_t)^{-1}}\right) \\
&= \exp\left(-\log n \frac{3+\epsilon_1}{2} \frac{3}{2\beta+1+\epsilon_2}\right).
\end{aligned}$$

With $\epsilon_1, \epsilon_2 \to 0$ we get $\mathbb{P}[Y^{(k_t)} \ge k_t - 1] \le n^{-\frac{3}{2}+\epsilon} = o(n^{-1})$.

Now, to compare between X_n and Y_n we will construct a different process $\{X_n^+\}$ that satisfies $X_n^+ \ge_S X_n$ but $\{X_n^+\}$ and $\{Y_n\}$ follow the same distribution.

Incidentally, $\{X_n\}$ fails to be a branching process with distribution $\text{Bin}(n-1,p)$ when a cycle occurs, i.e. when a vertex v has an already visited neighbour x. When such happens, we add a fake neighbour that will yield offspring according to $\text{Bin}(n-1,p)$ distribution.

Specifically, for each vertex x_i that is an already visited neighbour in the process $\{X_k\}$, arrived for the second time at generation $k(i)$, the new process $\{B_{j,i}\}_{j\in\mathbb{N}}$ with distribution $\text{Bin}(n-1,p)$ is added. Then, we have

$$X_n^+ = X_n + \sum_i B_{n-k(i),i}\,.$$

We may add several more processes regarding the same vertex x_i, with a different index, if the same happens again in a different cycle. However, a fixed cycle is only followed once, as we never consider the offspring form the repeated occurrences of a vertex.

It is obvious that $X_n^+ \geq_S X_n$ and that $\{X_n^+\}$ is a branching process with distribution $T \sim \mathrm{Bin}(n-1, p)$, so the process $\{X_n^+\}$ dies out after k_t generation with high probability, and consequently $\{X_n\}$ also dies out after k_t generation with high probability. This concludes the Lemma. □

Lemma 2.2.11. Let $np(n) = \beta$, where β is a constant. If $\beta > 1$, let $\alpha = \alpha(\beta)$ be the solution in $(0,1)$ of the equation (2.7).

Then $G(n,p)$ contains a component of size $n(1-\alpha(\beta)) + o(n)$. Furthermore, the size of the second largest component is a.a.s. at most $\frac{16\beta\log n}{(\beta-1)^2}$.

Sketch of proof. We will put the emphasis of the details of this proof in the construction of the relevant branching processes, and state some facts that can be checked in the source of this proof in [12].

Let $k_- = \frac{16\beta}{(\beta-1)^2}\log n$ and $k_+ = n^{2/3}$. As previously, start the process $\{X_i\}$ at the vertex v counting the number of active neighbours. Call K_v the number of vertices in the connected component of v. In this proof we will consider that at each step we collect only the offspring of one vertex per round, instead of the whole generation at the same time, which translates only to a difference in the indices, which get delayed (for instance, the process will take more rounds to stop, but will have the same number of vertices in the end).

Let us outline the proof. We will show that, if this process has survived to have at least k vertices with $k \in (k_-, k_+)$, then it still has some considerable amount of active vertices with probability $1 - o((nk_+)^{-1})$, using Chernoff bounds (Lemma 2.1.8). Then, a union bound argument implies that the branching process never stops in (k_-, k_+) for all vertices, this implies that either connected components are big, of size $\Omega(k_+)$, or small, of size $O(k_-)$. Afterwards, we show that any two big components will have too many active vertices after k_+ vertices have been evaluated, so there is an edge between these active vertices with high probability, which establishes that there is at most one big component. Finally, we show that one big component must exist by bounding the number of vertices outside such component, evaluating the extinction probability of their branching processes.

We will show, for each integer k such that $k_+ \geq k > k_-$, that we have with high probability

$$\sum_{i=1}^{k} X_k > k - 1 + (\beta - 1)\frac{k}{2}. \tag{2.10}$$

Let E_v^k be the event "Equation (2.10) holds for the branching process starting at v", which is to say that the branching process has, after k vertices have been observed, more than $(\beta-1)\frac{k}{2}$ active vertices. Let F_v^k be the negation event of E_v^k. Our goal is to show that E_v^k occurs for any vertex v and any k such that $k_- \le k \le k_+$, by proving the following bound:

$$\mathbb{P}[F_v^k] \le \exp\left(-\frac{(\beta-1)^2 k}{9\beta}\right), \tag{2.11}$$

and apply the union bound to obtain the desired:

$$\begin{aligned}
\mathbb{P}[\bigcup_{k=k_-}^{k_+} \bigcup_v F_v^k] &\le \sum_{k=k_-}^{k_+} \sum_v \exp\left(-\frac{(\beta-1)^2 k}{9\beta}\right) \\
&\le n k_+ \exp\left(-\frac{(\beta-1)^2 k_-}{9\beta}\right) \\
&= \exp(\log n + \frac{2}{3}\log n - \frac{16}{9}\log n) \\
&= n^{-1/16} = o(1).
\end{aligned}$$

To obtain Equation (2.11), we evaluate $\mathbb{P}[F_v^k] = \mathbb{P}[\sum_{i=1}^k X_i \le k-1+(\beta-1)\frac{k}{2}]$ by bounding each X_k from below: we count just the neighbours of a vertex that occur in a set of size $n - \frac{\beta+1}{2}k_+$ that does not contain any of the already visited vertices, which can always be done until the process reaches $k \le k_+$. So, if $X_k^- \sim \text{Bin}(n - \frac{\beta+1}{2}k_+, p)$ then $\sum_{i=1}^k X_k$ is stochastically bigger than $T := \sum_{i=1}^k X_k^- \sim \text{Bin}\left(k\left(n - \frac{\beta+1}{2}k_+\right), p\right)$. Note that

$$\mathbb{E}[T] = kp\left(n - \frac{\beta+1}{2}k_+\right) = k\beta(1-o(1)),$$

so by Chernoff bounds (Lemma 2.1.8) we have, for small $\epsilon > 0$ and n big enough:

$$\begin{aligned}
\mathbb{P}[T \le k-1+\frac{\beta-1}{2}k] &= \mathbb{P}[T \le \mathbb{E}[T] - k\beta(1-o(1)) + k + \frac{\beta-1}{2}k - 1] \\
&= \mathbb{P}[T \le \mathbb{E}[T] - k\beta(\frac{1}{2} - o(1)) + \frac{1}{2}k - 1] \\
&= \mathbb{P}[T \le \mathbb{E}[T] - \frac{1}{2}k(\beta(1-o(1)) - 1) - 1] \\
&\le \exp\left(-\frac{k^2\frac{1}{4}(\beta(1-o(1)) - 1 - 2k^{-1})^2}{k\beta(1-o(1))}\right) \\
&= \exp\left(-\frac{k}{4\beta}\frac{(\beta(1-o(1)) - 1 - 2k^{-1})^2}{(1-o(1))}\right) \\
&\le \exp\left(-\frac{k}{4\beta}\frac{4(\beta-1)^2}{9}\right) = \exp\left(-\frac{k(\beta-1)^2}{9\beta}\right),
\end{aligned}$$

which concludes the Equation (2.11).

The fact that there are no two different components of size $\Theta(n)$ follows: from the previous, after the process finds k_+ vertices in the connected component of v, $\frac{\beta-1}{2}k_+$ vertices are yet to be evaluated and are not saturated.

So, if v, v' are two vertices, then the probability that both belong to different connected components of size at least k_+ is smaller than the probability that the two sets of $\frac{\beta-1}{2}k_+$ non-saturated vertices at the k^+ step of each branching process have no edge between them, which has probability:

$$(1-p)^{\left(\frac{\beta-1}{2}k_+\right)^2} \leq \exp\left(-p\left(\frac{\beta-1}{2}k_+\right)^2\right) = \exp\left(-\frac{-\beta(\beta-1)^2}{4}n^{1/3}\right) = o(n^{-2/3})\,.$$

Here we used Proposition A.2.2. Now since there are at most $n^{1/3}$ disjoint connected components that reach size $k_+ = n^{2/3}$ in $G(n,p)$, by a union bound argument any two are connected by an edge a.a.s. Hence, there is only one big connected component.

Now it only remains to show how many vertices are in the big component, so let N be the number of vertices outside the big component.

It is clear that the probability $\mathbb{P}[K_v \geq k_-]$ that the process $\{X_i\}$ survives until it has k_- vertices is at least the probability of survival of the branching process with distribution $\mathrm{Bin}(n-k_-, p)$: at each vertex, we choose at random $n-k_-$ vertices among the non-visited ones and consider only the neighbours in that set, so that each offspring generation is an independent binomial distribution. As a consequence, $\mathbb{P}[K_v \geq k_-] \geq 1-\rho(\mathrm{Bin}(n-k_-,p)) = 1-\alpha(\beta)$ for $n \to \infty$, since $(n-k_-)p \to \beta$. Here we are using the notation from Remark 2.2.10.

On the other hand, we can bound $\mathbb{P}[K_v \geq k_-]$ from above with a branching process that has more chance of surviving at least k_- turns, namely the branching process $\{Y_i\}$ with offspring distribution $\mathrm{Bin}(n-1,p)$ as presented in the previous Lemma. Call $K = \sum_i Y_i$, so that $K \geq K_v$.

As it turns out, according to [12], the probability that the branching process $\{Y_i\}$ gets extinguished is highly concentrated in the first iterations, and we can then bound

$$\mathbb{P}[K_v \geq k_-] \leq \mathbb{P}[K \geq k_-] = \mathbb{P}[K=\infty]+\mathbb{P}[K \in (k_-,\infty)] = 1-\rho(\mathrm{Bin}(n-1,p))+o(1) = 1-\alpha(\beta)+o(1)\,,$$

for $n \to \infty$.

The idea of this approximation is that the branching process $\{Y_i(n)\}$ converges to the branching process $\{\mathscr{Y}_i\}$ with underlying distribution $\mathrm{Poi}(\beta)$, so $\mathbb{P}[K_v = k]$ can be computed explicitly by conditioning in the first k many variables, each of with with support in the finite set $\{0,\ldots,k\}$, converges for any k as well. Hence, $k_- \to \infty$ implies that $\mathbb{P}[K \in (k_-,\infty)] = o(1)$.

So $\mathbb{P}[K_v \geq k_-]$ is squeezed in two sequences both converging to $1-\alpha(\beta)$, then $\mathbb{E}[N] = n\alpha(\beta) + o(n)$. Our goal now is to show that $\mathrm{Var}[N] = o(\mathbb{E}[N]^2)$ to apply Chebichev's inequality (Lemma 2.1.2) showing, for any $\epsilon > 0$ constant, that $\mathbb{P}[|N - n\alpha(\beta)| \geq n\alpha(\beta)\epsilon] = o(1)$, concluding the Lemma with $N = n\alpha(\beta)(1+o(1))$.

Indeed, we have, with a simple counting estimate of the number of pairs (v,w) of vertices in small connected components, that

$$\mathbb{E}[N(N-1)] \leq n(\rho(\mathrm{Bin}(n,p)) + o(1))(k_- + n\rho(\mathrm{Bin}(n-k_-,p)) + o(n)) = (1+o(1))\mathbb{E}[N]^2\,,$$

The interested reader can see this computations in [12].

So $\mathrm{Var}[N] = \mathbb{E}[N^2] - \mathbb{E}[N]^2 = \mathbb{E}[N(N-1)] + \mathbb{E}[N] - \mathbb{E}[N]^2 \leq o(1)\mathbb{E}[N]^2 + \mathbb{E}[N]$.

Finally, the big component has $n-N$ vertices, which is

$$n - n(1-\alpha(\beta))(1+o(1)) = n\alpha(\beta)(1+o(1))\,,$$

and this concludes the proof of the Lemma. □

There is still to address the question of what happens when $c=1$. It is observed in [12] that the jump is not abrupt, but has a step in between where $c_1(G(n,p)) = \Theta(n^{2/3})$, for $p = n^{-1} + O(n^{-4/3})$. But for $np = 1 + \omega(n^{-4/3})$ and $np = 1 - \omega(n^{-4/3})$ the behaviour is determined by Lemmas 2.2.8 and 2.2.11.

This means that, between $G(1000,490)$ and $G(1000,510)$ there are huge differences on the size of the biggest component.

2.3 Tree models

For a given integer n, we are interested in the uniform model of trees in a fixed n vertex set. Recall that, in trees, the chromatic symmetric function over the coefficients has a simple combinatorial meaning, introduced in Theorem 1.2.8.

Definition 2.3.1 (Cayley Trees and Random trees)**.** Given a finite set V of vertices, a Cayley tree on V is a tree with vertex set V. If $V = \{1,\cdots,n\}$ we don't mention the set V. The set of Cayley trees of V is denoted by $Cay(V)$, or simply by $Cay(n)$ if $V = \{1,\cdots,n\}$.

The model of random trees is then written as T_n, where T_n takes a tree uniformly at random from the trees in $Cay(n)$, and according to Cayley's formula in [1], for a given tree T with $n \geq 2$ vertices, we have

$$\mathbb{P}[T_n = T] = n^{2-n}\,.$$

In fact, a more general form of Cayley's formula will be of use[2]

Lemma 2.3.2. For a Cayley tree T, let $x_T = \prod_{e\in T} x_{e_1}x_{e_2}$, where e_1 and e_2 represent the two endpoints of the edge e. Then we have that

$$\sum_{T\in Cay(n)} x_1^{\deg_T(1)}\cdots x_n^{\deg_T(n)} = \sum_{T\in Cay(n)} x_T = x_1\cdots x_n(x_1+\cdots+x_n)^{n-2}.$$

Proof. The first equality is clear since $x_T = \prod_i x_i^{\deg_T(i)}$ holds for any tree T.

Now for the second equality to be established, we will show that both sides are polynomials that satisfy several common properties. Call $P_n = \sum_{T\in Cay(n)} x_T = \sum_{T\in Cay(n)} x_1^{\deg_T(1)}\cdots x_n^{\deg_T(n)}$ and $Q_n = x_1\cdots x_n(x_1+\cdots+x_n)^{n-2}$.

Observe that both polynomials P_n and Q_n satisfy the following:

1. Both polynomials are divisible by $x_1\cdots x_n$, as $\deg_T(i)\geq 1$ for any tree T and vertex $i\in V(T)$.
2. For each monomial in each side, there are at least two vertices i,j such that both x_i^2, x_j^2 do not divide the monomial, as each tree has at least two leaves.
3. Both polynomials use only the variables $x_1,\dots,x_n$ and are symmetric in these variables.
4. Both satisfy the functional equation

$$\left.\frac{F_{n+1}}{x_{n+1}}\right|_{x_{n+1}=0} = (x_1+\cdots+x_n)\times F_n\,.$$

Only the last item in the case $F_n = P_n$ is not trivial: Note that the monomials divisible by x_{n+1}^2 vanish when the operator $\Big|_{x_{n+1}=0}$ is applied. So we are interested in the linear part of x_{n+1} and only the trees that have $n+1$ as a leaf contribute to that. So if we separate the sum over the different possible neighbours of $n+1$ we have:

$$\begin{aligned}\left.\frac{P_{n+1}}{x_{n+1}}\right|_{x_{n+1}=0} &= \sum_{\substack{T\in Cay(n+1)\\ n+1\text{ is a leaf}}}\frac{x_T}{x_{n+1}} = \sum_{i=1}^{n}\sum_{\substack{T\in Cay(n+1)\\ n+1\text{ is a leaf connected to } i}}\frac{x_T}{x_{n+1}}\\ &= \sum_{i=1}^{n}\sum_{T\in Cay(n)}\frac{x_{T\cup\{i,n+1\}}}{x_{n+1}} = \sum_{i=1}^{n}\sum_{T\in Cay(n)} x_T x_i\\ &= (x_1+\cdots+x_n)P_n\,.\end{aligned}$$

[2]The next Lemma's proof follows the lines presented in the lecture notes of Haiman available online at `https://math.berkeley.edu/~mhaiman/math172-spring10/`

Now we prove that any sequence of polynomials for $n \geq 2$ satisfying all these three conditions is uniquely determined. We act inductively and since the base case $n = 2$ is obvious, let's suppose that $P_n = Q_n$ and show that $P_{n+1} = Q_{n+1}$.

Since, from item 2, each monomial is linear in at least one term, it is enough to show that the linear part with respect to each variable on both polynomials is the same. From item 3 it is enough to show that the linear part of P_{n+1} and Q_{n+1} are the same with respect to one variable only, by symmetry. We will show it for the variable x_{n+1} using item 4: indeed the coefficient in the linear term of x_{n+1} on the polynomial F is exactly $\frac{F}{x_{n+1}}\Big|_{x_{n+1}=0}$.

So we have to show that

$$\frac{P_{n+1}}{x_{n+1}}\Big|_{x_{n+1}=0} = \frac{Q_{n+1}}{x_{n+1}}\Big|_{x_{n+1}=0}.$$

Or, equivalently,

$$(x_1 + \cdots + x_n) \times P_n = (x_1 + \cdots + x_n) \times Q_n,$$

which holds by induction hypothesis. □

The events "$e \in T_n$", where e is an edge, are clearly not independent, because the requirement that T_n is a tree has to some extent some underlying dependencies. If A is a set of edges then we denote the event "$A \subseteq T_n$" as E_A for short, and $p_A := \mathbb{P}[E_A]$ will be determined in this section.

Recall that given an integer partition $\lambda = (\lambda_1, \lambda_2, \cdots)$, we write $p(\lambda) = \prod_i \lambda_i$.

Proposition 2.3.3. If A has no cycles and its connected components form an integer partition of type $\lambda(A) = \lambda$, then

$$p_A = \frac{p(\lambda) n^{l(\lambda)-2}}{n^{n-2}} = p(\lambda) n^{-n+l(\lambda)}.$$

Proof. Let N_A be the number of Cayley trees T such that $A \subseteq E(T)$. Then the desired probability is $p_A = \frac{N_A}{n^{n-2}}$. We will now compute the number N_A.

Let $\frac{V}{A} = \{C_1, \cdots, C_k\}$ be the set of connected components of the graph with edge set A. Note that $k = l(\lambda(A))$.

A tree T that contains the edge set A maps to a tree $f(T) \in Cay(V/A)$ by collapsing the connected components of A to a point. Our goal now is to compute $\# f^{-1}(T)$ in order to use

$$N_A = \sum_{T \in Cay(V/A)} \# f^{-1}(T).$$

Note that to map T to $f(T)$ we transform all edges that are not in A into edges between connected components in V/A. It is now easy to see that if T is a tree in $Cay(V/A)$, then

$\#f^{-1}(T) = \prod_{\{e_1,e_2\}\in E(T_n)} \#e_1\#e_2 = (\#C_1)^{\deg_T(C_1)}\cdots(\#C_k)^{\deg_T(C_k)}$ by choosing, for each edge in T, which vertices are its endpoints.

Now, according to Lemma 2.3.2, we have that:

$$\begin{aligned}
N_A &= \sum_{T\in Cay(V/A)} \#f^{-1}(T) \\
&= \sum_{T\in Cay(V/A)} (\#C_1)^{\deg_T(C_1)}\cdots(\#C_k)^{\deg_T(C_k)} \\
&= \left(\prod_i \#C_i\right)\left(\sum_i \#C_i\right)^{k-2} \\
&= p(\lambda(A))\,n^{l(\lambda(A))-2}.
\end{aligned}$$

□

Chapter 3

Coefficients of the chromatic symmetric function in Random Graphs

In this chapter we are going to address the threshold problem for subgraph containment and for big matchings, presented in [12]. Afterwards, we will apply these theorems in the realm of the chromatic symmetric function.

3.1 Subgraph containments

In this section, we will study the threshold of $X_\lambda = 0$, where $\lambda = 1^s\mu$ is taken for μ fixed and $s \to \infty$. Namely, from Theorem 1.2.6 we get that X_λ counts the number of stable partitions of the graph G of type λ which is easily seen as counting the number of subgraphs isomorphic to $\uplus K_{\mu_i}$ in G^c, where $\uplus K_{\mu_i}$ is the graph composed from disjoint copies of complete graphs with sizes μ_i.

For that, we will introduce in this section the theory developed in [12] regarding subgraph containment of $G(n,p)$.

3.1.1 Subgraph containment in $G(n,p)$

Let us introduce a descriptive variable for the density of a graph.

Definition 3.1.1. Given a graph G with n vertices and e edges, we write $d(G) = \frac{e}{n}$ for its *density*.

We write as well

$$m(G) = \max\{d(H) | H \subseteq G,\ v(H) > 0\}.$$

In this manner, we say that G is *balanced* if $d(G) = m(G)$, and *strictly balanced* if, for any $H \subseteq G$ such that $v(H) > 0$, we have

$$d(H) = m(G) \Rightarrow H = G.$$

Balanced graphs is a topic studied in the literature with some enthusiasm. It is important to mention that, in 1985 Györi, Rothschild and Rucinski proved that any graph G is a subgraph of a balanced graph F with $d(F) = m(G)$, in [11].

The next theorem is a fundamental theorem from [12, Chapter 3]. Let X_H be the number of copies of a graph H in $G(n,p)$.

Theorem 3.1.2. Suppose that $e(G) > 0$, then

$$\mathbb{P}[X_G = 0] \to \begin{cases} 1 & \text{if } p << n^{-\frac{1}{m(G)}}; \\ 0 & \text{if } p >> n^{-\frac{1}{m(G)}}. \end{cases}$$

Proof. Denote by I_F, for all subgraphs $F \subseteq K_n$, the event "F is a subgraph of $G(n,p)$". Then $X_H = \sum_{\substack{F\subseteq K_n \\ F\cong H}} I_F$, so

$$\mathbb{E}[X_H] = \frac{n!}{(n-v(H))!}\,\text{aut}\,(H)^{-1} p^{e(H)} = \Theta\left((np^{d(H)})^{v(H)}\right).$$

For the case that $p << n^{-\frac{1}{m(G)}}$, let H be the densest subgraph $H \subseteq G$, i.e. $d(H) = m(G)$. Then, since H is a subrgraph of G, $X_H = 0 \Rightarrow X_G = 0$. Our goal is to use the 1st moment method (Lemma 2.1.3) to show that $X_H = 0$ a.a.s. to conclude that $\mathbb{P}[X_G = 0] \to 1$. Incidentally, if we have $p << n^{-\frac{1}{m(G)}}$ then $\mathbb{E}[X_H] = \Theta\left((np^{d(H)})^{v(H)}\right) = o(1)$ so $\mathbb{P}[X_H = 0] \to 1$, concluding the 0-statement.

For the 1-statement, we compute the Variance of $X_G = \sum_{\substack{F\subseteq K_n \\ F\cong H}} I_F$. In the next sum, G_1, G_2 will range over all subgraphs of K_n isomorphic to G. Write $\text{aut}_H(G)$ for the number of automorphisms of the graph G that fix each vertex of H. We will use the fact that $1-p \le 1-p^k \le (1-p)k$ for k a positive integer. We consider the sum over $H \subseteq G$ running over all isomorphic classes of subgraphs of G, however this wont make a difference in the asymptotic sense. Then,

$$\begin{aligned} \text{Var}[X_G] &= \sum_{G_1,G_2} \text{Cov}[I_{G_1}, I_{G_2}] \\ &= \sum_{H\subseteq G} \sum_{G_1\cap G_2 = H} p^{2e(G)-e(H)} - p^{2e(G)} \\ &= \sum_{H\subseteq G} \frac{n!(p^{2e(G)-e(H)} - p^{2e(G)})}{(n-2v(G)-v(H))!\,\text{aut}_H(G)^2\,\text{aut}\,(H)}. \end{aligned}$$

So,

$$\begin{aligned}\operatorname{Var}[X_G] &\asymp \sum_{H\subseteq G} n^{2v(G)-v(H)} p^{2e(G)-e(H)}(1-p)\\ &\asymp (1-p)\max_{H\subseteq G}\frac{n^{2v(G)}p^{2e(G)}}{n^{v(H)}p^{e(H)}}\\ &\asymp (1-p)\mathbb{E}[X_G]^2\max_{H\subseteq G}\left(np^{d(H)}\right)^{-v(H)}.\end{aligned}$$

Now, if $p >> n^{-\frac{1}{m(G)}}$ then $np^{d(H)} >> 1$ for any subgraph $H \subseteq G$, so $\frac{\operatorname{Var}[X_G]^2}{\mathbb{E}[X_G]^2} = o(1)$ so, by the 2nd moment method (Lemma 2.1.6) we get $\mathbb{P}[X_G = 0] \to 0$, so there exists a copy of G in $G(n,p)$ with high probability. □

Example 3.1.3. Here we study the example of the threshold for the random graph model $G(n,p)$ to contain a triangle. A triangle is a graph K_3 with $m(K_3) = 1$, so we should expect, according to Theorem 3.1.2, that if $p = o\left(n^{-1}\right)$, no triangle arises.

Indeed, we can compute the expected number of triangles and apply the 1st moment method (Lemma 2.1.3). Let $I_{x,v,w}$ be the indicator variable that all edges between vertices x, v and w are in $G(n,p)$. Then $X_{K_3} = \sum_{x,v,w} I_{x,v,w}$ is the number of triangles in the graph, where the sum runs over all different sets $\{x, v, w\}$. By linearity of expectation, we have that

$$\mathbb{E}[X_{K_3}] = \binom{n}{3}\mathbb{P}[X_{x,v,w} = 1] = \Theta(n^3)p^3.$$

So, by the 1st moment method Lemma (Lemma 2.1.3), if $np = o(1)$ then, with high probability, there are no triangles.

For the other side of the threshold, let $p = \omega\left(n^{-1}\right)$. We compute the variance in order to use the 2nd moment method (Lemma 2.1.6). Indeed, let I_T be the indicator variable for the event "T is a subgraph of $G(n,p)$", then

$$\operatorname{Var}[X_{K_3}] = \sum_{T_1,T_2}\operatorname{Cov}[I_{T_1}, I_{T_2}] \tag{3.1}$$

where the sum runs over all pairs of triangles, T_1, T_2.

In proof of Theorem 3.1.2 it is advised to group the sum (3.1) into pairs of triplets that satisfy $T_1 \cap T_2 = H$, for each H that is a subgraph of the triangle. There are four non-isomorphic subgraphs of the triangle, each with zero, one, two and three edges. It is impossible for two triangles to intersect in exactly two edges, so we get,

$$
\begin{aligned}
\mathrm{Var}[X_{K_3}] &= \sum_{T_1,T_2} \mathrm{Cov}[I_{T_1}, I_{T_2}] \\
&= \sum_{k=0}^{3} \sum_{\#E(T_1\cap T_2)=k} \mathbb{E}[I_{T_1} I_{T_2}] - \mathbb{E}[I_{T_1}]\mathbb{E}[I_{T_2}] \\
&= \sum_{k=0}^{3} \sum_{\#E(T_1\cap T_2)=k} p^{6-k} - p^6 \\
&= \left(p^5 - p^6\right) 6 \binom{n}{4} + \left(p^3 - p^6\right) \binom{n}{3} \\
&\asymp (np)^4 \frac{p(1-p)}{4} + (np)^3 \frac{(1-p^3)}{6}.
\end{aligned}
\tag{3.2}
$$

Now, since $\mathbb{E}[X_{K_3}]^2 \sim (np)^6 \frac{1}{36}$ we get, for $np = \omega(1)$, that $\mathrm{Var}[X_{K_3}] = o\left(\mathbb{E}[X_{K_3}]^2\right)$, which concludes the indicated threshold.

Note that X_{K_3} is also the number of stable partitions of type $\lambda = 1^{n-3}3$ in the complement graph of $G(n,p)$, which follows the same model as $G(n, 1-p)$. Hence, in this example, we find that the threshold value for X_λ is $1 - \Theta\left(\frac{1}{n}\right)$.

According to Theorem 1.2.6, this translates quite easily to the coefficients of the monomial basis of $\chi_{G(n,p)}$, X'_λ, via $X'_\lambda = \text{ aut } \lambda X_\lambda$.

Proposition 3.1.4. Let X'_λ be the coefficient of the chromatic symmetric function of $G(n,p)$ in m_λ, over the monomial basis.

- If $p = 1 - \omega(n^{-1})$ then $\mathbb{P}[X'_\lambda = 0] \to 0$.
- If $p = 1 - o(n^{-1})$ then $\mathbb{P}[X'_\lambda = 0] \to 1$.

We now recover a theorem from [12] that describes what happens for some graphs inside the threshold given by Theorem 3.1.2.

Theorem 3.1.5. Suppose that G is a strictly balanced graph, and let $p = \Theta(n^{-\frac{1}{m(G)}})$ with

$$\lim p n^{\frac{1}{m(G)}} = c.$$

Then

$$X_G \sim \text{ Poi } \left(\frac{c^{v(G)}}{\text{aut } G}\right).$$

Proof. This theorem is shown in [12, Theorem 3.19]. $\square$

Example 3.1.6. Let us recover here Example 3.1.3.

Suppose that $p = \Theta\left(n^{-1}\right)$ and indeed $\lim pn = c$. Then, by Theorem 3.1.5, the number of triangles X_T follows a Poisson distribution with parameter $\frac{c^3}{6}$.

Consequently, as the variable $X_{1^{n-3}3}$ in $G(n,p)$ counts the number of stable sets of type $1^{n-3}3$, it counts triangles on $G(n,p)^c \sim G(n,1-p)$ so it follows a Poisson distribution with parameter $\frac{c^3}{6}$ in the threshold $p = 1-\Theta(n^{-1})$, where $c = \lim(1-p)n$. In particular, $\mathbb{P}[X_{1^{n-3}3} = 0] \to \exp(-c^3/6)$ so $X_{1^{n-3}3}$ has a coarse threshold, similarly for any variable of the type $X_{1^{n-k}k}$ with a constant k.

3.1.2 Thresholds for coefficients of $\chi_{G(n,p)}$

The first result that we have presented in this section deals with the subgraph containment, and in Examples 3.1.3 and 3.1.6 we have seen that threshold for the property "contain a triangle" actually relates to the threshold of X_λ for $\lambda = 1^{n-3}3$. We are going to see here this connection in more generality and its consequences.

Example 3.1.7. Consider the graph $T = K_3$ and let $T_{3,3}$ be the disjoint union of two triangles T. We want to count the number of subgraphs isomorphic to $T_{3,3}$ in $G(n,p)$.

The idea explained in [12, Example 3.21] was to consider first the number of triangles X_T in the graph $G(n,p)$, which has threshold value $pn = \Theta(1)$. If $\lim np = c$, then $X_T \sim Poi\left(\frac{c^3}{3!}\right)$, according to Theorem 3.1.5, because T is a strictly balanced graph.

On the other hand, $X_{T_{3,3}} = \binom{X_T}{2}$ a.a.s: an embedded $T_{3,3}$ is just two embedded triangles that are disjoint. However, the number of pairs of triangles that are not disjoint correspond to an embedded T_1 or T_2, from Figure 3.1. But for $p = \Theta(n^{-1})$, according to Theorem 3.1.2, there are no embedded T_1 or T_2 a.a.s., so all triangles are disjoint.

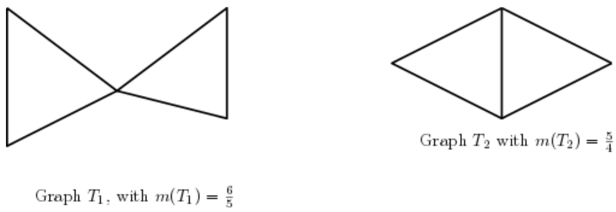

FIGURE 3.1: Different overlaps of two triangles

If $\mu = 3^2$ and $\lambda = 1^s\mu$ we get that the random variable X_λ in $G(n,p)$ has threshold value $p = \Theta(n^{-1})$ and if $\lim pn = c$,

$$X_\lambda \sim \binom{P}{2},$$

where $P \sim$ Poi $\left(\frac{c^3}{3!}\right)$. We also get that $\mathbb{P}[X_\lambda = 0] \to \exp(-c^3/6)$ and that the event "X_λ" also has a coarse threshold.

The example here discussed paves the way to a generalization that describes the variables X_λ for $\lambda = 1^s\mu$ with μ fixed.

Theorem 3.1.8. Let $\lambda = 1^s\mu$ be a partition and suppose that μ is fixed and $s \to \infty$. Suppose further that the biggest part of μ has size k.

Then X_λ satisfies the following:

- If $1-p = o\left(n^{-\frac{2}{k-1}}\right)$ then $\mathbb{P}[X_\lambda = 0] \to 1$.
- If $1-p = \omega\left(n^{-\frac{2}{k-1}}\right)$ then $\mathbb{P}[X_\lambda \le a] \to 0$ for any real a, i.e. $X_\lambda \to \infty$ a.a.s.

Write $\mu = k^s\mu'$, where μ' is a partition in which all parts are of size smaller than k. Additionally, if $(1-p)^{k-1}n^2 = \Theta(1)$ and $\lim(1-p)^{\frac{k-1}{2}}n = c$ then there are two cases:

- If μ' is non-empty, then $\mathbb{P}[X_\lambda \le a] \to \alpha$ for a suitable parameter α independent of $a \ge 0$ (i.e., the probability is concentrated at 0 and at values that approach ∞).

  Namely,
  $$\alpha = \mathbb{P}[X_\lambda = 0] = \sum_{x=0}^{s-1} e^{-c^k/k!}\frac{x^{c^k/k!}}{x!}.$$
- If $\mu = k^s$ then
  $$X_\lambda \sim \binom{P}{s},$$
  where $P \sim$ Poi $\left(\frac{c^k}{k!}\right)$.

Let's first introduce our notion of K_μ, and then prove Theorem 3.1.8.

Definition 3.1.9. Given a partition μ, where k is the biggest part of μ, consider the graph $K_\mu = \biguplus_i K_{\mu_i}$. Note that the number of automorphisms of this graph is $\lambda!$ aut (μ).

We have that, $m(K_\mu) = d(K_k) = \frac{\binom{k}{2}}{k} = \frac{k-1}{2}$, since the densest subgraph of K_μ is a k-complete graph.

Note that if $\mu = (m,n)$, this is dangerously similar to the notation for the complete bipartite graph $K_{m,n}$, when in fact we have $K_{(m,n)} = K^c_{m,n}$.

Proof of Theorem 3.1.8. Note that a stable set in $G(n,p)$ of type λ is a K_μ-subgraph in $G(n,p)^c$. So $X_\lambda = X_{K_\mu}$ where X_{K_μ} counts the number of K_μ-subgraphs in $G(n,p)^c \sim G(n,1-p)$. According to Theorem 3.1.2, if $1-p = o(n^{-\frac{2}{k-1}})$ then $X_{K_\mu} = 0$ a.a.s, whereas if $1-p = \omega(n^{-\frac{2}{k-1}})$ we get $X_{K_\mu} \to \infty$ a.a.s.

Consequently, the threshold value for X_λ is $1-\Theta(n^{-\frac{2}{k-1}})$, which concludes the first part.

Now suppose that $p = 1-\Theta(n^{-2/(k-1)})$ and that $\lim(1-p)^{\frac{k-1}{2}} n = c$. Then, from Theorem 3.1.5 and since K_k is a strictly balanced graph, we have that the number of K_k-subgraphs in $G(n,p)^c$ follows the distribution $X_{K_k} \sim \mathrm{Poi}(c^k/k!)$.

We now show that there are no two K_k-subgraphs in $G(n,p)^c$ that overlap a.a.s. Note that two K_k-subgraphs that overlap in t vertices such that $k > t > 0$ form a subgraph $G_t = K_k \cup K_k$ contained in $G(n,p)^c \sim G(n,1-p)$ with $v(G_t) = 2k-t$ and $e(G_t) = \binom{2k-t}{2} - (k-t)^2$. Now observe that the graph G_t is too dense to occur as a subgraph of $G(n,p)^c$: write $v(G_t) = l = 2k-t$ and note that

$$
\begin{aligned}
2d(G_t) &= \frac{l(l-1)-2(l-k)^2}{l} \\
&= l-1-2l+4k-2k^2/l \\
&= k-1-2k+t+3k-k\left(1+\frac{t}{l}\right) \\
&= k-1+t-kt/l \\
&> k-1\,,
\end{aligned}
$$

because $k < l$. Hence $m(G_t) \geq d(G_t) > \frac{k-1}{2}$.

Consequently $\mathbb{P}[\cup_t X_{G_t} > 0] \leq \sum_t \mathbb{P}[X_{G_t} > 0] = o(k)$, and since k is a constant, we get that no two embedded K_k intersect a.a.s.

This immediately concludes the case $\mu = k^s$ since a.a.s. $X_{K_\mu} \sim \binom{X_{K_k}}{s}$ and $X_{K_k} \sim \mathrm{Poi}(c^k/k!)$. Additionally, we have that

$$
\alpha = \mathbb{P}[X_\lambda = 0] = \mathbb{P}[X_{K_k} < s] = \sum_{x=0}^{s-1} e^{-c^k/k!}\frac{x^{c^k/k!}}{x!}\,.
$$

For the remaining case, if there is a part of μ of size smaller than k, let μ' be the partition resulting from μ by disregarding all parts of size k, so $\mu = k^s\mu'$.

Now $K_{\mu'}$ has density strictly smaller than $m(K_\mu) = \frac{k-1}{2}$, as it is easily observed, so there are arbitrarily many occurrences of $K_{\mu'}$ in $G(n,1-p)$. Additionally, if we divide the vertex set of $G(n,1-p)$ into $ks+1$ disjoint blocks $B_{1,n},\dots,B_{ks+1,n}$, all of these blocks will have arbitrarily many occurrences of $K_{\mu'}$ for $n \to \infty$.

Now, there are two cases: either there is no occurrence of s disjoint copies of K_k, which as we have seen occurs with probability

$$
\alpha = \sum_{x=0}^{s-1} e^{-c^k/k!}\frac{x^{c^k/k!}}{x!}\,,
$$

and implies that $X_{K_\mu} = 0$, or there are s disjoint occurrences of K_k, which means that there is one block $B_{i,n}$ that is disjoint from such s disjoint copies. Since in $B_{i,n}$ there are arbitrarily many copies of $K_{\mu'}$, there are arbitrarily many copies of K_μ in $G(n,1-p)$.

Hence, we conclude that $\mathbb{P}[X_{K_\mu} \leq a] \to \mathbb{P}[X_{K_\mu} = 0] = \alpha$, so according to the relation between the random variable X_{K_μ} in $G(n,1-p)$ and X_λ in $G(n,p)$ pointed our earlier, this concludes the proof. □

3.2 Big matchings

The goal of this section is to introduce some threshold studies for the case $\lambda = 1^s 2^r$ for $r = \Theta(s)$.

To prelude that, we introduce the problem of perfect matchings on $G(n,p)$, which is thoroughly studied in the literature and makes use of either Tutte's theorem or Hall's theorem.

In this work, we will use Hall's Theorem following [12], which we recall now:

Theorem 3.2.1 (Hall's Theorem)**.** Let $G = (A \uplus B, E)$ be a bipartite graph with $\#A = \#B$. Then, there is a perfect matching in G if and only if there is no set $S \subseteq A$ such that $\#N(S) < \#S$.

To a set $H \subseteq A$ such that $\#N(H) < \#H$ we call a *Hall set*. We say that H satisfies the *Hall property* if it's not a Hall set.

Proof. One of the implications is trivial: if there is a perfect matching, there is a bijective function $f : A \to B$ such that $\{a, f(a)\}$ is an edge. If $S \subseteq A$, then $f(S) \subseteq N(S)$ so $\#S = \#f(S) \leq \#N(S)$ as desired.

For the other implication, we refer to any Graph Theory introductory book, for instance [4] or [5], where the main idea is to greedily construct the Hall set H. □

Now take $G(n,n,p)$ and let's compute the probability that a perfect matching arises. For that, we compute the probability of a set S to arise such that $\#N(S) < \#S$.

Proposition 3.2.2. The probability $\mathbb{P}[G(n,n,p)$ has a perfect matching$]$

- converges to 0, if $np - \log n \to -\infty$.
- converges to $\exp(-2e^{-c})$ if $np - \log n \to c$.
- converges to 1, if $np - \log n \to \infty$.

We won't go over the case $np - \log n \to c$ here. As for the other cases, we will prove them after we settle some remarks.

In the proof of Proposition 3.2.2 we will check the existence of Hall sets. It's natural to narrow such search to the minimal Hall sets. Incidentally, we know that a minimal Hall set X:

- Satisfies $\#X = \#N(X) + 1$: Start from any Hall set and erase any vertex in X until we have $\#X = \#N(X) + 1$ preserving the Hall property.
- Is such that each vertex in $N(X)$ is adjacent to at least two vertices in X. We can obtain that if, for each $v \in N(X)$ connected only to $x \in X$, we erase x from X, preserving Hall's property because $v \notin N(X \setminus x)$.
- Satisfies $\#S \le \frac{\#A}{2}$. This is so because if $S \subseteq A$ is a Hall set with $\#S > \frac{\#A}{2}$ then $N(B \setminus N(S)) \subseteq A \setminus S$, so $B \setminus N(S)$ is a smaller Hall set.

With this, if we show that there is no Hall set satisfying any of the above properties, from Hall's theorem there exists a perfects matching. We will deal with the Hall sets of size at least three, the Hall sets of size two (called *cherries*) and the Hall sets of size one (the isolated vertices) separately. Here we remark already for the study of cherries.

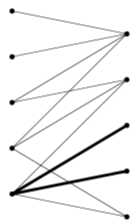

FIGURE 3.2: The fat lines are a cherry in a bipartite graph.

Remark 3.2.3 (Cherries count)**.** If $\#S = 2$, then S is called a *cherry*, as an example we have Figure 3.2. Let $Ch(G)$ be the number of cherries in the graph G, then

$$\mathbb{E}[Ch(G(n,n,p))] = 2\binom{n}{2} np^2(1-p)^{2n-2} \le n^3 p^2 \exp(-2(n-1)p)\,.$$

In the interval $0.5 \log n + \log\log n + \omega(1) \le np \le 2 \log n$, to count the number of cherries we have:

$$\begin{aligned}
\mathbb{E}[Ch(G(n,n,p))] &= n\Theta\left((\log n)^2\right)\exp(-2np) \\
&\le \Theta(1)\, n(\log n)^2 \exp(-2(\log n + \log\log n) - \omega(1)) \qquad (3.3) \\
&= o(1)\,.
\end{aligned}$$

Remark 3.2.4. If $\#S = 1$, then S consists of an isolated vertex. Let $IsoV(G)$ be the number of isolated vertices on G. Note that we have proved in Examples 2.1.4 and 2.1.7 that for $G = G(n,p)$ and $p = n^{-1}\log n + \omega(n^{-1})$, then $IsoV(G) = 0$ a.a.s., whereas if $p = n^{-1}\log n - \omega(n^{-1})$ then $IsoV(G) > 0$ a.a.s.

Consequently, a perfect matching in $G(n,p)$ can't occur for $p = n^{-1}\log n - \omega(n^{-1})$.

In fact, for $G(n,n,p)$ the case isn't much different, as for $np = \log n + \omega(1)$ we have:

$$amthbbE[IsoV(G(n,n,p))] = 2n(1-p)^{n-1} \leq 2\exp(\log n - pn + p)$$
$$= 2\exp -\omega(1) = o(1)\,.$$

So by the 1st moment method (Lemma 2.1.3) there are a.a.s. no isolated vertices for $np = \log n + \omega(1)$. We will see latter on that for $np = \log n - \omega(1)$, there are isolated vertices a.a.s.

In the course of the proof of Proposition 3.2.2 we will see that $IsoV(G(n,n,p))$ is non-zero with high probability when $np = \log n - \omega(1)$, unlike $Ch(G(n,n,p))$. We will see in fact that no other Hall sets occur at $np \geq 0.7\log n$, which means that the existence of isolated vertices is the ultimate hurdle for the existence of a perfect matching in $G(n,n,p)$, according to Hall's Theorem 3.2.1.

Proof Of Proposition 3.2.2. The way we are going to argue is the following: We show that no Hall set of size at least three occurs when $0.7\log n \leq np \leq 2\log n$. With that and with what we have seen in Remark 3.2.3, there are no cherries for $np = 0.5\log n + \log\log n + \omega(1)$, so a perfect matching must exist when all vertices become non-isolated, according to Hall's Theorem 3.2.1 if $np \geq 0.7\log n$. Recall that we have seen in Remark 3.2.4, with the 1st moment method (Lemma 2.1.3) that no isolated vertices exists for $np = \log n + \omega(1)$. Hence, once we use the 2nd moment method (Lemma 2.1.6) to show that there are some isolated vertices for $p = \log n - \omega(1)$, we conclude that a perfect matching arises exactly when the last isolated vertex gets a neighbour, and always after $np - \log n = -\omega(1)$.

Let $\mathscr{A}$ be the event that some Hall set occurs, and $\mathbb{A}_{X,N}$ the event that X is a Hall set with $N(X) = N$. Conditioning to the minimal Hall sets and bounding the probability that X is a Hall set of size s with a given $N(X)$ by only choosing two neighbours for each $v \in X$, we get that $\mathbb{P}[\mathbb{A}_{X,N}] \leq \binom{s}{2}^{s-1} p^{2s-2}(1-p)^{s(n-s+1)}$.

Recall the bound $\binom{n}{s} \leq \frac{n^s}{s!} \leq \left(\frac{en}{s}\right)^s$ from Proposition A.2.3. Recall also that we are assuming $0.7\log n \leq np \leq 2\log n$.

Recall as well the identity $\sum_{s\geq 3} x^s/s \leq \sum_{s\geq 3} x^s = \frac{x^3}{1-x}$ and call $L = \frac{e^2 8(\log n)^2}{n^{1.05}} \to 0$. Also note that for $s \leq n/2$ we have $n - s + 1 > n/2$. So, a union bound argument tells us:

$$\begin{aligned}
\mathbb{P}[\mathscr{A}] &\leq \sum_{s=3}^{n/2} 2\binom{n}{s}\binom{n}{s-1}\binom{s}{2}^{s-1} p^{2s-2}(1-p)^{s(n-s+1)} \\
&\leq \sum_{s=3}^{n/2} 2\left(\frac{en}{s}\right)^{s}\left(\frac{en}{s-1}\right)^{s-1} s^{2(s-1)} \times \left(\frac{2\log n}{n}\right)^{2s-2}\left(1-\frac{0.7\log n}{n}\right)^{s\frac{n}{2}} \\
&\leq \sum_{s=3} e^{2s-1}2^{2s-1}\left(\frac{s}{s-1}\right)^{s-1} sn(\log n)^{2s-2}\exp(-0.7\log n)^{s/2} \\
&\leq \sum_{s=3} e^{2s-1}2^{2s-1}2^{s-1}s^{-1}n(\log n)^{2s-2}n^{-0.35s} \\
&= \frac{n}{4e(\log n)^2}\sum_{s=3}^{n/2} s^{-1}\left(\frac{e^2 2^2 2(\log n)^2}{n^{0.35}}\right)^{s} \\
&= \frac{n}{4e(\log n)^2}\sum_{s=3}^{n/2} s^{-1}L^s \\
&\leq \frac{n}{4e(\log n)^2}\left(\frac{L^3}{1-L}\right) \\
&= O\left(\frac{n(\log n)^6}{(\log n)^2 n^{3\times 0.35}}\right) = O\left(\frac{(\log n)^4}{n^{0.05}}\right).
\end{aligned} \tag{3.4}$$

Concluding that, for $2\log n \geq np \geq 0.7\log n$ there are no cherries or Hall sets of size at least 3. We know as well that for $np = \log n + \omega(1)$ there are no isolated vertices according to Remark 3.2.4, concluding that if $\log n + \omega(1) = np \leq 2\log n$ then there is a perfect matching with probability $1 - o(1)$. Note that the property "Graph G has a perfect matching" is a monotone one, so we conclude that for any $p \geq n^{-1}\log n + \omega(n^{-1})$, we have

$$\mathbb{P}[\text{ "Graph } G(n,p) \text{ has a perfect matching"}] \to 1\,.$$

Now we use the 2nd moment method (Lemma 2.1.6) to show that for $np - \log n \to -\infty$, there are isolated vertices. Take $np = \log n - c$ for $c = \omega(1)$ but naturally $c = o(\log n)$, and recall some useful inequalities in Proposition A.2.3 as well as in Proposition A.2.2. Then,

$$\begin{aligned}
\mathbb{E}[Iso(G(n,n,p))] = \sum_v \mathbb{E}[I_v] = 2n(1-p)^n &\geq 2n\left(e^{-p} - p^2/2\right)^n \\
&= 2ne^{-np} - n^2p^2e^{-(n-1)p} + \sum_{k=2}^{n} 2n\binom{n}{k}\frac{(-p)^{2k}}{2^k}e^{-(n-k)p} \\
&\geq 2e^c\log n - (\log n)^2 e^p n^{-1}e^c - \sum_{k=2}^{n} 2n\left(\frac{ne}{k}\right)^k \frac{p^{2k}e^{kp}}{2^k} n^{-1}e^c\log n \\
&\geq 2\log n - o(1) - 2e^c\log n\sum_k\left(\frac{enp^2e^p}{2k}\right)^k.
\end{aligned}$$

Now since $np^2 = (\log n - c)p \le (\log n)^2/n$ we have that $\frac{e^{p+1}np^2}{2k} \le \frac{e^{p+1}(\log n)^2}{2n}$ and so,

$$\begin{aligned}
\mathbb{E}[Iso(G(n,n,p))] &\ge 2e^c \log n - o(1) - 2e^c \log n \sum_{k=2}^{n} \left(\frac{e^{p+1}(\log n)^2}{2n}\right)^k \\
&= 2e^c \log n - o(1) - 2e^c \log n \left(\frac{e^{p+1}(\log n)^2}{2n}\right)^2 \frac{1}{1-\frac{e^{p+1}(\log n)^2}{2n}} \\
&= 2e^c \log n - o(1) - e^c (\log n)^5 n^{-2} \Theta(1) \to \infty.
\end{aligned}$$

So $\mathbb{E}[Iso(G(n,n,p))] \to \infty$.

Let I_v be the indicator variable for the event "The vertex v is isolated". If $v \neq w$ are in the same side of the bipartition, I_v, I_w are independent. So we expect to cut down a great bunch of the computation of $\mathrm{Var}[X]$ for $p = o(1)$:

$$\begin{aligned}
\mathrm{Var}[Iso(G(n,n,p))] &= \sum_v \mathrm{Var}[I_v] + 2 \sum_{v \in A} \sum_{w \in B} \mathrm{Cov}[I_v, I_w] \\
&= 2n((1-p)^n - (1-p)^{2n}) + 2n^2\left((1-p)^{2n-1} - (1-p)^{2n}\right) \\
&= 2n(1-p)^n(1-(1-p)^n) + 2n^2(1-p)^{2n}\frac{p}{1-p} \\
&= \mathbb{E}[X]O(1) + o(\mathbb{E}[X]^2) = o\left(\mathbb{E}[X]^2\right).
\end{aligned} \tag{3.5}$$

So when $np = \log n - \omega(1)$, from the 2nd moment method (Lemma 2.1.6) we get that $\mathbb{P}[X = 0] \to 0$, concluding that a perfect matching does not exist with high probability.

The claim that $\mathbb{P}[IsoV(G(n,n,p)) = 0] \to \exp(-2e^{-c})$, when $np - \log n \to c \in \mathbb{R}$ uses the Stein Chen Method [Chapter 6 12, Theorem 6.24], however, we don't need it for the remaining of this work. Such method yields, additionally, that $X \sim Poi(\lambda)$ where $\lambda = 2e^{-c}$. □

Since we have seen that the first perfect matching appears exactly at the same time as the last isolated vertex gets connected, and according to the duality of the models $G(n,n,p)$ and $G(n,n,M)$ we get that:

Corollary 3.2.5. In the random process $\{G(n,n,M)\}_{M=0}^{n^2}$, the hitting time of the increasing property "There is a perfect matching" is the same as the hitting time of the increasing property "All vertices have an incident edge".

Now we will focus on passing the following reasoning to the realm of $G(n,p)$.

The obvious way to translate Corollary 3.2.5 and Proposition 3.2.2 to the $G(n,p)$ realm is to fix a bipartition of type $(n/2, n/2)$ of the vertices and apply Proposition 3.2.2. So we know that if $pn/2 - \log(n/2) = \omega(1)$ then a matching to occurs in $G(n,p)$, or equivalently, $np = 2\log n + \omega(1)$.

Now, the threshold here obtained for the existence of a perfect matching doesn't match the one for the number of isolated vertices in Example 2.1.7, so we should be able to improve this result. In fact, we will be able to see a similar theorem from Corollary 3.2.5 for the $G(n,M)$ model, where the hitting time for a perfect matching is exactly the same as the hitting time of the last isolated vertex to get connected, a result due to Bollobás and Thomason (1985) in [3].

We are going to introduce here an improved result, by finding a matching on the non-isolated vertices of $G(n,p)$: since it is clear that the existence of isolated vertices is the ultimate hurdle to obtain a perfect matching in $G(n,n,p)$, it is intuitive that we are able find one perfect matching in the non-isolated vertices of $G(n,p)$ once the threshold for existence of cherries has been passed.

We will, then, put that intuition into a demonstration following [12].

Definition 3.2.6 (PM property)**.** A graph G is said to have the *PM property* if there is a matching covering all but at most one non-isolated vertex.

Note that the expected number of cherries $Ch(G(n,p))$ in our graph is, for $np = 0.5\log n + \log\log n + \omega(1)$

$$\begin{aligned}
\mathbb{E}[Ch(G(n,p))] = &\sum_{x,v,w} \mathbb{P}[x,v,w \text{ forms a cherry }]\\
=&n\binom{n-1}{2}p^2(1-p)^{2(n-3)+1}\\
\leq&n^3p^2\exp(-2pn+5p)\\
=&\left(\frac{1}{2}\log n + \log\log n + \omega(1)\right)^2 n\exp(-\log n - 2\log\log n - \omega(1))\\
=&(\log n)^2(\log n)^{-2}\exp(-\omega(1))\\
=&o(1)\,.
\end{aligned}$$

From the bipartite case we get the intuition that once there are no cherries left, we have the PM property. This means that once we have disregarded the isolated vertices, the threshold to guarantee a matching is

$$np = \frac{1}{2}\log n + \log\log n + \omega(1)\,.$$

Our task in the following is to make this intuition clear.

Lemma 3.2.7. Let $np = \Theta(\log n)$ and $c > 0$.

Then, a.a.s. every bipartite subgraph induced in $G(n,p)$ by two sets of the same size with minimum degree $c\log n$ has a perfect matching.

Proof. Let $u = \frac{n(\log\log n)^2}{\log n}$. Most of the work has been already done, namely recall that in Example 2.1.5 and we obtained that every pair of disjoint subsets of vertices of size bigger than u has an edge in $G(n,p)$.

In [12] it is also mentioned that every set S of size at most $2u$ has fewer than $\#S(\log\log n)^3$ edges in $G(n,p)$ with both endpoints in S.

With that the proof becomes non-probabilistic. Take a bipartite subgraph $B \subseteq G(n,p)$ with parts of equal size and suppose it has no perfect matching, despite all vertices $v \in B$ having degree $\deg_B v \geq c\log n$. Write $B = W_1 \uplus W_2$ and $w = \#W_1 = \#W_2$. Then, according to Hall's Theorem 3.2.1 there is a Hall set $X \subseteq W_1$ such that $\#N_B(X) < \#X$.

There are three cases:

- Case 1: $\#X \leq u$.

  Then $\#(X \cup N_B(X)) < 2u$, but the number of edges generated by $X \cup N_B(X)$ is, at least

  $$\sum_{v\in X} \deg_B v \geq \#X c\log n \geq \#(X \cup N_B(X))\frac{c}{2}\log n,$$

  which is a contradiction with the fact that any subgraph S of at most $2u$ edges has fewer than $\#S(\log\log n)^3$ edges with both endpoints in S.

- Case 2: $\#N_B(X) \geq w - u$.

  Then we can apply Case 1 to $W_2 \setminus N_B(X)$.

- Case 3: $\#X > u$ and $\#NB(X) < w - u$.

  Then $\#(W_2 \setminus N_B(X)) > u$ and there is no edge between X and $W_2 \setminus N_B(X)$, a contradiction with the fact that any pair of disjoint sets of size bigger than u has an edge connecting them. □

Theorem 3.2.8. In the $G(n,p)$ model:

- If $np - 0.5\log n - \log\log n \to -\infty$, then $\mathbb{P}[G(n,p)$ has PM property $] \to 0$.
- If $np - 0.5\log n - \log\log n \to \infty$, then $\mathbb{P}[G(n,p)$ has PM property $] \to 1$.

Additionally, we have that the following hitting times coincide.

Corollary 3.2.9. Consider the random process $\{G(n,M)\}_{M=0,\cdots,\binom{n}{2}}$ where $G(n,M) \subseteq G(n,M+1)$, an let

$$\tau_{pm} = \min\{M|G(n,M) \text{ has a perfect matching }\},$$

and

$$\tau_{iv} = \min\{M|\forall v \in V(G(n,M)), \deg(v) > 0\},$$

which represent the hitting times of the respective properties.

Then, we have that $\tau_{pm} = \tau_{iv}$ a.a.s with $n \to \infty$.

Here we will just present a sketch of the proof, which can be found at [12, Chapter 4]. The idea is the following: We will split the set of vertices into a bipartition $\mathscr{B}$ with sets of equal size (± 1 if needed) uniformly at random and show that a.a.s. $G(n,p)$ with such bipartition satisfies some useful properties, for instance that there are only a few vertices that don't satisfy the precondition on Lemma 3.2.7. With that bipartition, our goal is to apply Lemma 3.2.7 to find a perfect matching. For that we go through a first fase of matching where we will find a matching for the "bad vertices" that don't satisfy the degree condition $\deg(v) \geq c \log n$, and then go to a second matching fase where we find a matching for the remaining vertices by finally applying Lemma 3.2.7.

We will now deal with the technicalities regarding the "bad vertices" with small degree.

Definition 3.2.10 (*Bad* and *small* vertices)**.** A *bad* vertex, for a given bipartition $\mathbb{P}$, is a vertex v that has less than $\log n/200$ neighbours in one of the sides of the bipartition $\mathscr{B}$.

A *small* vertex is a vertex that has degree smaller than four.

Lemma 3.2.11. Fix an integer k and a bipartition $\mathscr{B}$. Then, in $G(n,p)$ there is a.a.s. no k-vertex tree with more than two small vertices or more than four bad vertices.

Proof. Can be found in [12, Chapter 4]. □

Definition 3.2.12 (*Common* graphs)**.** A common graph is a graph G with n vertices and a fixed bipartition $\mathscr{B}$ on the vertices, that satisfies the following:

- Has no cherries.
- The graph satisfies Lemma 3.2.7 for $c = \frac{1}{300}$, i.e. every bipartite subgraph induced in G by two sets of equal size and with minimal degree at least $\log n/300$ has a perfect matching.
- There are fewer than $n^{0.98}$ bad vertices.
- For the fixed bipartition $\mathscr{B}$, it satisfies Lemma 3.2.11 for $k \leq 11$.

- The graph has maximum degree $8\log n$.
- The difference in non-isolated vertices between the sides is at most $n^{0.75}$.

Note we have already shown that most of this properties hold a.a.s., we will now see the rest of them here, to conclude the following Lemma:

Lemma 3.2.13. $G(n,p)$ is a.a.s. a common graph, for p such that $\frac{\log n}{2n} \le p \le \frac{2\log n}{n}$ and for a bipartition $\mathscr{B}$ taken uniformly at random among the bipartitions of $V(G(n,p))$ of sets of equal size (± 1 if needed).

Proof. We already know that a.a.s. a graph G satisfies, for the fixed bipartition $\mathbb{P}$, Lemma 3.2.11 for $k \le 11$.

From Equation (3.3) we know that a.a.s. there are no cherries in $G(n,p)$.

From Lemma 3.2.7 we know as well that every bipartite subgraph induced in G by two sets of equal size and with minimal degree at least $\log n/300$ has a perfect matching.

In Example 2.1.9 we have seen that the maximum degree is smaller than $8\log n$ a.a.s.

Let C_1, C_2 be the number of non-isolated vertices on each side, and note that $C_i \sim \mathrm{Bin}(n/2, p)$. A simple variation computation shows that C_1 and C_2 are concentrated around their mean by a margin of $O(\sqrt{n\log n})$. So, Markov's inequality (Lemma 2.1.1) tells us that the difference is smaller than $n^{0.75}$ a.a.s.

So we only have to show here that there are fewer than $n^{0.98}$ bad vertices.

Now to show that there are at most $n^{0.98}$ bad vertices, let $Y = \sum_v J_v$ be the number of bad vertices, and let J_v be the indicator variable for "v is a bad vertex ". For a vertex v, let C_v, D_v be the number of neighbours on each side of $\mathscr{B}$ and both follow a distribution stochastically smaller than $\mathrm{Bin}(n/2+1, p)$.

Recall that $\frac{\log n}{2n} \le np \le \frac{2\log n}{n}$. So we have that $\mathbb{E}[C_v] = 0.5np + p$ and so $\frac{\log n}{4} \le \mathbb{E}[C_v] \le \log n + p$. Then, from Chernoff bounds (Lemma 2.1.8) and a union bound argument we have,

$$\begin{aligned}
\mathbb{P}[v \text{ is a bad vertex }] &\le 2\mathbb{P}[C_v < \log n/200] \\
&\le 2\mathbb{P}\left[C_v < \mathbb{E}[C_v] - \log n\left(\frac{1}{4} - \frac{1}{200}\right)\right] \\
&\le 2\exp\left(-\frac{\left(\frac{49}{200}\log n\right)^2}{2\mathbb{E}[C_v]}\right) \\
&\le 2\exp\left(-\log n \frac{0.06\log n}{\log n + p/2}\right) = \Theta(n^{-0.03}).
\end{aligned}$$

So $\mathbb{E}[Y] = n\mathbb{P}[v \text{ is a bad vertex }] \leq \Theta(n^{0.97})$. Hence, from Markov's inequality 2.1.1,

$$\mathbb{P}[Y \geq n^{0.97}] \leq \frac{\mathbb{E}[Y]}{n^{0.98}} = O(n^{-0.01}) = o(1)\,.$$

Which concludes that a.a.s. there are not more than $n^{0.98}$ bad vertices. □

Now that we have shown that a.a.s. $G(n,p)$ is a common graph, we will find a way to construct a matching in common graphs that covers all but the isolated vertices in $G(n,p)$. This concludes that $G(n,p)$ has PM property in the conditions for Theorem 3.2.8.

Description of the algorithm to show Theorem 3.2.8. We will describe her the algorithm to find a matching that covers all but at most one non-isolated vertex in a common graph G.

Number all non-isolated bad vertices in increasing degree order, i.e. $v_1, v_2, \dots$ such that $\deg(v_1) \leq \deg(v_2) \leq \cdots$. We will find a matching iteratively for v_1, v_2, ..., until all bad vertices are matched.

First we match all v_i of degree one, which is possible because there are no cherries. Suppose now that $u_1, u_2, \dots, u_l$ are matched with $v_1, v_2, \dots, v_l$ respectively.

If v_{l+1} is already matched, i.e. $v_{l+1} = u_i$ for some i, then we add nothing and set $u_{l+1} = v_i$. If v_{l+1} has not yet been matched we choose u_{l+1} as follows:

- If $\deg(v_i) \in \{2,3\}$ then at most one of its neighbours is matched, as otherwise we would have a small tree with at least three small edges of size at most five, which is impossible because the graph G is common.
- If $\deg(v_i) \geq 4$ then there are at most three matched neighbours, for otherwise there would be a small tree with at most eleven vertices and at least five bad vertices, which is impossible again.

This concludes the matching of all bad vertices. Note that for any remaining vertex v, a big bunch of its neighbours is still to be matched. In fact, if for sake of contradiction, nine neighbours of v are already matched, then there are at least five different matchings neighbouring v, so there are at least five bad vertices in a tree containing v of size at most eleven.

So any remaining vertex v should have at most eight neighbours removed.

However, the remaining bipartition may not have the same number of non-isolated vertices in each side. Now there are at most $2n^{0.98}$ vertices removed, plus the difference in non-isolated vertices on the bipartition $\mathscr{B}$ to begin with, which is $O(n^{0.98})$. So, we have to move a vertex set $\mathscr{V}$ of size at most $3n^{0.98}$ from one side to the other, preserving the condition in Lemma 3.2.7 that the degree in each vertex in the induced bipartite graph is at least $\log n/300$. Note that we relaxed the degree requirement from Lemma 3.2.7, from $c = 200$ to $c = 300$, so we can account for the lost neighbours.

To choose $\mathscr{V}$ we simply choose a 2-independent set, i.e. a set of vertices in which no two share a neighbour. It's easy to show that there is such a set of size $\frac{n/2}{1+\Delta^2} >> n^{0.98}$, where $\Delta \leq 8\log n$ is the biggest degree.

Hence, we can choose a 2-independent set of vertices to change the side of bipartition $\mathscr{B}$ so that we get an even bipartition $\mathscr{B}'$ on the non-isolated vertices. All degrees are changed by at most eight, as we have noted before, so if we denote $\deg_{\mathscr{B}'}(v)$ by the minimum of the number of neighbours of v on each side of $\mathscr{B}'$, each vertex satisfies

$$\deg_{\mathscr{B}'}(v) \geq \log n/200 - 8 \geq \log n/300\,,$$

for n big enough.

With this and Lemma 3.2.7 we conclude the proof of Theorem 3.2.8 □

So, we have that once there are no isolated vertices, there is a perfect matching in $G(n,p)$ and the thresholds coincide. From Example 2.1.7 the threshold is exactly $p = \frac{\log n}{n} + \Theta(n^{-1})$:

Corollary 3.2.14. Let n run over only even numbers, and let $\mathscr{P}$ be the probability that $G(n,p)$ has a perfect matching. Then,

- $\mathscr{P} \to 0$, if $np - \log n \to -\infty$.
- $\mathscr{P} \to 1$, if $np - \log n \to \infty$.

3.2.1 Thresholds with matchings

In this section we will show two results that use the 1st moment method (Lemma 2.1.3) to establish lower bounds on the thresholds of the coefficients of chromatic symmetric functions. We will also see a conjecture on the upper bound of the threshold and discuss some applications of the theorems indicated in the previous section.

Theorem 3.2.15. Let $\lambda = 1^s 2^r$ be a partition of n. Suppose that $r = \Theta(s)$ and assume that $\alpha = \lim \frac{r}{s}$ exists.

Write for simplicity $C = \frac{2e}{(2+\alpha^{-1})(2\alpha+1)^{\alpha^{-1}}}$, $D = (2+\alpha^{-1})^{-1}$ and $P = \sqrt{2\pi(2+\alpha^{-1})(2\alpha+1)}$, which will play the role of constants. Set also $p_0 = 1-p$.

Then,

$$\mathbb{P}[X_\lambda = 0] \to 1 \text{ if } p_0 = \frac{C}{n} + \frac{C\log n}{2Dn^2} - \omega\left(\frac{1}{n^2}\right).$$

The upper part of the threshold is still a conjecture.

Conjecture 3.2.16. For suitable choice of α we have as well that

$$\mathbb{P}[X_\lambda = 0] \to 0 \text{ if } p_0 = \frac{C}{n} + \frac{C\log n}{2Dn^2} + \omega\left(\frac{1}{n^2}\right).$$

Proof of Theorem 3.2.15. We use 1st moment method (Lemma 2.1.3) to get this theorem.

First, recall that we write $\binom{n}{\lambda} = \frac{n!}{\lambda_1!\cdots\lambda_{l(\lambda)}!} = \frac{n!}{\lambda!}$. We recall also Theorem 1.2.16 to get that $\mathbb{E}[X_\lambda] = \frac{1}{\operatorname{aut}\lambda}\binom{n}{\lambda}(1-p)^r = \frac{n!p_0^r}{s!r!2^r}$.

Using Stirling's approximation A.2.1, recall that $n = 2r + s$ and $(\alpha + o(1))s = r$ to get

$$\begin{aligned}
\mathbb{E}[X_\lambda] &= \frac{1}{r!s!}\frac{n!}{2^r}p_0^r \\
&= \frac{e^s e^r n^n}{e^n s^s r^r}\frac{\sqrt{2\pi n}}{2\pi\sqrt{sr}}\frac{p_0^r}{2^r}(1+o(1)) \\
&= \left(\frac{p_0 n}{2e}\right)^r\left(\frac{n}{r}\right)^r\left(\frac{n}{s}\right)^s\sqrt{\frac{n}{2\pi rs}}(1+o(1)) \\
&= \left(\frac{p_0 n}{2e}\right)^r\left(2+\alpha^{-1}\right)^r(2\alpha+1)^{r\alpha^{-1}}\frac{1}{\sqrt{n}}\frac{1}{\sqrt{2\pi(2+\alpha^{-1})(2\alpha+1)}}(1+o(1)) \\
&= \left(p_0 n C^{-1}\right)^{nD}\frac{1}{P\sqrt{n}}(1+o(1)).
\end{aligned} \tag{3.6}$$

Now, for $p_0 = \frac{C}{n} + \frac{C\log n}{2Dn^2} - \omega\left(\frac{1}{n^2}\right)$ we get for n big and using Proposition A.2.2, that

$$\begin{aligned}
\mathbb{E}[X_\lambda] &= \left(p_0 n C^{-1}\right)^{nD}\frac{1}{P\sqrt{n}}(1+o(1)) \\
&= \left[1 + \frac{1}{n}\left(\frac{\log n}{2D} - \omega(1)\right)\right]^{nD}\frac{1}{P\sqrt{n}}(1+o(1)) \\
&\le \exp\left(\frac{\log n}{2D} - \omega(1)\right)^D\frac{1}{P\sqrt{n}}(1+o(1)) \\
&= \frac{\exp\left(-D\omega(1)\right)}{P}(1+o(1)) \\
&= o(1).
\end{aligned}$$

So, by the 1st moment method (Lemma 2.1.3), we get that $P[X_\lambda = 0] = 1 - o(1)$, concluding the theorem. □

For the upper threshold, the 2nd moment method (Lemma 2.1.6) is a good approach by only computing the variance of the variable X_λ. It is a good approach because the only inequality, from Proposition A.2.2, is very tight, so it is easy to change the reasoning above to show that if $p_0 = \frac{C}{n} + \frac{C\log n}{2Dn^2} + \omega\left(\frac{1}{n^2}\right)$ then $\mathbb{E}[X_\lambda] = \omega(1)$, as we have seen, for instance, in the proof of Proposition 3.2.2.

For that we separate the sum exactly the same way as in the proof of Theorem 3.1.2. It is easy to enumerate subgraphs of matchings: it's simply the matchings with k edges, where $k \le r$.

$$\begin{aligned}
\mathrm{Var}[X_\lambda] &= \sum_{M,N} \mathrm{Cov}[X_M, X_N] \\
&= \sum_{k=0}^{r} \sum_{\#E(M)\cap E(N)=k} \mathbb{E}[X_M X_N] - \mathbb{E}[X_M]\mathbb{E}[X_N] \\
&\le \sum_{k=0}^{r} \binom{\binom{n}{2}}{k, r-k, r-k, \binom{n}{2}-2r+k} \left(p_0^{2r-k} - p_0^{2r}\right),
\end{aligned}$$

where the sum ranges over matchings of r independent edges in the vertex set of $G(n,p)$.

$$\begin{aligned}
\mathrm{Var}[X_\lambda] &\le \sum_{k=0}^{r} \binom{\binom{n}{2}}{k, r-k, r-k, \binom{n}{2}-2r+k} \left(p_0^{2r-k} - p_0^{2r}\right) \\
&\le p_0^{2r} \sum_{k=0}^{r} \frac{\binom{n}{2}^{2r-k}}{k!(r-k)!^2} \left(p_0^{-k} - 1\right) \\
&\asymp p_0^{2r} \sum_{k=1}^{r} \frac{n^{4r-2k}(n^k-1)2^{-2r+k}}{k!(r-k)!^2} \\
&= (p_0 n)^{2r} \sum_{k=1}^{r} \frac{(n/2)^{r-k}}{(r-k)!} \frac{(n/2)^{r-k}}{(r-k)!} \frac{(n/2)^{k}}{k!} \\
&\le (p_0 n)^{2r} \left(\sum_{k=1}^{r} \frac{(n/2)^{r-k}}{(r-k)!}\right)\left(\sum_{k=1}^{r} \frac{(n/2)^{r-k}}{(r-k)!}\right)\left(\sum_{k=1}^{r} \frac{(n/2)^{k}}{k!}\right) \\
&\le (p_0 n)^{2r} \exp(3n/2) = (p_0 n)^{2r} \exp(1.5(2+\alpha^{-1}))^r .
\end{aligned} \tag{3.7}$$

Now recall that $\mathbb{E}[X_\lambda]^2 = \left(p_0 n C^{-1}\right)^{2r} \frac{1}{P^2 n}$, according to (3.6). Hence, we get that $\mathrm{Var}[X_\lambda] = o(\mathbb{E}[X_\lambda]^2)$ when $C^{-2} > \exp(1.5(2+\alpha^{-1}))$.

Unfortunately, the expression $\exp(1.5(2+\alpha^{-1}))$ grows much larger than C^{-2}, possibly because the inequalities used in (3.7) are very rough.

With this same reasoning we can indeed find more 1-statements for different rates of growth of r, for instance the following proposition:

Proposition 3.2.17. Let $\lambda = 1^s 2^r$ be an integer partition of $n = s+2r$. Suppose that $r = \Theta(\sqrt{n})$ and assume that $\alpha = \lim \frac{r}{\sqrt{n}}$ exists.

Call also $p_0 = 1-p$.

Then, if $p_0 = \frac{2\alpha}{e} n^{-3/2} + \frac{n^{-2}\log n}{2e} - \omega(n^{-2})$, we have

$$\mathbb{P}[X_\lambda = 0] \to 1\,.$$

Proof. We know from Theorem 1.2.16 that $\mathbb{E}[X_\lambda] = \frac{1}{\operatorname{aut}\lambda}\binom{n}{\lambda} p_0^r$ which, for

$$n^2 p_0 = \frac{2\alpha}{e} n^{1/2} + \frac{\log n}{2e} - \omega(1)\,,$$

gives us, with Proposition A.2.3:

$$\begin{aligned}
\mathbb{E}[X_\lambda] &= \frac{1}{r!s!}\frac{n!}{2^r} p_0^r \\
&\asymp n^{2r}\frac{p_0^r e^r (1+o(1))}{2^r r^r \sqrt{2\pi r}}\Theta(1) \\
&= \left(\frac{en^2 p_0}{2r}\right)^r \frac{1+o(1)}{\sqrt{2\pi r}} \\
&= \left(1 + \frac{(\log n)/4 - \omega(1)}{r}\right)^r \frac{\Theta(1))}{\sqrt{2\pi\sqrt{n}(\alpha+o(1))}} \\
&\le \exp((\log n)/4 - \omega(1))\frac{\Theta(1)}{\sqrt{2\pi\alpha} n^{1/4}} \\
&= \Theta(1)\exp(-\omega(1)) = o(1)\,.
\end{aligned} \tag{3.8}$$

So, by the 1st moment method (Lemma 2.1.3) we conclude that $\mathbb{P}[X_\lambda = 0] \to 1$. $\square$

The topic of big matchings discussed in the realm of $G(n,p)$ has direct applications in the discussion of thresholds in the variables X_λ where $\lambda = 1^r 2^s$. Namely, we know that there is a matching that covers all non-isolated vertices for $np_0 = \frac{\log n}{2} + \log log n + \omega(1)$ according to Theorem 3.2.8, then if there are at least $2r$ non-isolated vertices in $G(n, p_0)$, there is a stable set of type $1^s 2^r$ in $G(n,p)$.

Theorem 3.2.18. Fix reals ϵ and k such that $\epsilon > 0$ and $k \in (0.5, 1)$. Let $s = (1+\epsilon)n^{1-k}$ and $r = \frac{n-s}{2}$, corrected to integers if needed, such that $\lambda = 1^s 2^r \vdash n$. Then, for $p = 1 - \frac{k\log n}{n}$ where $k \in (0.5, 1)$ we have that

$$\mathbb{P}[X_\lambda = 0] \to 0\,.$$

Proof. We wont bother with remarks regarding the corrections for s and r to be integers.

Let X be the number of isolated vertices and I_v the indicator variable for the event "v is an isolated vertex". Let $p_0 = 1 - p = 1 - \frac{k \log n}{n}$.

We know that $G(n, p_0)$ satisfies the PM property a.a.s. for $p_0 = \frac{\log n}{2n}$, according to Theorem 3.2.8. So there is a matching that covers all but one non-isolated vertex. Hence, it is enough to show that $X \le s - 1$ a.a.s.

From Proposition A.2.2, and setting $p_0 = \frac{k \log n}{n}$ we have,

$$\begin{aligned}
\mathbb{E}[X] &= \sum_v \mathbb{E}[I_v] = n(1 - p_0)^{n-1} \\
&\le n \exp\left(-\frac{nk \log n}{n} + p_0\right) \\
&\le (1 + o(1)) n^{1-k}.
\end{aligned}$$

Recall from Example 2.1.7 that $\mathrm{Var}[X] = o(\mathbb{E}[X]^2)$ for any $p_0 = o(1)$. Then, we know that the number of isolated vertices is concentrated around its mean, since for $a = \Theta(\mathbb{E}[X])$, Chebichev's inequality (Lemma 2.1.2) gives us

$$\mathbb{P}[|X - \mathbb{E}[X]| > a] \le \frac{\mathrm{Var}[X]}{a^2} = o(1).$$

Hence $X < \mathbb{E}[X] + a = (1 + \epsilon)\mathbb{E}[X]$ a.a.s. for any small constant ϵ. Or, for any $\epsilon > 0$,

$$X < (1 + \epsilon/2) n^{1-k} \text{ a.a.s.}$$

This concludes that there is a matching that covers all but $(1 + \epsilon/2) n^{1-k} + 1 < (1 + \epsilon) n^{1-k}$ vertices a.a.s. $\square$

So we have the following table describing the bounds on $p_0 = 1 - p$ for the thresholds, where the variables C_i are to be read as suitable constants, and $k \in (0.5, 1)$:

Upper Bound	Number of edges in a Matching	Lower Bound
$\frac{\log n}{n} + \omega(n^{-1})$	Perfect Matching	$\frac{\log n}{n} - \omega(n^{-1})$
$\frac{k \log n}{n}$	$0.5n - (1 + o(1)) n^{1-k}$	?
?	$\Theta(n)$	$C_1 n^{-1} + C_2 n^{-2} \log n - \omega(n^{-2})$
?	$\Theta(\sqrt{n})$	$C_1 n^{-3/2} + C_2 n^{-2} \log n - \omega(n^{-2})$

Note that the lower bounds on the last two columns are not surprising, specially if we consider that there are, in expected value, $M \asymp n^2 p$ many edges.

3.3 Coefficients over the power sum basis

We have seen that the power sum coefficients are given by Theorem 1.2.7, as

$$A_\mu = \sum_{\substack{A \subseteq E(G(n,p)) \\ \lambda(A)=\mu}} (-1)^{\#A}.$$

For the tree case, we have the simpler expression given by Theorem 1.2.8, as

$$A_\mu = (-1)^{n-l(\mu)}\Theta_\mu = (-1)^{n-l(\mu)} \sum_{\substack{A \subseteq E(T_n) \\ \lambda(A)=\mu}} 1\,.$$

For that reason, we consider here the Tree Model T_n introduced in Definition 2.3.1.

If we denote by E_A the indicator variable for the event $A \in E(T_n)$ we get that $E_A \sim \mathrm{Ber}(p_A)$, where $p_A = \mathbb{P}[A \subseteq E(T_n)] = p(\mu)n^{-n+l(\mu)}$ is given from Proposition 2.3.3.

Then

$$\Theta\mu = \sum_{\substack{A \subseteq E(K_n) \\ \lambda(A)=\mu}} E_A\,, \tag{3.9}$$

is a sum of Bernoulli variables which is reasonably simple to handle, for instance, we can compute its expected value, obtaining:

Proposition 3.3.1. In the uniform random tree model T_n, the expected value of the variables Θ_μ are

$$\mathbb{E}[\Theta_\mu] = n^{-n+l(\mu)} \frac{\prod_i \mu_i^{\mu_i-1}}{\operatorname{aut}\mu}\binom{n}{\mu}.$$

And for the coefficients on the power-sum basis:

$$\mathbb{E}[A_\mu] = (-1)^{-n+l(\mu)} n^{-n+l(\mu)} \frac{\prod_i \mu_i^{\mu_i-1}}{\operatorname{aut}\mu}\binom{n}{\mu} = (-n)^{-n+l(\mu)} \frac{\prod_i \mu_i^{\mu_i-1}}{\operatorname{aut}\mu}\binom{n}{\mu}.$$

Proof. By linearity of expectation and $\mathbb{E}[E_A] = \mathbb{P}[A \subseteq E(T_n)] = p_A$ if A has no cycles, obtained in Proposition 2.3.3, we get:

$$\begin{aligned}
\mathbb{E}[\Theta_\mu] &= \sum_{\substack{A\subseteq E(K_n)\\ \lambda(A)=\mu\\ A \text{ has no cycles}}} p_A \\
&= \frac{1}{\operatorname{aut}\mu}\binom{n}{\mu}\prod_i \mu_i^{\mu_i-2} p(\mu) n^{-n+l(\mu)} \\
&= n^{-n+l(\mu)} \frac{\prod_i \mu_i^{\mu_i-1}}{\operatorname{aut}\mu}\binom{n}{\mu}.
\end{aligned}$$

Here we count the number of subsets A of $E(K_n)$ of type μ without cycles in the following manner: for each set partition of type μ we count the number of possible trees to span the different blocks, and add it all up to obtain $\frac{1}{\operatorname{aut}\mu}\binom{n}{\mu} \times \prod_i \mu_i^{\mu_i-2}$.

The results follow from $A_\mu = (-1)^{n-l(\mu)}\Theta_\mu$. □

This Proposition can be readily applied to compute basic properties regarding trees, as we are going to see in Proposition 3.3.4.

Remark 3.3.2. We have obtained in (3.9) that Θ_μ is a sum of Bernoulli random variables. Hence, it's normal to try to compute its variance, specially since we have already a formula for the expected value in Proposition 3.3.1. This is as far as this work goes in this topic:

$$\begin{aligned}
\mathbb{E}[\Theta_\mu^2] &= \sum_{\substack{A,B\subseteq E(K_n)\\ \lambda(A)=\lambda(B)=\mu}} \mathbb{E}[E_A E_B] \\
&= \sum_{\substack{A,B\subseteq E(K_n)\\ \lambda(A)=\lambda(B)=\mu}} \mathbb{E}[E_{A\cup B}] \\
&= \sum_{C\subseteq E(K_n)} \mathbb{E}[E_C] \left(\sum_{\substack{A,B\subseteq E(K_n)\\ \lambda(A)=\lambda(B)=\mu\\ A\cup B=C}} 1 \right).
\end{aligned}$$

Hence, to obtain concentration results we need only to compute, for the edge sets C without cycles,

$$\#\{A,B\subseteq E(K_n) | \lambda(A)=\lambda(B)=\mu \text{ and } A\cup B = C\}.$$

There are some issues in counting these objects, specially because the behaviour of $A\cup B$ only with the information that $\lambda(A)=\lambda(B)$ is hectic. But a naturally further work to tackle this problem.

3.3.1 Application to count number of centroids

In [18] and in [13] we are presented with the notion of the centroid of a tree and it is introduced an algorithm to construct a given tree knowing only its chromatic symmetric function and reduced additional information. This comes along from a conjecture present in [19], that says that any two non-isomorphic trees have different chromatic symmetric function. In the following we will compute quite easily the probability that T_n has two centroids from Proposition 3.3.1.

Definition 3.3.3 (Centroid)**.** Given a vertex v in a tree T, call the *weight* of a vertex v, or $s(v)$, to the maximum size of a connected component in $T \setminus v$. The *weight* of a tree $s(T)$ is the minimum weight among its vertices $s(T) = \min_{v \in T} s(v)$. A vertex v is called a *centroid* if $s(v) = s(T)$

In [18], [13] and [14] it is shown that a tree has either one or two centroids, and has exactly two centroids if and only if T has an even number of n vertices and $\theta_{(n/2,n/2)}(T) = 1$. A tree has exactly one centroid if and only if $\theta_{(n/2,n/2)} = 0$.

Hence we have that $\mathbb{P}[T_n \text{ has two centroids }] = \mathbb{E}[\Theta_{(n/2,n/2)}]$ and with this we get:

Proposition 3.3.4. For n even, big enough, we have that:

$$\mathbb{P}[T_n \text{ has two centroids }] \sim \frac{2\sqrt{2}}{\sqrt{n}\sqrt{\pi}}.$$

Proof. We just apply Proposition 3.3.1 to

$$\mathbb{P}[T_n \text{ has two centroids }] = \mathbb{E}[\Theta_{(n/2,n/2)}],$$

yielding,

$$\begin{aligned}
\mathbb{P}[T_n \text{ has two centroids }] &= \mathbb{E}[\Theta_{(n/2,n/2)}] \\
&= n^{-n+2} \frac{\left((n/2)^{n/2-1}\right)^2}{2!} \binom{n}{n/2, n/2} \\
&= \frac{\binom{n}{n/2,n/2}}{2^{n-1}}.
\end{aligned}$$

Now recall from Proposition A.2.3 that we have $\binom{n}{n/2,n/2} \sim \frac{2^n}{\sqrt{n\pi/2}}$ so,

$$\mathbb{P}[T_n \text{ has two centroids }] \sim \frac{2^n}{\sqrt{n\pi/2}2^{n-1}} = \frac{2\sqrt{2}}{\sqrt{n\pi}}. \qquad \square$$

3.4 Biggest connected component problem

In chapter 2 we dealt with the phase transition events, with special emphasis on the evolution of the connected components of $G(n, p)$. It can be readily applied in the chromatic symmetric function realm.

Incidentally, directly from Lemmas 2.2.8 and 2.2.11 we get that:

Proposition 3.4.1. Suppose that we have $p = \frac{c}{n}$ such that c is constant. Let $k = k(n) = \Omega(\log n)$. Let $\lambda(n) \vdash n$ any partition of n with a part of size k.

- If $c < 1$ then $k > \frac{3\log n}{(1-\beta)^2} \Rightarrow \Theta_\lambda = 0$ a.a.s.
- If $c > 1$ and $k = n(1-\alpha(\beta)) + \Omega(n)$ then $\Theta_\lambda = 0$ a.a.s.

Proof. This is a Corollary from Lemmas 2.2.8 and 2.2.11. $\square$

Chapter 4

Summary and further work

4.1 Summary and Conclusions

The main goal of this thesis is to apply some results from the broad literature of random graphs to the chromatic symmetric function: an invariant introduced chromatic invariant by Stanley in [19].

With this thesis we discuss some aspects on the Erdös-Rényi model for random graphs, like thresholds for a perfect matching or for connectedness, and relate them with the coefficients from the chromatic symmetric function over two of the most important bases from the symmetric function space.

We establish some notation and the basic properties of thresholds in Chapter 1, where Theorem 1.2.14 is the main result that guarantees the existence of a threshold for monotone properties. Afterwards, we introduce some of the most useful machinery in Random graphs: the 1st and 2nd moment methods (Lemmas 2.1.3 and 2.1.6 respectively) as well as Chernoff bounds ((Lemma 2.1.8).

Then, we introduce the main properties of the models $G(n, p)$ and $G(n, M)$, by clarifying the duality between the models, in particular Theorem 2.2.3. We also introduce the phase transition with special emphasis on the connectedness concluding with Lemmas 2.2.8 and 2.2.11.

We conclude Chapter 2 with an intricate proof of Cayley's formula (Lemma 2.3.2) and a small introduction to tree models leading up to Proposition 2.3.3. Afterwards, in Chapter 3, we discuss the subgraph containment with Theorem 3.1.2 playing a main role of inspiration, and its connections to the coefficients $X_{1^s\mu}$, described in Theorem 3.1.8.

Then we focus on finding big matchings in random bipartite graphs, in Proposition 3.2.2, and in the Erdös-Rényi model, in Proposition 3.2.8. The conclusions of this discussion are the hitting time Corollaries 3.2.5 and 3.2.9. This preludes the discussion of the behaviour of coefficients of type $X_{1^s 2^r}$ for simple applications of the 1st moment method (Lemma 2.1.3) in Theorems 3.2.15 and 3.2.17, and is readily applied to find an upper bound for the threshold of such coefficients where $s = \Theta(n^{1-k})$ in Theorem 3.2.18.

We finally conclude Chapter 3 with the discussion of the case of trees and the coefficients A_λ.

The main conclusions in this work are, unsurprisingly, the results regarding the chromatic symmetric function in the random graphs realm. There are a description of the thresholds for $X_{1^s\mu} = 0$ in Theorem 3.1.8 and the results on thresholds for big matchings (and consequently on for $X_{1^s 2^r}$) in Theorems 3.2.15, 3.2.17 and 3.2.18.

4.2 Original work and Further work

Most of the work presented is inspired in [12], namely the discussions on the existence of thresholds, on the perfect matching in $G(n,n,p)$ and the PM property on $G(n,p)$, the subgraph containment thresholds, the duality of the models $G(n,p)$ and $G(n,M)$ and the connectedness threshold. Some of the remaining Lemmas are well known in the literature of random graphs, namely the ones presented in "probabilistic tools" such as the 1st and 2nd moment methods (Lemmas 2.1.3 and 2.1.6 respectively) as well as Chernoff bounds (Lemma 2.1.8).

The remaining work is original and inspired in the results presented in the literature and in the idea, brought up by Prof. Dr. Valentin Féray, of applying the chromatic symmetric function in the random graphs realm.

There are a few open problems left here for further work, mostly that rely on effective development of intricate sums. Main examples are Conjecture 3.2.16 and Remark 3.3.2.

The main idea behind Theorem 3.2.18 is to count the number of isolated vertices and use Theorem 3.2.8 to find when those are the only vertices that aren't covered. So, going over the computations in Equation (3.4) we find that the threshold for occurrence of bigger Hall sets is smaller than the one regarding the cherries, so there is an interval where the only hurdle for the existence of a big matching are the isolated vertices and the cherries. Is it possible to bound the number of cherries and isolated vertices to get a claim similar to Theorem 3.2.18?

We note that Theorem 3.1.8 provides no interval where the coefficients X_λ of type $\lambda = 1^s\mu = k^s\mu'$ for μ' non-trivial partition converge to a non-trivial distribution over $\mathbb{R}$. So another open problem is to find the exact threshold expression for such coefficients X_λ.

Appendices

Appendix A

Asymptotic Notation and useful inequalities

A.1 Notation

In this thesis we use the following conventions regarding notation on asymptotics:

- We say that $a = \omega_n(b)$ if $\lim_{n\to\infty} \frac{|a|}{|b|} = \infty$.
- We say that $a = o_n(b)$ if $\lim_{n\to\infty} \frac{|a|}{|b|} = 0$.
- We say that $a = \Omega_n(b)$ if $\liminf_{n\to\infty} \frac{|a|}{|b|} > 0$.
- We say that $a = O_n(b)$ if $\limsup_{n\to\infty} \frac{|a|}{|b|} < \infty$.
- We say that $a = \Theta_n(b)$ if $0 < \liminf_{n\to\infty} \frac{|a|}{|b|} \le \limsup_{n\to\infty} \frac{|a|}{|b|} < \infty$.

We also say that $a >> b$ or $b << a$ if $a = \omega(b)$. We also write $a \asymp b$ if $a = \Theta(b)$, mainly because the relation defined like this is an equivalence relation. If, additionally, we have that $\lim_{n\to\infty} \frac{|a|}{|b|} = 1$ then we also write $a \sim b$.

A.2 Useful propositions

We begin with the well known Stirling's approximation.

Proposition A.2.1.

$$n! = \left(\frac{n}{e}\right)^n \sqrt{2\pi n}(1 + o(1)).$$

Proof. The proof of this theorem can be found in [16]. $\square$

Now follows some trivial inequalities that will turn out to be useful. Namely, Bernoulli's inequality:

Proposition A.2.2. We have $1-x \le e^{-x}$, or equivalently, $1+x \le e^x$, for all $x \in \mathbb{R}$.

We also have, for $x \ge 0$, that $1-x+x^2/2 \ge e^{-x}$.

We have as well Bernoulli's inequality, $1+nx \le (1+x)^n$ for $x > -1$ and n a non-negative integer.

Proof. Let $f(x) := e^x - x - 1$ and note that $f'(x) = e^x - 1$ vanishes only when $x = 0$ which is a local minimum because $f''(0) = e^0 = 1 > 0$. Since $\lim_{x\to-\infty} f(x) = \lim_{x\to+\infty} f(x) = +\infty$, $x = 0$ is the global minimum and $f(x) \ge f(0) = 0\ \forall x \in \mathbb{R}$.

Now call $g(x) = 1-x+x^2/2$ and suppose that $x \ge 0$. Then

$$g(x) = \int_0^x g'(t)dt + g(0) = \int_0^x (-1+t)dt + 1 \ge \int_0^x -e^{-t}dt + 1 = e^{-x},$$

or for short, $g(x) \ge e^{-x}$.

To obtain that $1+nx \le (1+x)^n$ for $x > -1$ and $n \ge 0$ integer, we use induction on n, where the base case $n = 0$ is trivial. Now suppose that it holds for n, then

$$(1+x)^{n+1} = (1+x)^n(1+x) \ge (1+nx)(1+x) = 1+(n+1)x+nx^2 \ge 1+(n+1)x,$$

concluding the induction argument. $\square$

Proposition A.2.3. If $s = O(\sqrt{n})$ then $\binom{n}{s} \asymp \frac{n^s}{s!}$.

For all n, s positive integers we have

$$\left(\frac{n}{s}\right)^s \le \binom{n}{s} \le \left(\frac{en}{s}\right)^s.$$

Additionally,

$$\binom{2n}{n} \sim \frac{4^n}{\sqrt{n\pi}}.$$

Proof. Note that

$$\frac{n!}{(n-s)!} = n^s \prod_{i=0}^{s-1}\left(1-\frac{i}{n}\right).$$

Now according to Proposition A.2.2, we have that $1-\frac{i}{n} \ge \exp(-i/n) - i^2/n^2$ so we have,

$$\frac{n!}{(n-s)!} \geq n^s \prod_{i=0}^{s-1} \exp(-i/n) - i^2/n^2 = n^s \left(\exp(\Theta(1)) - O(\exp(\Theta(1)) \sum_{i=0}^{s-1} i^2/n^2) \right) = n^s (\Theta(1) + O(n^{-1/2})),$$

Which concludes that $\binom{n}{s} \asymp \frac{n^s}{s!}$.

Now note that for $a > b > 0$ and $i \geq 0$ integers we have $\frac{a}{b} \geq \frac{a+i}{b+i}$ so

$$\binom{n}{s} = \prod_{i=0}^{s-1} \frac{n-i}{s-i} \geq \prod_{i=0}^{s-1} \frac{n}{s} = \left(\frac{n}{s}\right)^s.$$

On the other hand, recall that $e^s = \sum_{k \geq 0} \frac{s^k}{k!} \geq \frac{s^s}{s!}$ hence

$$\binom{n}{s} \leq \frac{n^s}{s!} \leq \frac{n^s e^s}{s^s}.$$

No obtain that $\binom{2n}{n} \asymp \frac{4^n}{\sqrt{n}}$ we use Stirling's approximation A.2.1 to obtain:

$$\begin{aligned}
\binom{2n}{n} &= \frac{(2n)!}{n!^2} \\
&= \left(\frac{2n}{e}\right)^{2n} \left(\frac{e}{n}\right)^{2n} \frac{\sqrt{2\pi 2n}}{2\pi n} (1 + o(1)) \\
&= 2^{2n} \frac{1 + o(1)}{\sqrt{n\pi}}.
\end{aligned}$$

This concludes that $\binom{2n}{n} \sim \frac{4^n}{\sqrt{n\pi}}$. □

Appendix B

Probabilistic notation

The notation here presented follows the lines on [12]. Additionally, we also write $X \sim Y$ for two random variables if they have the same distribution.

We say that $X_n = a$ a.a.s. or $X_n \to a$ if $\mathbb{P}[X_n \neq a] = o(1)$.

We also say that $X_n = \infty$ a.a.s. or$X_n \to \infty$ if $\mathbb{P}[X_n \leq a] = o(1)$ for any real number a.

We also note that $G(n, 1-p) \sim G(n, p)^c$.

We use a notion of *stochastic order* among the random variables. We say that for two real valued random variables X and Y, X is stochastically bigger than Y, or $X \geq_S Y$ if $\mathbb{P}[X \geq a] \geq \mathbb{P}[Y \geq a]$ for any real a. Trivially, this implies that if $X \geq_S Y$ then $\mathbb{E}[X] \geq \mathbb{E}[Y]$.

Remark B.0.4. Equivalently, $X \geq_S Y$ if there are two random variables X_0, Y_0 in the same probability space such that $X \sim X_0$, $Y \sim Y_0$ and $\mathbb{P}[X_0 \geq Y_0] = 1$. This notion of stochastic order also makes clear that $\mathbb{E}[X] \geq \mathbb{E}[Y]$, as $X_0 - Y_0$ is a non-negative random variable.

In Chapter 2 we claim that the size X of the branching process where we delete some cells before having offspring is stochastically smaller than the size Y of the original branching process. We simply use the criteria in Remark B.0.4, since in the joint probability space we have $\mathbb{P}[X \leq Y] = 1$.

Bibliography

[1] Aigner, M., Ziegler, G. M., Hofmann, K. H., and Erdos, P. (2010). *Proofs from the Book*, volume 274. Springer.

[2] Alon, N. and Spencer, J. H. (2004). *The probabilistic method*. John Wiley & Sons.

[3] Bollobás, B. and Thomason, A. (1985). Random graphs of small order. *North-Holland Mathematics Studies*, 118:47–97.

[4] Bondy, A. and Murty, U. (1991). *Graph Theory with Applications*. Wiley.

[5] Diestel, R. (2000). *Graph Theory*, volume 1. Springer-Verlag New York, 2nd edition.

[6] Erdös, P. and Rényi, A. (1959). On random graphs i. *Publ. Math. Debrecen*, 6:290–297.

[7] Erdős, P. and Rényi, A. (1966). On the existence of a factor of degree one of a connected random graph. *Acta Mathematica Hungarica*, 17(3-4):359–368.

[8] Fougere, J. (2003). On symmetric chromatic polynomials of trees, undergraduate thesis. Master's thesis.

[9] Friedgut, E., Bourgain, J., et al. (1999). Sharp thresholds of graph properties, and the k-sat problem. *Journal of the American mathematical Society*, 12(4):1017–1054.

[10] Guerreiro, J. (2009). Partições de inteiros - presentation for novos talentos em matemática, fundação calouste gulbenkian.

[11] Györi, E., Rothschild, B., and Runciński, A. (1985). Every graph is contained in a sparsest possible balanced graph. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 98, pages 397–401. Cambridge Univ Press.

[12] Janson, S., Luczak, T., and Rucinski, A. (2011). *Random graphs*, volume 45. John Wiley & Sons.

[13] Orellana, R. and Scott, G. (2014). Graphs with equal chromatic symmetric functions. *Discrete Mathematics*, 320:1–14.

[14] Penaguião, R. (2016). Algebraic invariants in graphs. ETH Zürich, Semester Project.

[15] Privault, N. (2013). *Understanding Markov Chains: Examples and Applications*. Springer Undergraduate Mathematics Series. Springer Singapore.

[16] Robbins, H. (1955). A remark on Stirling's formula. *The American Mathematical Monthly*, 62(1):26–29.

[17] Ross, S. M. et al. (1996). *Stochastic processes*, volume 2. John Wiley & Sons New York.

[18] Smith, I., Smith, Z., and Tian, P. (2015). Symmetric chromatic polynomial of trees. *arXiv preprint arXiv:1505.01889*.

[19] Stanley, R. P. (1995). A symmetric function generalization of the chromatic symmetric polynomial of a graph. *Advances in Mathematics*, 111(1):166 – 194.

[20] Stanley, R. P. (1999). *Enumerative Combinatorics*, volume 2 of *Cambridge Studies in advanced Mathematics*. Chambridge University Press.

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

The chromatic symmetric function on random graphs

Authored by (in block letters):
For papers written by groups the names of all authors are required.

Name(s):	First name(s):
Penaguião	Raúl

With my signature I confirm that
- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date

Zürich, 6th September

Signature(s)

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.