Random Planar Structures and Random Graph Processes

Habilitationsschrift

zur Erlangung der Lehrbefähigung
für das Fach Informatik

vorgelegt dem Rat der Mathematisch-Naturwissenschaftlichen Fakultät II
der Humboldt-Universität zu Berlin

von

Dr. Mihyun Kang
geboren am 5. August 1973 in Jeju, Südkorea

Prof. Dr. Christoph Markschies
Präsident
der Humboldt-Universität zu Berlin

Prof. Dr. Wolfgang Coy
Dekan

Berlin, den 20. Juni 2007

Gutachter:
1. Prof. Dr. Michał Karoński
2. Prof. Dr. Hans Jürgen Prömel
3. Prof. Dr. Günter M. Ziegler

Antrag auf Zulassung zum Habilitationsverfahren:	31.10.2006
Zulassung zum Habilitationsverfahren:	27.11.2006
Annahme der schriftlichen Habilitationsleistung:	23.04.2007
Öffentlicher Vortrag:	20.06.2007

Contents

Part I

Basics

Chapter 1

Introduction

This thesis focuses on two kinds of discrete structures: planar structures, such as planar graphs and subclasses of them, and random graphs, particularly graphs generated by random processes.

A *planar graph* is a graph which can be embedded in the plane without crossing edges. Kuratowski's theorem states that planar graphs can be characterized in terms of forbidden minors, namely $K_{3,3}$ and K_5 (see e.g., [54]). Planar structures, in particular planar graphs, have been extensively studied during the last few decades, including the proofs of the famous four colour theorem [5, 6, 123]. *Random* planar structures, however, have been investigated only during the last few years [13, 48, 69, 70, 71, 99, 112]. In this thesis we study the following aspects of (random) planar structures:

- How many of them are there (exactly or asymptotically)?
- How can we efficiently sample a random instance uniformly at random?
- What properties does a random planar structure have? E.g., what is the probability of connectedness? How many edges are there in average? What is the chromatic number?

Random graphs were first introduced by Erdős and Rényi [59, 60] and studied extensively since (see e.g., the monographs [33, 82]).

Classical random graph models include a uniform random graph, a binomial random graph, and a random graph process. In a uniform random graph model, a graph $G(n, m)$ is chosen uniformly at random among all graphs with n vertices and m edges. In a binomial random graph model, each of the possible pairs of vertices is adjacent in a graph $G(n, p)$ independently with probability p. A random graph process $(\mathcal{G}_0, \mathcal{G}_1, \cdots, \mathcal{G}_{\binom{n}{2}})$ is an algorithmic version of the uniform random graph model $G(n, M)$, where the graph generation proceeds as follows: Starting with n isolated vertices $\mathcal{G}_0$, a new graph $\mathcal{G}_{M+1}$ is obtained from a current graph $\mathcal{G}_M$ by adding a new random edge in each step. The dynamical viewpoint of the random graph process provides a picture of how the random graph process evolves or how a structural property changes as the graph evolves from empty to full. Random graph processes *with degree restrictions* recently attracted a lot attention [73, 74, 121, 122, 124, 125]. In the thesis, we study

random graph processes where the minimum degree grows quite quickly with the following questions in mind:

- How does the connectedness of a graph generated by a random graph process change as the number of edges increases?
- How does the structure of its components evolve?
- When does the phase transition occur?
- How big is the largest component?

1.1 Classes, models, and methods

In this section we specify the type of planar structures and the models of random graph processes with degree restrictions studied in this thesis, and the methods employed to study them.

Graph classes. Planar structures studied in this thesis are

- forests,
- outerplanar graphs,
- cubic planar graphs, and
- planar graphs.

We study *labeled* planar structures and *unlabeled* planar structures separately, since the latter requires more techniques. One of the most well-studied subclasses of planar graphs are *forests*, i.e., graphs without cycles. In the thesis forests are discussed for the illustration of the methods that we will use for other classes of graphs. Another interesting subclass of planar graphs are *outerplanar graphs*, i.e., graphs that can be embedded in the plane in such a way that there is a face containing all the vertices. In terms of forbidden minors, forests are the graphs without K_3 minors, and outerplanar graphs are the graphs without $K_{2,3}$ and K_4 minors. A *cubic planar graph* is a planar graph where each vertex has exactly three neighbors, i.e., a 3-regular planar graph.

Graph process models. Random graph processes of interest in this thesis are

- the minimum degree random multigraph process and
- the min-min random graph process.

The *minimum degree random multigraph process* $(G_{\min}(n,0), G_{\min}(n,1), \cdots)$ is defined by the rule that $G_{\min}(n, M+1)$ is obtained from $G_{\min}(n,M)$ by first choosing uniformly at random a vertex of minimum degree and then connecting it with another vertex chosen uniformly at random among all the remaining vertices. The *min-min random graph process* $(G_0, G_1, \cdots)$ is defined by the rule that G_{M+1} is obtained from G_M by choosing a pair $\{v,w\}$ of distinct nonadjacent vertices of G_M of minimum degree uniformly at random among all such pairs and adding a new edge $\{v,w\}$. (If it impossible to continue this way, we restart the process from the empty graph.)

Methods. The main tools to investigate *planar structures* are the recursive method, the singularity analysis, and the probabilistic method, based on the decomposition of planar structures along the connectivity (see Figure 1.1).

For exact enumeration and uniform sampling of planar structures we use the *recursive method*:

- Decompose the planar structures along the connectivity, yielding a decomposition tree.
- Derive recursive counting formulas according to the decomposition tree.
- Sample a graph as a reversed procedure of decomposition according to the probabilities computed by recursive counting formulas.

For asymptotic enumeration we use the *singularity analysis* of generating functions:

- Interpret the decomposition of *labeled* structures directly in terms of exponential generating functions. In case of *unlabeled* structures we interpret the decomposition of them in terms of cycle indices, from which we derive ordinary generating functions.
- Determine the dominant singularities of generating functions and their singularity types.
- Derive the asymptotic numbers.

Finally, using the *probabilistic method* we derive typical properties of a *random planar structure*.

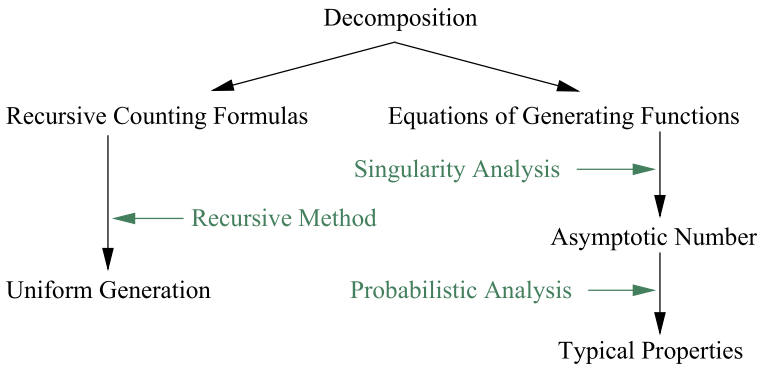


Figure 1.1: Decomposition of a planar structure

To investigate the *random graph processes* we use the *probabilistic method*, Wormald's differential equation method, multi-type branching processes, and the singularity analysis:

- *Wormald's differential equation method* describes the average change of a certain random variable when a new edge is added in the random graph process as a differential equation.
- *Multi-type branching processes* approximate random graph processes with several types of vertices involved in the evolution, and can be applied to study the phase transition of the random graph process via the branching process.
- The *dominant singularity* of a probability generating function describing the evolution of the random graph process determines the order of the largest component.

1.2 Summary of results

In this section we present the main results of this thesis, which answer the typical questions about the considered graph classes and graph process models.

Planar structures. The main results on planar structures are about their asymptotic numbers, typical properties, and efficient uniform sampling algorithms.

First, using the singularity analysis we determine that the number g_n of *labeled* planar structures of certain classes on n vertices is asymptotically of the form $(1+o(1))\, c\, n^{-\alpha}\, \beta^n\, n!$, for suitable constants c, α, and β, while the number of *unlabeled* planar structures is asymptotically of the form $(1+o(1))\, c\, n^{-\alpha}\, \beta^n$. The constant α is called the *critical exponent* and β the *growth constant*. We compute the values of α and β for (labeled and unlabeled) outerplanar graphs and labeled cubic planar graphs (see Table 1.1).

Curiously, the critical exponents α for labeled outerplanar graphs and for labeled cubic planar graphs differ by one: α is $5/2$ for outerplanar graphs and $7/2$ for cubic planar graphs. These critical exponents are directly computed from the singular expansions of the generating functions $g(z) = \sum_n g_n/n! z^n$ with singular type $(1-\beta z)^{3/2}$ for outerplanar graphs (see (5.2.3)) and with singular type $(1-\beta z)^{5/2}$ for cubic planar graphs (see (6.4.9) and (6.4.10)). However, we do not know what structural properties of planar structures determine the singular type.

Next, we investigate asymptotic properties of a random planar structure. The expected number of edges in a random planar structure on n vertices is of the form $(1+o(1))\, \mu\, n$ for a positive constant μ. A random planar structure on n vertices is connected with probability tending to a constant p_{con} as $n \to \infty$, and the chromatic number of a random planar structure on n vertices is three with probability tending to a constant p_χ as $n \to \infty$. We compute the values of μ, p_{con} and p_χ for outerplanar graphs and cubic planar graphs (see Table 1.1).

For example, we prove that the chromatic number of a random cubic planar graph on n vertices is three with probability tending to 0.999 as $n \to \infty$, while the chromatic number of a random *connected* cubic planar graph on n vertices is three with probability tending to *one* as $n \to \infty$. To this end, we show that the number of isolated K_4's in a random cubic planar graph on n vertices has asymptotically Poisson distribution and that a random cubic planar graph contains linearly many triangles with probability tending to one as $n \to \infty$. As a consequence, together with Brooks' theorem, we obtain the result on the chromatic number of a random cubic planar graph.

Table 1.1 provides the values of α, β, μ, $p_{\rm con}$, and p_χ for trees, outerplanar graphs, cubic planar graphs, and planar graphs. Note that every cubic planar graph on n vertices has *exactly* $3n/2$ edges (provided that n is even), and that every tree on n vertices has $n-1$ edges and is connected, and its chromatic number is two ($n \geq 2$). The values for trees (see [63]) and for planar graphs (see [71]) are included in the table for comparison.

Classes	α	β	μ	$p_{\rm con}$	p_χ
Labeled trees	5/2	2.71	1	1	
Unlabeled trees	5/2	2.95	1	1	
Labeled outerplanar graphs	5/2	7.32	1.56	0.861	1
Unlabeled outerplanar graphs	5/2	7.50	1.54	0.845	1
Labeled cubic planar graphs	7/2	3.13	1.50	≥ 0.998	0.999
Labeled planar graphs	7/2	27.2	2.21	0.963	

Table 1.1: The (first few digits of the) values of the critical exponent α, the growth constant β, the edge density μ, the probability $p_{\rm con}$ of connectedness, and the probability p_χ of the chromatic number being three.

Furthermore, using the recursive method we design the *first* polynomial time algorithms that sample an outerplanar graph, a cubic planar graph, and a planar graph on n vertices, uniformly at random, with running time and memory requirement presented in Table 1.2.

Classes	Running time	Memory
labeled or unlabeled trees	$\tilde{O}(n^4)$	$\tilde{O}(n^3)$
unlabeled trees	$\tilde{O}(n^4)$	$\tilde{O}(n^3)$
Labeled outerplanar graphs	$\tilde{O}(n^4)$	$\tilde{O}(n^3)$
Unlabeled outerplanar graphs	$\tilde{O}(n^7)$	$\tilde{O}(n^3)$
Labeled cubic planar graphs	$\tilde{O}(n^6)$	$\tilde{O}(n^3)$
Labeled planar graphs	$\tilde{O}(n^7)$	$\tilde{O}(n^4)$

Table 1.2: Running time and memory requirement of uniform samplers.

Finally, we provide a complete proof of the (labeled and unlabeled) map enumeration obtained from the Gaussian matrix integral. We also apply the method of the Gaussian matrix integral to other problems, e.g., we show that the enumeration of the graphs *embeddable* on a given 2-dimensional surface, particularly planar graphs with a given degree sequence, can also be formulated by the Gaussian matrix integral.

Random graph processes. The main results on the random graph processes with degree restrictions concern the connectedness, the phase transition, and the giant component.

First, we show that the graph $G_{\min}(tn)$ generated by the minimum degree random multigraph process and the graph G_{tn} generated by the min-min ran-

dom graph process become *connected* as soon as the minimum degree reaches three, with probability tending to one as $n \to \infty$.

Next, we prove that there is a constant h_{cr} such that $G_{\min}(tn)$ enjoys the *phase transition* when $t = h_{\mathrm{cr}}$. That is, $G_{\min}(tn)$ consists of small components of order $O(\log n)$ when $t < h_{\mathrm{cr}}$, while it consists of one giant component of order $\Theta(n)$ and small components of order $O(\log n)$ when $t > h_{\mathrm{cr}}$, with probability tending to one as $n \to \infty$.

Finally, using Fourier transforms we determine the asymptotic distribution of the order of the *giant component* in G_{tn}. More precisely, we show that for a constant $0 < t < 1/2$ and for any positive integer l, the giant component in G_{tn} is of order $n - 2l$ with probability tending to $2\exp(2t(t-1))\sqrt{\frac{t(1-t)}{\pi l}}(1-2t)^{2l}$ as $n \to \infty$.

1.3 Overview of the chapters

The details of the main results stated above can be found in the relevant chapters. In this section we overview the thesis and present the main theorems in each chapter.

The thesis consists of four parts. In the rest of Part I we recall concepts and facts on planar structures, random graphs, and some other fields relevant to or necessary for the thesis. In Parts II and III we present the enumeration results, uniform sampling algorithms, and properties of labeled and unlabeled planar structures. Finally, in Part IV we discuss the minimum degree random multigraph process and the min-min random graph process.

Notations for asymptotics. Before stating the results, we introduce some notations.

The expression *asymptotically almost surely*, abbreviated by *a.a.s.*, means *with probability tending to one as* $n \to \infty$.

For two sequences $\{a_n\}_{n\geq 0}$ and $\{b_n\}_{n\geq 0}$ and two numbers a, b, we write $a_n \sim b_n$ if $\lim_{n\to\infty} a_n/b_n = 1$ and $a \approx b$ if a is numerically approximated by b. Given a real number x up to the first l digits (in decimal expansion) we write $a \doteq x$ if the first l digits of a equal x. For example, $a \doteq 1.23$ if the first three digits of a equal 1.23. Furthermore the notation $\tilde{O}(\cdot)$ denotes the growth up to logarithmic factors.

Part I. Basics. In Chapters 2 and 3 we provide backgrounds on planar structures and random graphs.

In Chapter 2 we first dicuss the current state of the art in the study of planar structures from the aspects of enumeration and uniform sampling. We then explain how to count *labeled* planar structures using the recursive method and exponential generating functions. We also illustrate how to count *unlabeled* planar structures using *cycle indices* and ordinary generating functions. Next we show how to determine the asymptotic numbers from the generating functions. To this end, we have to determine the dominant singularities and the singular types of the generating functions. Finally, we discuss techniques for uniform sampling, such as Markov chain Monte Carlo method, Boltzmann sampler, and the recursive method.

In Chapter 3 we first review well-known results on classical random graphs, such as evolution, the phase transition, planarity, and connectedness. We then move to random regular graphs and random graphs with a given degree sequence. We discuss also recent development of random graph processes with degree constraints. Finally, we present several probability distributions, basic inequalities, and facts from Wormald's differential equation method and multi-type branching processes.

Part II. Labeled planar structures. In Part II, from Chapter 4 to Chapter 7, we study *labeled* planar structures, such as forests, outerplanar graphs, cubic planar graphs, and planar graphs.

In Chapter 4 we study labeled *forests* and labeled *trees*. We present counting formulas and a uniform sampling algorithm based on the recursive method (see Figure 4.2). Note, however, that for the uniform generation of labeled trees, a linear-time algorithm was already known [4]. The main purpose of this chapter is to illustrate how to use the generating functions and the recursive method for labeled structures.

Theorem 4.4.1. *Labeled trees on n vertices can be sampled uniformly at random in deterministic time $\tilde{O}(n^4)$ with space $O(n^3 \log n)$. This can also be done in deterministic time $\tilde{O}(n^2)$ if we apply a precomputation step.*

In Chapter 5 we study labeled *outerplanar* graphs. Outerplanar graphs are relatively simple compared to planar graphs, but rich enough so that we have to understand the properties of outerplanar graphs when we want to decompose them into smaller parts and apply the recursive method. The property that every 2-connected outerplanar graph contains a unique Hamiltonian cycle yields that the decomposition can be interpreted in terms of a simple generating function of square root type (see (5.1.1)).

From a complete set of decompositions of labeled outerplanar graphs along the connectivity, we derive the equations of generating functions. Using the singularity analysis of generating functions (see (5.2.2) and (5.2.3)) we determine the asymptotic number of labeled outerplanar graphs, and investigate typical properties of a random outerplanar graph. Furthermore, using the recursive method we derive the counting formulas for labeled outerplanar graphs with various connectivity and derive an algorithm that samples a random outerplanar graph in polynomial time.

Theorem 5.2.1. *The number of labeled outerplanar graphs on n vertices is asymptotically*

$$(1 + o(1))c_1 n^{-5/2} \gamma^n n!,$$

where $c_1 \doteq 0.008095$ and $\gamma \doteq 7.32098$.

Let $G(n)$ denote a random outerplanar graph chosen uniformly at random among all the labeled outerplanar graphs on vertices $1, \ldots, n$.

Theorem 5.3.2. *A random outerplanar graph $G(n)$ satisfies that*

$$\lim_{n\to\infty} \Pr(G(n) \text{ is connected}) = 0.861666,$$

whereas $\lim_{n\to\infty} \Pr(G(n)$ is 2-connected$) = 0$.

Theorem 5.3.3. *Let Y_n be the number of edges in $G(n)$. Then*

$$\mathbb{E}[Y_n] \sim 1.56251n, \quad \mathrm{Var}(Y_n) \sim 0.223992n.$$

Theorem 5.4.1. *A labeled outerplanar graph on n vertices can be sampled uniformly at random in deterministic time $\tilde{O}(n^4)$ with space $O(n^3 \log n)$. This can be done in deterministic time $\tilde{O}(n^2)$ if we allow a precomputation step to evaluate the numbers from the counting formulas.*

In Chapter 6 we study labeled *cubic* planar graphs. The restriction on the degree, i.e., the 3-regularity of graphs yields a nice characterization of cubic graphs with one distinguished vertex, called *rooted cubic graphs*. At the last step of the decomposition of rooted cubic graphs, the 3-regularity further involves a well-studied class of planar graphs, *triangulations*, as the dual of rooted 3-connected cubic planar graphs.

Using the characterization of rooted cubic planar graphs and triangulations we derive the equations of generating functions. We then determine the singular types of the generating functions of rooted cubic planar graphs (see (6.4.8)) and cubic planar graphs (see (6.4.9) and (6.4.10)). Using the singularity analysis we determine the asymptotic number of labeled cubic planar graphs, and investigate typical properties of a random cubic planar graph. Using the recursive method we also derive the counting formulas and a uniform sampling algorithm.

Theorem 6.4.1. *The number of labeled cubic planar graphs on n vertices is asymptotically*

$$(1 + o(1))c_2 n^{-7/2} \rho^{-n} n!,$$

for a constant $c_2 > 0$ and $\rho^{-1} \doteq 3.132595$.

For $k = 0, 1, 2, 3$ let $G_n^{(k)}$ denote a random cubic planar graph chosen uniformly at random among all the *k-vertex-connected* cubic planar graphs on vertices $1, \ldots, n$ for even n.

Theorem 6.5.1. *For a constant $\zeta(\rho)$ satisfying $0.998 \leq \zeta(\rho) < 1$,*

$$\lim_{n\to\infty} \Pr(G_n^{(0)} \text{ is connected}) = \zeta(\rho),$$

whereas $\lim_{n\to\infty} \Pr(G_n^{(0)}$ is 2-connected$) = 0$.

Lemma 6.5.2. *Let $\nu = \rho^4/4! \doteq 0.000432$. Let X_n be the number of components of $G_n^{(0)}$ isomorphic to K_4 for even n. Then X_n has asymptotically the Poisson distribution with mean ν. That is, for $k = 0, 1, 2, \ldots$*

$$\Pr(X_n = k) \to e^{-\nu} \frac{\nu^k}{k!} \quad \text{as } n \to \infty.$$

In particular, the probability that $G_n^{(0)}$ has at least one component isomorphic to K_4 tends to $1 - e^{-\nu}$ as $n \to \infty$ with n even.

Lemma 6.5.6. *Let $Y_n^{(k)}$ be the number of triangles in $G_n^{(k)}$ for $k = 0, 1, 2, 3$. Then for even n there exists $\delta > 0$ such that*

$$\Pr(Y_n^{(k)} \geq \delta n) = 1 - e^{-\Omega(n)}.$$

Theorem 6.5.7. *Let ν be as in Lemma 6.5.2, and let $\chi(G_n^{(k)})$ denote the chromatic number of $G_n^{(k)}$. Then we have*

$$\begin{aligned}\lim_{n\to\infty} \Pr(\chi(G_n^{(0)}) = 4) &= 1 - e^{-\nu},\\ \lim_{n\to\infty} \Pr(\chi(G_n^{(0)}) = 3) &= e^{-\nu} \doteq 0.999568,\end{aligned}$$

and for $k = 1, 2, 3$ we have $\lim_{n\to\infty} \Pr(\chi(G_n^{(k)}) = 3) = 1$.

Theorem 6.6.1. *A labeled cubic planar graph can be sampled uniformly at random in deterministic time $O(n^6 \log^3 n)$ and in $O(n^2)$ with pre-computation, which requires $O(n^3 \log n)$ space.*

In Chapter 7 we study labeled *planar* graphs. The main idea is to decompose graphs into smaller parts involving graphs with higher connectivity. In particular, for the decomposition of 2-connected graphs we use the characterization of 2-connected graphs with one distinguished oriented edge, called *networks*, due to Trakhtenbrot [136]. From the decomposition along the connectivity we derive the recursive counting formulas for labeled planar graphs with various connectivity and derive an algorithm that samples a random planar graph in polynomial time.

Theorem 7.7.1. *A labeled planar graph on n vertices and m edges can be sampled uniformly at random in deterministic time $\tilde{O}(n^7)$ with space $O(n^4 \log n)$. If we apply a preprocessing step, this can also be done in deterministic time $\tilde{O}(n^3)$.*

Part III. Unlabeled planar structures. In Part III, from Chapter 9 to Chapter 11, we study unlabeled planar structures (i.e., isomorphism classes of planar structures), such as forests, outerplanar graphs, 2-connected planar graphs, and maps and planar graphs with a given degree sequence.

Note that unlabeled structures are much more difficult to handle due to symmetry, and that the problem of determining the asymptotic number of *unlabeled planar graphs* is still open. One way of enumerating unlabeled structures uses *cycle indices* introduced by Pólya [118]: The symmetry of a graph is closely related to the orbits of the automorphism group of the graph, and Burnside's lemma can be used to express the number of orbits in terms of the number of objects fixed by the automorphism group. Cycle indices, which is a formal power series encoding the cycle decompositions of the automorphism group of a graph, is further translated as the number of unlabeled graphs.

In Chapter 8 we study unlabeled *trees* and present counting formulas and a uniform sampling algorithm.

Theorem 8.4.1. *Unlabeled trees on n vertices can be sampled uniformly at random in deterministic time $\tilde{O}(n^4)$ with space $\tilde{O}(n^3)$.*

In Chapter 9 we study unlabeled *outerplanar* graphs. We derive the asymptotic number of unlabeled outerplanar graphs, using cycle indices and the singularity analysis of generating functions. This is an important new step toward the enumeration of unlabeled planar graphs. We also study typical properties of a random unlabeled outerplanar graph. Furthermore, using the recursive method we derive the counting formulas for unlabeled outerplanar graphs and derive an algorithm that samples a random unlabeled connected outerplanar graph in *expected* polynomial time.

Theorem 9.3.4. *The number of unlabeled outerplanar graphs on n vertices is asymptotically*

$$(1 + o(1))c_3 n^{-5/2} \eta^{-n} n!,$$

where $c_3 \approx 0.021504$ and $\eta^{-1} \approx 7.503597$.

Theorem 9.4.1. *The probability that a random unlabeled outerplanar graph is connected is asymptotically ≈ 0.845721.*

Theorem 9.4.7. *The distribution of the number of edges in a random unlabeled outerplanar graph on n vertices is asymptotically* Gaussian *with mean μn and variance $\sigma^2 n$, where*

$$\mu \approx 1.54894, \quad \sigma^2 \approx 0.227504.$$

The same holds for a random connected *outerplanar graph.*

Theorem 9.4.8. *Let $\chi(G_n)$ denote the chromatic number of a random unlabeled outerplanar graph G_n on n vertices. Then we have*

$$\lim_{n\to\infty} \Pr(\chi(G_n) = 3) = 1.$$

Theorem 9.5.1. *An unlabeled outerplanar graph on n vertices can be sampled uniformly at random in expected time $\tilde{O}(n^7)$ with space $O(n^3 \log n)$.*

In Chapter 10, we study unlabeled *2-connected planar* graphs. We derive the counting formulas and a uniform sampling algorithm, using the recursive method. The main step is to control the *symmetry* of the graphs and find bijections between symmetric graphs and certain type of graphs called *coloured networks* that we can further decompose.

Theorem 10.6.1. *An unlabeled 2-connected planar graph with m edges can be sampled uniformly at random, in expected $O(m^9)$ time. If the algorithm has direct access to the values of appropriately chosen counting formulas, the algorithm can generate such an object in expected cubic time.*

In Chapter 11, we study maps and planar graphs with a given degree sequence. A technique of theoretical physics called *Wick's theorem* (see Theorem 11.1.1) interprets the Gaussian matrix integral of the product of the traces of powers of Hermitian matrices as the number of maps with a given degree sequence, while it is common in combinatorics to use the decomposition of maps yielding the equations of generating functions.

We first complete a proof of the map enumeration obtained from the Gaussian matrix integral. We then show that the number of planar graphs with a given degree sequence can be expressed as the Gaussian matrix integral. To be more precise, let $M = (M_{ij})$ be an $N \times N$ Hermitian matrix and $dM = \prod_i dM_{ii} \prod_{i<j} d\,\mathrm{Re}(M_{ij}) d\,\mathrm{Im}(M_{ij})$ denote the standard Haar measure, where $\mathrm{Re}(M_{ij})$ and $\mathrm{Im}(M_{ij})$ denote the real part and the imaginary part of M_{ij}. Then the Gaussian Hermitian matrix integral of an arbitrary function f is defined as

$$< f > \quad = \quad \frac{\int e^{-N \,\mathrm{Tr}(\frac{M^2}{2})} f(M) dM}{\int e^{-N \,\mathrm{Tr}(\frac{M^2}{2})} dM},$$

where the integration is over all $N \times N$ Hermitian matrices. Let $\mathcal{A}(r_3, \ldots, r_k)$ denote the set of all subsets of edges which have a decomposition into r_i cycles of length i $(i = 3, \ldots, k)$ and define a function $g_{r_3,\ldots}(M) = \sum_{c \in \mathcal{A}(r_3,\ldots,r_k)} \prod_{e \in c} M_e$.

Theorem 11.6.6. *For every z_i with $|z_i| \in (0, \varepsilon_i)$ with suitable $\varepsilon_i > 0$ we have*

$$\lim_{N\to\infty} N^{-2} \log \sum_{r_3,\ldots} < g_{r_3,\ldots}(M) > \prod_i \frac{(Nz_i)^{r_i}}{r_i!} \quad = \quad \sum_{r_3,\ldots} \sum_{[\Gamma]^*} \prod_i \frac{z_i^{r_i}}{r_i!},$$

where Γ is a 2-vertex-connected 3-edge-connected planar fat graph with r_i vertices of degree i, $i \geq 0$, and $[\,\cdot\,]^*$ is the isomorphism equivalence class of Γ^*, the dual of Γ.

Note that the coefficient of $\prod_i z_i^{r_i}/r_i!$ in the right hand side of the power series in Theorem 11.6.7 is the same as the number of unlabeled planar graphs with r_i faces of length i, whose dual is 2-connected and 3-edge-connected.

Part IV. Random graph processes. In Part IV, from Chapter 12 and Chapter 13, we study random graph processes with degree constraints, such as the minimum degree random multigraph process and the min-min random graph process.

In Chapter 12 we study the minimum degree random multigraph process. This graph process makes the minimum degree of a graph generated by the process grow very fast, and makes components merge together quickly, creating the giant component quicker than in the classical random graph model. Furthermore, each of the possible small components outside the giant component is in some sense forced to be attached to the giant component, resulting in a single connected component with relatively small number of edges $((1 + o(1))1.73n)$, compared to the number of edges required in the classical random graph model $((1 + o(1))n \log n/2)$.

Using Wormald's differential equation method (Lemma 3.4.5) we determine the number of vertices of small degrees in the graph $G_{\min}(n, M)$ generated by the minimum degree random multigraph process at time M. We investigate also the component structure and the connectedness of $G_{\min}(n, M)$.

Furthermore, we show that the minimum degree random multigraph process enjoys the *phase transition* around a constant $h_{\mathrm{cr}} \doteq 0.8607$, using a multi-type branching process and the singularity analysis of probability generating functions. To study the phase transition of the classical random graph process $G(n, M)$ it is common to approximate the random graph process by a usual branching process (with a single type): Starting with a single vertex v we create the component containing v by exposing the neighbourhood of v first and then exposing the neighbourhood of a neighbour of v, and so on. This is possible, since each edge is present *independently* in $G(n, M)$. However, in $G_{\min}(n, M)$ it is not the case. To overcome this problem we distinguish the types of vertices in $G_{\min}(n, M)$ and approximate the process using a multi-type branching process.

Theorem 12.1.1. *Let $X_k(M)$ be the number of vertices of degree k in the minimum degree process $G_{\min}(n, M)$ and $H_k := \min\{M : \delta(G_{\min}(n, M)) \geq k\}$. Then a.a.s. $G_{\min}(n, M)$ is such that for every $M \geq 0$ and for $k = 0, 1, 2$,*

$$|X_k(M) - n\alpha_k(M/n)| = o(n)\,,$$

and for $k = 1, 2, 3$,

$$H_k = n\, h_k + o(n)\,,$$

where $\alpha_k(t)$*'s are deterministically given functions and* h_k*'s are constants, in particular* $h_1 \doteq 0.6931$, $h_2 \doteq 1.2197$ *and* $h_3 \doteq 1.7316$.

Theorem 12.3.1. *Let* $\delta > 0$. *Then with probability* $1 - O(1/n)$ *each component of* $G_{\min}(n, (1+\delta)n)$ *smaller than* $n/2$ *has at most* $(2/\delta)\log n$ *vertices and contains at most one cycle. Moreover, for every function* $\omega = \omega(n) \to \infty$, *a.a.s. the number of all vertices contained in unicyclic components of* $G_{\min}(n, (1+\delta)n)$ *is smaller than* ω.

Theorem 12.4.1. *Let* $\rho_n(t)$ *denote the probability that* $G_{\min}(n, tn)$ *is connected. Then, for every constant* $t \neq h_2$, *the limit*

$$\rho(t) = \lim_{n\to\infty} \rho_n(t)$$

exists and $\rho(t) = 0$ *for* $t < h_2$ *while* $\rho(t) = 1$ *for* $t \geq h_3$. *If* $t \in (h_2, h_3)$, *then* $0 < \rho(t) < 1$, *where*

$$\rho^+ = \lim_{t\to h_2+} \rho(t) > 0 \quad \text{and} \quad \lim_{t\to h_3-} \rho(t) = 1\,.$$

Theorem 12.7.1. *Let*

$$h_{\rm cr} = \log\left(\frac{16\log 2 - 2}{3\log 2 - 1 + \log 2\,\sqrt{27 - 16\log 2}}\right) \doteq 0.8607.$$

(1) If $t < h_{\rm cr}$, *then* a.a.s. *every component in* $G_{\min}(n, tn)$ *has* $O(\log n)$ *vertices.*

(2) If $t = h_{\rm cr}$, *and* $\omega(n) \to \infty$, *then* $G_{\min}(n, tn)$ a.a.s. *contains no component of order greater than* $n^{2/3}\omega(n)$, *and at least one component of order greater than* $n^{2/3}/\omega(n)$.

(3) If $t > h_{\rm cr}$, *then* a.a.s. *the largest component in* $G_{\min}(n, tn)$ *has* $\Theta(n)$ *vertices and every other component, if any, has* $O(\log n)$ *vertices.*

In Chapter 13 we study the min-min random graph process. A uniformly distributed random graph with degree constraints is a natural model to study, however there is no obvious way to define the corresponding graph process model. The min-min random graph process is relevant to a random graph with a given degree sequence, though a graph generated by this process may not be uniformly distributed.

We study the connectedness of the graph G_M generated by the min-min random graph process at time M. Using Fourier transforms and the singularity analysis of probability generating functions we investigate the order of the *giant component* of G_M. For that, let $X = X(M)$ be the number of vertices *outside* of the giant component of G_M. Then we can precisely determine the limiting distribution of X as $n \to \infty$. Finally, we investigate the distribution of tX if $t = t(n)$ tends to 0 as $n \to \infty$.

Theorem 13.4.1. *Let* $M = n + tn$.

(1) If $M \leq n$, *then* a.a.s. G_M *is disconnected.*

(2) Suppose that $0 < t < 1/2$ remains fixed as $n \to \infty$. Then G_M a.a.s. *has a largest component consisting of at least $\frac{n}{2}$ vertices. Furthermore, the number Y of components of order at most $\frac{n}{2}$ is asymptotically Poisson with mean $\mu_t = \frac{1}{2}(-(1-2t)^2 - \ln(4(t-t^2)))$. That is,*

$$\lim_{n\to\infty} \Pr[Y = k] = \mu_t^k \exp(-\mu_t)/k! \quad \text{for any } k.$$

In particular, letting

$$\Psi(t) = \exp(-\mu_t) = 2\sqrt{t-t^2}\ \exp((2t-1)^2/2),$$

we have $\lim_{n\to\infty} \Pr(G_M \text{ is connected}) = \lim_{n\to\infty} \Pr(Y = 0) = \Psi(t)$.

(3) If $t \geq 1/2$, then a.a.s. *G_M is connected.*

Theorem 13.5.1. *Let $M = n + tn$ for a constant $0 < t < 1/2$. Then as $n \to \infty$, X converges in distribution to the distribution given by the probability generating function*

$$q(z) = \sum_{l=0}^{\infty} q_l z^l = \exp\left[\frac{1}{2}(1-2t)^2(1-z^2)\right]\sqrt{\frac{1-(1-2t)^2}{1-(1-2t)^2z^2}}\ .$$

As a consequence, for any positive integer l,

$$\begin{aligned}
&\lim_{n\to\infty} \Pr(X = 2l) = q_{2l} \\
&= 2\sqrt{t(1-t)}\exp\left[\frac{(1-2t)^2}{2}\right](1-2t)^{2l}\sum_{0\leq m\leq l}\binom{2m}{m}\frac{(-1)^{l-m}}{2^{l+m}(l-m)!} \\
&= (1+O(1/l))\ 2\exp(2t(t-1))\sqrt{\frac{t(1-t)}{\pi l}}(1-2t)^{2l},
\end{aligned}$$

while $\lim_{n\to\infty} \Pr(X = 2l-1) = q_{2l-1} = 0$. *Furthermore,* a.a.s. *all components on at most $\frac{n}{2}$ vertices are cycles of even lengths.*

Theorem 13.5.2. *Suppose that $M = n + tn$, where $t = t(n)$ with $n^{-1}\ln^4 n \leq t = o(1)$. Then tX converges in distribution to Gamma distribution with both shape and scale parameter equal to $\frac{1}{2}$. That is,*

$$\lim_{n\to\infty} \Pr(tX \leq b) = \frac{1}{\sqrt{\pi}}\int_0^{2b} \frac{\exp(-s)}{\sqrt{s}}ds.$$

As a consequence, for any $x > 0$,

$$\begin{aligned}
\lim_{n\to\infty} \Pr(tX \geq x) &= \frac{\exp(-2x)}{\sqrt{2\pi x}}\sum_{m=0}^{\infty}(-1)^m(2m-1)!!2^{-2m}x^{-m} \\
&= (1+O(1/x))\ (2\pi x)^{-1/2}\exp(-2x),
\end{aligned}$$

where $(2m-1)!! = \prod_{i=1}^{m}(2m-2i+1)$.

Acknowledgement. The main results of this thesis are from joint work with Manuel Bodirsky, Amin Coja-Oghlan, Éric Fusy, Omer Giménez, Clemens Gröpl, Youngmee Koh, Martin Loebl, Mike Löffler, Tomasz Łuczak, Colin McDiarmid, Marc Noy, Sangwook Ree, Taral Guldahl Seierstad and Stefan Vigerske: [24, 21] (Chapter 5), [30] (Chapter 6), [22] (Chapter 7), [26] (Chapter 9), [25] (Chapter 10), [85] (Chapter 11), [88, 86] (Chapter 12), [46] (Chapter 13). Chapters 2, 4, and 8 are based mainly on the books [63, 76], and Chapter 3 on the books [33, 82].

I would like to thank my coauthors and my other colleagues in the research group "Algorithms and Complexity" of Humboldt University Berlin and the research group "Algorithms, Structure, Randomness" of Humboldt University Berlin, Technical University Berlin and Zuse-Institute Berlin for stimulating and fruitful discussions. I have been supported by the Deutsche Forschungsgemeinschaft through the European Graduate Program "Combinatorics, Geometry, and Computation", and the research group "Algorithms, Structure, Randomness".

Chapter 2

Planar Structures

In this chapter we recall some known facts on planar graphs. First, we briefly survey the current state of research on planar structures from the viewpoint of enumeration and uniform sampling (Section 2.1). Next, we study the enumeration methods for labeled and unlabeled planar structures. In particular we review the recursive counting methods through decomposition, the exponential generating functions for labeled structures (Section 2.2), the cycle indices and the ordinary generating functions for unlabeled structures as well as Bernside's lemma, Pólya's theorems (Section 2.3), and the singularity analysis to determine the asymptotics (Section 2.4). Finally, we discuss the techniques for uniform sampling, such as Markov chain Monte Carlo method, Boltzmann sampler and the recursive method (Section 2.5).

2.1 Planar graphs and subclasses

We know fairly well about labeled planar graphs from the viewpoint of the enumeration, the uniform sampling and typical properties, due to recent research on this field, some of which we list below.

Denise, Vasconcellos, and Welsh [48] were the first to investigate the number of labeled planar graphs and typical properties of a random planar graph, i.e., a graph that is chosen uniformly at random among all the labeled planar graphs on n vertices. They proved that the number of labeled planar graphs on n vertices is at most $75.8^{n+o(n)}n!$ for large n, based on a Markov chain whose stationary distribution is the uniform distribution on all labeled planar graphs. They showed also that a random planar graph is connected with asymptotic probability bounded away from zero, and that the expected number of edges is at least $3n/2$.

Further investigating this Markov chain, Gerke and McDiarmid [69] proved that the limit of the expected edge density μ of a random planar graph is at least $13/6 \doteq 1.86$, and McDiarmid, Steger, and Welsh [99] proved that the quantity $(g(n)/n!)^{1/n}$, where $g(n)$ denotes the number of labeled planar graphs on n vertices, converges to a limit γ, which is called the *growth constant*, as $n \to \infty$ and that the asymptotic fraction p_{con} of labeled *connected* planar graphs is between $1/e$ and 1. Gerke, McDiarmid, Steger, and Weißl [70] proved that the quantity $(g(n, qn)/n!)^{1/n}$, where $g(n, qn)$ denotes the number of labeled planar

graphs on n vertices with qn edges for $1 < q < 3$, converges to a limit γ_q as $n \to \infty$.

Using the triangulations and probabilistic methods, Osthus, Prömel, and Taraz [112] improved the upper bound of $g(n)$ to $37.3^{n+o(n)}n!$. Bender, Gao and Wormald [13] showed that the number of labeled 2-connected planar graphs on n vertices is asymptotically $26.1^{n+o(n)}n!$, which is a lower bound of $g(n)$, using the singularity analysis of generating functions arising from the decomposition of graphs along connectivity. Further analysing the singularities and singular types of the generating functions in [13], Giménez and Noy [71] determined the asymptotic number of labeled planar graphs on n vertices to be $c\, n^{-7/2}\, 27.2^n\, n!$ for some constant $c > 0$, the asymptotic fraction of connected graphs to be $p_{\mathrm{con}} \doteq 0.963$, and the limit of the expected edge density to be $\mu \doteq 2.21$.

Bodirsky, Gröpl, Johannsen, and Kang [28] determined the asymptotic number of labeled *3-connected* planar graphs to be $21.049^{n+o(n)}n!$, which was originally derived by Bender and Richmond [16]. In Chapter 6 (see [30]) we determine the asymptotic number of labeled cubic planar graphs on n vertices to be $c\, n^{-7/2}\, 3.13^n\, n!$ for some constant $c > 0$, and show that the chromatic number of a random connected cubic planar graph is asymptotically almost surely three. Furthermore, in Chapter 5 ([24, 31]) we show that the number of labeled outerplanar graphs on n vertices is asymptotically $c\, n^{-5/2}\, 7.32^n\, n!$ for some constant $c > 0$, and a random outerplanar graph has $1.56n$ edges in average and is connected with probability tending to 0.86.

For the uniform sampling of *labeled* planar graphs, the Markov chain suggested by Denise, Vasconcellos, and Welsh [48] is a good candidate. However, its mixing time is unknown, and even if it were rapidly mixing, it would only *approximate* the uniform distribution. In Chapter 7 ([22, 29]) we develop the *first* polynomial time algorithm to sample labeled planar graphs uniformly at random, based on the recursive method. In Chapters 5 ([21]) and 6 ([30]) we also design the first polynomial time algorithms for sampling labeled outerplanar graphs and labeled cubic planar graphs. The best known sampling algorithm for labeled planar graphs is due to Fusy [66], who developed a Boltzmann sampler that samples a labeled planar graph of an approximate size in linear time and an exact size in quadratic time. Bodirsky, Gröpl, Johannsen, and Kang [28] presented a polynomial time algorithm for sampling labeled 3-connected planar graphs.

Unlabeled planar structures, i.e., isomorphism classes of planar graphs, is considered more difficult to study than the labeled ones, due to symmetry. Though the picture of unlabeled planar structures is far from being completeted, there are recent progress on subclasses of them.

Otter [113] first applied the cycle indices introduced by Pólya to estimate the asymptotic number of unlabeled trees, i.e., unlabeled acyclic connected graphs. In Chapter 9 ([26]) we make a new step toward the enumeration of unlabeled planar graphs. Using cycle indices we show that the number of unlabeled outerplanar graphs on n vertices is asymptotically $c\, n^{-5/2}\, 7.5^n$ for some constant $c > 0$, and that a random unlabeled outerplanar graph has $1.54n$ edges in average and is connected with probability tending to 0.84.

Most of the known uniform sampling algorithms for *unlabeled* planar structures use the recursive method. In Chapter 9 ([21]) we design an expected polynomial time algorithm to sample unlabeled connected outerplanar graphs, and in Chapter 10 ([23]) to sample unlabeled 2-connected planar graphs. Bodirsky,

Gröpl, and Kang [25] developed an expected polynomial time algorithm to sample unlabeled connected cubic planar graphs. Instead of relying on the recursive method, which requires sometimes expensive pre-computation time, Bodirsky, Fusy, Kang, and Vigerske [27] coined a new concept of cycle-pointing to count and sample unlabeled structures including trees, outerplanar graphs, cactus graphs, and planar maps in an unbiased way, which yields generating functions for a Boltzmann sampler that runs in linear time.

While the combinatorial methods for enumeration employ the generating functions, a technique of theoretical physics called Wick's theorem interprets the Gaussian matrix integral of the product of the trace of powers of Hermitian matrices as the number of maps with a given degree sequence [51], which yields the map enumeration results analogous to those obtained by combinatorial methods [14]. In Chapter 11 ([85]) we show that the enumeration of the graphs embeddable on a given 2-dimensional surface, say the plane, can also be formulated as a Gaussian matrix integral.

2.2 Enumeration of labeled planar structures

To determine the number of labeled (not necessarily planar) graphs we can employ two methods. One is *recursive counting*, based on the decomposition of graphs into smaller graphs. The other is to use an *exponential generating function*, which is a formal power series whose coefficients are the counting sequences of the number of graphs.

Recursive counting. Let $g(n)$, $c(n)$, and $b(n)$ denote the number of labeled graphs, the number of labeled connected graphs, and the number of labeled 2-connected graphs on vertex set $[n] := \{1, 2, \cdots, n\}$ for a nonnegative integer n, respectively.

As a convention we let $g(0) = 1$ and $c(0) = 0$. Further we let $g_c(n)$ be the number of labeled graphs consisting of c connected components. Clearly $g_1(n) = c(n)$. We start with a very simple observation that a graph consists of at least one labeled connected graph but at most n labeled connected graphs. Thus $g_c(n) = 0$ for $c = 0$ or $c > n \geq 1$ and $g(n) = \sum_{c=1}^{n} g_c(n)$ for $n \geq 1$. We observe also that a labeled graph with one distinguished vertex (e.g., the vertex 1) as a root can be decomposed into the *split-off-graph* that is the labeled connected graph containing the second smallest vertex, and the labeled graph with the remaining parts. Summing over all the possible number of vertices in the split-off-graph, which is counted by $g_1(i) = c(i)$, we obtain that $g_c(n) = \sum_{i=1}^{n-c+1} \binom{n-1}{i-1} g_1(i) g_{c-1}(n-i)$ for $2 \leq c \leq n$. Putting these altogether we obtain

$$g(n) = \begin{cases} 1 & \text{for } n = 0 \\ \sum_{c=1}^{n} g_c(n) & \text{for } n \geq 1\,, \end{cases} \tag{2.2.1}$$

$$g_c(n) = \begin{cases} 0 & \text{for } n = 0,\ c = 0 \text{ or } c > n \geq 1 \\ c(n) & \text{for } c = 1 \leq n \\ \sum_{i=1}^{n-c+1} \binom{n-1}{i-1} g_1(i) g_{c-1}(n-i) & \text{for } 2 \leq c \leq n\,. \end{cases} \tag{2.2.2}$$

These formulas will later be used for labeled forests, labeled outerplanar graphs, and labeled planar graphs. For each of such graphs, we will derive appropriate

recursive counting formulas for $c(n)$, which may involve other counting formulas for 2-connected graphs and 3-connected graphs if necessary.

To compute these numbers efficiently, we use a well-known technique *dynamic programming*, i.e., we store all of their values in a table to avoid recomputation. Note that the number of labeled planar structures that we will consider, say g_n, is known to be "small", in the sense that the quantity $(g_n/n!)^{1/n}$ converges to a limit γ (see [99]). Thus the values in the table can be stored with $O(n \log n)$ many bits. Assuming an $O(n \log n \log\log n)$ multiplication algorithm (see e.g., [44]), the number of computation steps needed to fill the table is in $\tilde{O}(\text{poly}(n))$.

Exponential generating functions. Let $G(x), C(x)$, and $B(x)$ be the exponential generating functions corresponding to the numbers of labeled graphs, labeled connected graphs, and labeled 2-connected graphs, defined by

$$G(x) := \sum_{n\geq 0} g_n \frac{x^n}{n!}\,, \quad C(x) := \sum_{n\geq 0} c_n \frac{x^n}{n!}\,, \quad B(x) := \sum_{n\geq 0} b_n \frac{x^n}{n!}\,.$$

Let $C'(x) = dC(x)/dx$ and $B'(x) = dB(x)/dx$ denote the formal derivatives. Between these generating functions the following relation, due to the block decomposition, holds (see [63, 76, 144]).

Proposition 2.2.1. *Let $G(x), C(x), B(x)$ be as above. Then*

$$G(x) = \exp(C(x))\,, \tag{2.2.3}$$

$$xC'(x) = x\exp(B'(xC'(x)))\,. \tag{2.2.4}$$

The best way to see these relations is through a proof, which we sketch below (see [76] for the details).

Proof. (Sketch) The first relation (2.2.3) holds because a graph is a collection of connected graphs.

The idea for (2.2.4) is to use the *block-decomposition* of a graph. The *blocks of a graph* are either the maximal 2-connected subgraphs, the edges of the graph that are not contained in such a subgraph, or isolated vertices. The blocks and cut-vertices of a graph form a forest on two types of vertices: the blocks and the cutvertices of the graph (see Figure 2.1). A block and a cutvertex are said to be adjacent in the forest if the block contains the vertex.

We consider a *rooted graph*, which is a graph with one distinguished vertex as a root. Since there are n choices to select the root, the numbers of labeled rooted connected and 2-connected graphs on n vertices are nc_n and nb_n, respectively, and hence their exponential generating functions are $xC'(x)$ and $xB'(x)$. By using an auxiliary counting formula depending on the number of blocks incident to a root and a counting formula which enumerates the possible ways of composing a block and rooted connected graphs in such a way that these rooted connected graphs are incident to non-root vertices of the block, one can derive (2.2.4). □

These identities hold for arbitrary classes of labeled graphs, in particular, labeled outerplanar graphs, labeled cubic planar graphs, and labeled planar graphs. To complete the relation between exponential generating functions we

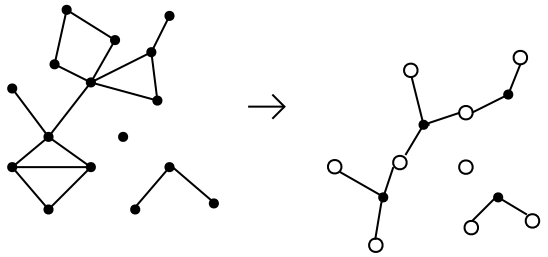


Figure 2.1: The block structure of a graph.

may further need counting formulas for 2-connected graphs and 3-connected graphs. Having complete set of such identities we can determine the exact numbers using Taylor series expansions at $x = 0$. To determine the asymptotic number we think of the generating functions as complex valued functions and apply singularity analysis presented in Section 2.4.

2.3 Enumeration of unlabeled planar structures

In the enumeration of *unlabeled* graphs cycle index sums introduced by Pólya and Burnside's lemma play essential roles [76, 118]. To determine the number of unlabeled graphs, the problem is reformulated, so that the answer can be obtained by finding the number of orbits of the appropriate permutation group. Burnside's lemma can then be used to express the number of orbits in terms of the number of objects fixed by permutations in the group. Pólya's enumeration theorem incorporates Burnside's lemma in terms of an appropriate cycle index and a polynomial called figure counting series. The results in this section are based on the book by Harary and Palmer [76].

Cycle index of a permutation group. Let A be a group of permutations on object set $X = \{1, \cdots, n\}$. Note that each permutation $\sigma \in A$ can be written uniquely as a product of disjoint cycles. For each integer k from 1 to n let $i_k(\sigma)$ denote the number of cycles of length k in the disjoint cycle decomposition of σ. The cycle index $Z(A)$ of A is a polynomial in the formal variables $s_1, \cdots, s_n$ defined by

$$Z(A) := Z(A; s_1, \cdots, s_n) := \frac{1}{|A|} \sum_{\sigma \in A} \prod_{k=1}^{n} s_k^{i_k(\sigma)}.$$

For example the cycle index of all the symmetric group is

$$\sum_{n \geq 0} Z(S_n) = \exp\left(\sum_{k \geq 1} \frac{s_k}{k} \right). \tag{2.3.1}$$

For convenience we take $Z(S_0) = 1$.

Burnside's lemma. Before stating Burnside's lemma let us recall facts on a permutation group. Let A be a permutation group on object set $X = \{1, 2, \cdots, n\}$. We say that x and y in X are *similar* if there is a permutation $\sigma \in A$ such that $\sigma x = y$. This is an equivalent relation and the equivalent classes are called the *orbits* of A. For each $x \in X$ the set $A(x) = \{\sigma \in A | \sigma x = x\}$ is called a *stabilizer* of x. If x and y in X are similar, then $|A(x)| = |A(y)|$. Further for any element y of an orbit Y of A, $|A| = |A(y)||Y|$.

Lemma 2.3.1 (Burnside's lemma)**.** *The number $N(A)$ of orbits of A satisfies*

$$N(A) = \frac{1}{|A|} \sum_{\sigma \in A} i_1(\sigma).$$

Consider the graph G in Figure 2.2 and denote by $\Gamma(G)$ its automorphism group. Then $\Gamma(G)$ consists of the following four permutations

$$\begin{aligned} \sigma_1 &= (1)(2)(3)(4)(5)(6) \\ \sigma_2 &= (1)(23)(4)(5)(6) \\ \sigma_3 &= (1)(2)(3)(4)(56) \\ \sigma_4 &= (1)(23)(4)(56), \end{aligned}$$

and $i_1(\sigma_1) = 6$, $i_1(\sigma_2) = 4$, $i_1(\sigma_3) = 4$, and $i_1(\sigma_4) = 2$. Thus

$$\frac{1}{|\Gamma(G)|} \sum_{\sigma \in \Gamma(G)} i_1(\sigma) = \frac{1}{4}(6 + 4 + 4 + 2) = 4.$$

Obviously there are four orbits of $\Gamma(G)$: $\{1\}$, $\{2, 3\}$,$\{4\}$, and $\{5, 6\}$.

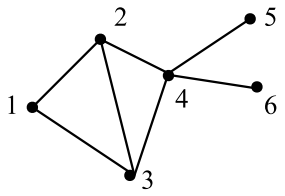


Figure 2.2: A graph with two fixed points and four orbits.

We may sometimes restrict A to a subset Y of X, where Y is a union of orbits of A. We denote by $A|Y$ the set of permutations on Y obtained by restricting those of A to Y. For each $\sigma \in A$, we denote by $i_1(\sigma|Y)$ the number of elements in Y fixed by σ. Then we obtain a restricted form of Burnside's lemma saying that the number $N(A|Y)$ of orbits of A restricted to Y satisfies

$$N(A|Y) = \frac{1}{|A|} \sum_{\sigma \in A} i_1(\sigma|Y). \tag{2.3.2}$$

Pólya's theorems. Let A be a permutation group on object set $X = [n]$ and let I be an identity group on a countable object set Y with at least two elements. The power group I^A is the collection Y^X of functions from X into Y as its object sets. The permutations of I^A consist of all ordered pairs (σ, id) of $\sigma \in A$ and $\mathrm{id} \in I$. The image of any function f in Y^X under (σ, id) is given by $(\sigma, \mathrm{id})f(x) = f(\sigma x)$ for each $x \in X$, considering that I^A acts on Y^X. Let $\omega : Y \to \{0, 1, \cdots, \}$ be a weight function such that $\omega^{-1}(k) < \infty$ for all $k = 0, 1, \cdots$ and let $c_k = |\omega^{-1}(k)|$ be the number of figures with weight k. The formal power seires in the variable x, defined by

$$c(x) = \sum_{k \geq 0} c_k x^k,$$

enumerates the elements of Y by weight and is called *figure counting series.*

The weight of a function in Y^X is defined by

$$\omega(f) = \sum_{x \in X} \omega(f(x)). \tag{2.3.3}$$

Thus functions in the same orbit of the power group I^A have the same weight. Let C_k be the number of orbits of weight k. The formal power series $C(x) = \sum_{k \geq 0} C_k x^k$ is called the *configuration counting series* or the *ordinary generating function* with counting sequence $\{C_k\}_k$. The following Pólya's theorem expresses $C(x)$ in terms of $Z(A)$ and $c(x)$.

Theorem 2.3.1. *The configuration counting series is obtained by replacing each variable s_k in $Z(A)$ by the figure counting series $c(x^k)$, which we denote by*

$$C(x) = Z(A, c(x)) := Z(A; c(x), c(x^2), \cdots, c(x^n)).$$

For illustration let us count the number of unlabeled pentagon whose vertices are coloured either red or blue. Let $X = \{1, 2, \cdots, 5\}$ and $Y = \{\mathrm{red}, \mathrm{blue}\}$. Each function f from X to Y corresponds to a labeled pentagon with coloured vertices where the vertex labeled with x has colour $f(x)$. Thus the pentagon represented by f has $f^{-1}(\mathrm{red})$ vertices coloured red and $f^{-1}(\mathrm{blue})$ vertices coloured blue. We now consider the identity group I acting on Y. To determine the number of unlabeled pentagons whose vertices are coloured either red or blue we should identify the pentagons when one differs from the other only by a rotation or reflection of the pentagon, that is, we should equip the pentagon with the dihedral group of degree 5, denoted by D_5. To remove the labels we should identify two labeled pentagons with coloured vertices whenever their corresponding functions are in the same orbit of I^{D_5}. We define the weight function $\omega : Y \to \{0, 1\}$ by $\omega(\mathrm{red}) = 0$ and $\omega(\mathrm{blue}) = 1$. Then $1 + x$ is the figure counting series for Y and a function of weight k represents a pentagon with $5 - k$ red vertices and k blue vertices. Hence the configuration counting series $C(x) = \sum_{k \geq 0} C_k x^k$ counts the number of unlabeled pentagons, where the coefficient C_k is the number of unlabeled pentagons with k blue vertices. From Theorem 2.3.1 we have that

$$C(x) = Z(D_5, 1 + x).$$

But it is known that

$$Z(D_5) = \frac{s_1^5 + 4s_5}{10} + \frac{s_1 s_2^2}{2},$$

and therefore the ordinary generating function for the counting sequence of the number of unlabeled pentagon whose vertices are coloured either red or blue is

$$\begin{aligned} C(x) = Z(D_5, 1+x) &= \frac{(1+x)^5 + 4(1+x^5)}{10} + \frac{(1+x)(1+x^2)^2}{2} \\ &= 1 + x + 2x^2 + 2x^3 + x^4 + x^5, \end{aligned}$$

as we can see in Figure 2.3.

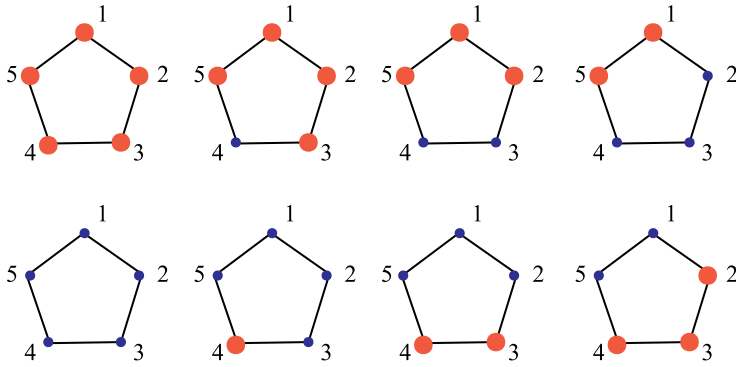


Figure 2.3: Pentagon coloured with two colours.

Next let us consider the composition of two permutation groups. Let A and B be permutation groups with objects sets $X = \{x_1, x_2, \cdots, x_n\}$ and $Y = \{y_1, y_2, \cdots, y_m\}$. The composition of A with B, denoted by $A[B]$, has object set $X \times Y$ and is defined as follows. For each $\sigma \in A$ and each sequence $\beta_1, \beta_2, \cdots, \beta_k$ of k permutations in B, there is a permutation in $A[B]$, denoted by $[\sigma; \beta_1, \beta_2, \cdots, \beta_k]$, such that for every ordered pair $(x_i, y_j) \in X \times Y$,

$$[\sigma; \beta_1, \beta_2, \cdots, \beta_k](x_i, y_j) = (\sigma x_i, \beta_j y_j).$$

The following Pólya's composition theorem shows that the cycle index of the composition of A with B is obtained by substituting the cycle index of B into the cycle index of A.

Theorem 2.3.2. *The cycle index $Z(A[B])$ of the composition of A with B is the polynomial obtained from $Z(A)$ by replacing each variable s_k in $Z(A)$ by $Z(B; s_k, s_{2k}, s_{3k}, \cdots)$, which is denoted by $Z(A)[Z(B)]$.*

Cycle index for a graph. For a graph G on n vertices with the automorphism group $\Gamma(G)$, we write $Z(G) := Z(\Gamma(G))$, and for a set of graphs $\mathcal{C}$, we write $Z(\mathcal{C})$ for the *cycle index sum* for $\mathcal{C}$ defined by

$$Z(\mathcal{C}) := Z(\mathcal{C}; s_1, \cdots, s_n) := \sum_{G \in \mathcal{C}} Z(G; s_1, \cdots, s_n). \tag{2.3.4}$$

As shown in [18], if $\bar{\mathcal{C}}$ is the set of graphs of $\mathcal{C}$ equipped with distinct labels, then

$$Z(\mathcal{C}) = \sum_{n \geq 0} \frac{1}{n!} \sum_{G \in \bar{\mathcal{C}}_n} \sum_{\sigma \in \Gamma(G)} \prod_{k=1}^{n} s_k^{i_k(\sigma)},$$

which coincides with (2.3.4) and shows the close relationship of cycle index sums to exponential generating functions in labeled counting.

The composition of graphs corresponds to the composition of the associated cycle indices. Consider an object set $X = \{1, \cdots, n\}$ and a permutation group A on X. A composition of n graphs from $\mathcal{C}$ is a function $f : X \to \mathcal{C}$. Two compositions f and g are similar, $f \sim g$, if there exists a permutation $\sigma \in A$ with $f \circ \sigma = g$.

Theorem 2.3.3. *We write $\mathcal{G}$ for the set of equivalence classes of compositions of n graphs from $\mathcal{C}$ (with respect to the equivalence relation $\sim$). Then*

$$Z(\mathcal{G}) = Z(A)[Z(\mathcal{C})] := Z(A; Z(\mathcal{C}; s_1, s_2, \cdots), Z(\mathcal{C}; s_2, s_4, \cdots), \cdots), \quad (2.3.5)$$

that is, $Z(\mathcal{G})$ is obtained from $Z(A)$ by replacing each s_i by

$$Z(\mathcal{C}; s_i, s_{2i}, \cdots) = \sum_{G \in \mathcal{C}} Z(G; s_i, s_{2i}, \cdots).$$

Hence, (2.3.5) makes it possible to derive the cycle index sum for a class of graphs by decomposing the graphs into simpler structures with known cycle index sum.

In many cases, such a decomposition is only possible when, for example, one vertex is distinguished from the others in the graphs. A graph with a distinguished vertex is called a *vertex-rooted* graph. The automorphism group of a vertex-rooted graph consists of all permutations of the group of the *unrooted* graph that fix the root vertex. Hence, one can expect a close relation between the cycle indices of unrooted graphs and the cycle indices of their rooted counterparts. As shown in [76], if $\mathcal{G}$ is an unlabeled set of graphs and $\hat{\mathcal{G}}$ is the set of graphs in $\mathcal{G}$ rooted at a vertex, then

$$Z(\hat{\mathcal{G}}) = s_1 \frac{\partial}{\partial s_1} Z(\mathcal{G}). \quad (2.3.6)$$

This relationship can be inverted to express the cycle index sum for the unrooted graphs in terms of the cycle index sum for the rooted graphs,

$$Z(\mathcal{G}) = \int_0^{s_1} \frac{1}{s_1} Z(\hat{\mathcal{G}}) ds_1 + Z(\mathcal{G})|_{s_1=0}. \quad (2.3.7)$$

Observe that permutations without fixed points are not counted by the cycle indices of the rooted graphs, so that their cycle indices are added as a boundary term to $Z(\mathcal{G})$.

Ordinary generating functions. Once the cycle index sum $Z(\mathcal{G})$ for a class $\mathcal{G}$ of graphs of interest is known, the corresponding ordinary generating function can be derived by replacing the formal variables s_i in the cycle index sums by x^i. For we know that for a graph G

$$Z(G; x, x^2, \cdots) = x^{|G|}. \quad (2.3.8)$$

Thus letting g_n be the number of graphs $G \in \mathcal{G}$ of given size n, the ordinary generating function for $\mathcal{G}$ defined by

$$G(x) := \sum_{n \leq 0} g_n x^n = \sum_{G \in \mathcal{G}} x^{|G|} \quad (2.3.9)$$

is obtained from the cycle index sum by

$$G(x) = \sum_{G \in \mathcal{G}} Z\left(G; x, x^2, \cdots\right). \tag{2.3.10}$$

More generally, for a group A and an ordinary generating function $f(x)$ we define

$$Z\left(A; f(x)\right) := Z\left(A; f(x), f(x^2), f(x^3), \cdots\right)$$

as the ordinary generating function obtained by substituting each s_i in $Z\left(A\right)$ by $f(x^i)$, $i \geq 1$.

Once we have ordinary generating functions for graphs we can compute the exact numbers using Taylor series expansions at $x = 0$ and the asymptotic number using singularity analysis.

2.4 Singularity analysis

To determine the asymptotics of the coefficients of a generating function we use singularity analysis. The results in this section are based on the book by Flajolet and Sedgewick [63].

The fundamental observation is that the exponential growth of the coefficients of a generating function $f(x)$ that is analytic at the origin is determined by the radius of convergence R of the coefficients of f, i.e.,

$$[x^n]f(x) = \theta(n)R^{-n}, \quad \text{with} \quad \limsup |\theta(n)|^{1/n} = 1,$$

where $[x^n]f(x)$ denotes f_n (resp. $f_n/n!$) for the ordinary generating function $f(x) = \sum_{n\geq 0} f_n x^n$ (resp. the exponential generating function $f(x) = \sum_{n\geq 0} f_n x^n/n!$). Such a function f necessarily has a singularity on the boundary of its disc of convergence $|x| = R$, which is called a *dominant singularity*. If in addition f has non-negative Taylor coefficients, then the point $x = R$ is a dominant singularity of f (Pringsheim's Theorem; see e.g., Theorem IV.3 in [63]).

It will turn out that a generating function $f(x)$ determinded implicitly by the equation of the type

$$f(x) = x\phi(f(x)) \tag{2.4.1}$$

for an appropriate generating function $\phi(u)$ plays an important role.

To determine the exact coefficient of the generating function $f(x)$ defined by (2.4.1), we may apply the following Lagrange Inversion Theorem (see Appendix A in [63]).

Theorem 2.4.1. *The coefficients of all the powers of $f(x)$ satisfy*

$$[x^n]f(x)^k = \frac{k}{n}[u^{n-k}]\phi(u)^n, \tag{2.4.2}$$

for any positive integer k.

Dominant singularity and asymptotics. A closer look at the type of the dominant singularity, for example, the order of the pole, enables us to compute subexponential factors as well. The following lemma describes the singular expansion for a common case [63, Thm. VI.1].

Lemma 2.4.2. *Let $f(x) = (1-x)^{-\sigma}$ with $\sigma \notin \{0, -1, -2, \ldots\}$. Then the coefficients $[x^n]f(x)$ have a full asymptotic development in descending powers of n,*

$$[x^n]f(x) = \binom{n+\sigma-1}{n} \sim \frac{n^{\sigma-1}}{\Gamma(\sigma)} \left(1 + \sum_{k=1}^{\infty} \frac{e_k(\sigma)}{n^k}\right)$$

where $\Gamma(\sigma)$ is the Gamma-Function, $\Gamma(\sigma) := \int_0^\infty e^{-t}t^{\sigma-1}dt$ where σ is a complex number with $\sigma \notin \{0, -1, -2, \ldots\}$, and $e_k(\sigma)$ is a polynomial in σ of degree $2k$.

Due to the rescaling rule

$$[x^n](1 - x/x_0)^{1/2} = x_0^{-n}[x^n](1-x)^{1/2}, \tag{2.4.3}$$

we can directly apply Lemma 2.4.2, once we have determined the singular expansion of a generating function near the dominant singularity.

To determine the asymptotic estimate of the coefficients of the generating function $f(x)$ defined by (2.4.1) we can apply the following lemma, which is a combination of Theorem IV.3, Theorem VI.1, Proposition VI.1, and Theorem VI.4 in [63]. A generating function ϕ is called *periodic*, if $\phi(u) = u^e h(u^d)$ for some power series h and some integer $d \geq 2$, and *aperiodic* otherwise.

Lemma 2.4.1. *Let $\phi(u)$ be analytic at the origin with non-negative Taylor coefficients satisfying $\phi(0) \neq 0$ and aperiodic, and let r be its radius of convergence. Then the functional inverse of $\psi(u) := u/\phi(u)$, say $f(x)$, exists and is analytic at the origin. Moreover, if there exists $u_0 \in (0, r)$ such that $\psi'(u_0) = 0$ and $\psi''(u_0) \neq 0$, then $f(x)$ has a dominant singularity at $x_0 = \psi(u_0) = u_0/\phi(u_0)$ and its coefficients satisfy*

$$[x^n]f(x) = (1 + O(n^{-1}))\frac{1}{\sqrt{2\pi}}\sqrt{\frac{\phi(u_0)}{\phi''(u_0)}}n^{-3/2}{x_0}^{-n}. \tag{2.4.4}$$

Proof. (Sketch) The Taylor expansion of $x = \psi(u)$ at $u = u_0$ (where $\psi'(u_0) = 0$) is of the form

$$\psi(u) = \psi(u_0) + \frac{1}{2}\psi''(u_0)(u - u_0)^2 + \cdots,$$

which implies a locally quadratic dependency between x and u:

$$(u - u_0)^2 \sim \frac{2}{\psi''(u_0)}(x - x_0) = -\frac{2}{\psi''(u_0)}(x_0 - x).$$

Since $\phi(u)$ is analytic at the origin with non-negative Taylor coefficients satisfying $\phi(0) \neq 0$, $f(x)$ has non-negative Taylor coefficients, and hence it is in particular increasing along the positive real axis. Therefore we obtain

$$f(x) - f(x_0) = u - u_0 \sim -\sqrt{-\frac{2}{\psi''(u_0)}}(x_0 - x)^{1/2} = -\sqrt{\frac{2\phi(u_0)}{\phi''(u_0)}}(1 - x/x_0)^{1/2}. \tag{2.4.5}$$

Lemma 2.4.2 implies that

$$[x^n](1-x)^{1/2} = (1+O(n^{-1}))\frac{n^{-3/2}}{\Gamma(-1/2)} = (1+O(n^{-1}))\frac{1}{-2\sqrt{\pi}}n^{-3/2}. \quad (2.4.6)$$

Using (2.4.5), (2.4.6), and the rescaling rule (2.4.3), we get (2.4.4). □

In some of our calculations, it will appear that a generating function $f(x)$ is given only implicitly by an equation $H(x, f(x)) = 0$, for a bivariate function $H(x, y)$. Theorem VII.3 in [63] describes how to derive a full singular expansion of $f(x)$ in this case. We state it here in a slightly modified version.

Theorem 2.4.3. *Let $H(x,y) := \sum_{n,m} h_{n,m}x^n y^m$ be a bivariate function that is analytic in a complex domain $|x| < R$, $|y| < S$ and satisfies $H(0,0) = 0$, $\frac{\partial}{\partial y}H(0,0) = -1$, and whose coefficients $h_{m,n}$ satisfy the following positivity conditions: They are nonnegative except for $h_{0,1} = -1$ (because $\frac{\partial}{\partial y}H(0,0) = -1$) and $h_{m,n} > 0$ for at least one pair (m,n) with $n \geq 2$. Assume that there are two numbers $r \in (0,R)$ and $s \in (0,S)$ such that*

$$H(r,s) = 0, \qquad \frac{\partial}{\partial y}H(r,s) = 0, \quad (2.4.7)$$

$\frac{\partial^2}{\partial y^2}H(r,s) \neq 0$ and $\frac{\partial}{\partial x}H(r,s) \neq 0$. Assume further that $H(x, Y(x)) = 0$ admits a solution $Y(x)$ that is analytic at 0, has non-negative coefficients, and is aperiodic. Then r is the unique dominant singularity of $Y(x)$ and $Y(x)$ converges at $x = r$, where it has the singular expansion

$$Y(x) = s + \sum_{i\geq 1} Y_i \left(1 - \frac{x}{r}\right)^{i/2}$$

with $Y_1 = -\sqrt{\frac{2r\frac{\partial}{\partial x}H(r,s)}{\frac{\partial^2}{\partial y^2}H(r,s)}} \neq 0$ and computable constants $Y_2, Y_3, \cdots$. Hence,

$$[x^n]\,Y(x) = -\frac{Y_1}{2\sqrt{\pi n^3}}r^{-n}\left(1 + O\left(\frac{1}{n}\right)\right).$$

The formulas that express the coefficients Y_i in terms of partial derivatives of $H(x,y)$ at (r,s) can be found in [117].

Asymptotic properties of random structures. When a parameter ξ of a combinatorial structure is studied, the generating function $f(x)$ has to be extended to a bivariate generating function $f(x,y) = \sum_{n,m} f_{n,m}x^n y^m$ where the second variable y marks the parameter ξ. We can determine the asymptotic distribution of ξ from $f(x,y)$ by varying y in some neighbourhood of 1. The following theorem follows from the so-called quasi-powers theorem [63, Thm. IX.7].

Theorem 2.4.4. *Let $f(x,y)$ be a bivariate generating function of a family of objects $\mathcal{F}$, where the power in y corresponds to a parameter ξ on $\mathcal{F}$, i.e.,*

$$[x^n y^m]\, f(x,y) = |\{f \in \mathcal{F} :\ |f| = n, \xi(f) = m\}|.$$

Assume that, in a fixed complex neighbourhood of $y = 1$, $f(x, y)$ has a singular expansion of the form

$$f(x,y) = \sum_{k \geq 0} f_k(y) \left(1 - \frac{x}{x_0(y)}\right)^{k/2} \tag{2.4.8}$$

where $x_0(y)$ is the dominant singularity of $x \mapsto f(x, y)$. Furthermore, assume that there is an odd $k_0 \in \mathbb{N}$ such that for all y in the neighbourhood of 1, $f_{k_0}(y) \neq 0$ and $f_k(y) = 0$ for $0 < k < k_0$ odd. Assume that $x_0(y)$ and $f_{k_0}(y)$ are analytic at $y = 1$, and that $x_0(y)$ satisfies the variance condition*,*

$$x_0''(1)\, x_0(1) + x_0'(1)\, x_0(1) - x_0'(1)^2 \neq 0.$$

Let X_n be the restriction of ξ onto all objects in $\mathcal{F}$ of size n. Under these conditions, the distribution of X_n is asymptotically Gaussian with mean $\mathbb{E}(X_n) \sim \mu n$ and variance $\mathrm{Var}(X_n) \sim \sigma^2 n$, where

$$\mu = -\frac{x_0'(1)}{x_0(1)}, \quad \sigma^2 = -\frac{x_0''(1)}{x_0(1)} - \frac{x_0'(1)}{x_0(1)} + \left(\frac{x_0'(1)}{x_0(1)}\right)^2.$$

2.5 Uniform sampling

There are several fields of applications of efficient algorithms that sample random combinatorial structures. We can use such a generation procedure as an experimental tool to investigate properties of combinatorial structures that hold almost always. It can also be used to produce test instances for other algorithms on these structures. We can then measure the average running time of these algorithms on random instances.

In this thesis we will focus on the generation of random planar structures according to their *uniform distribution*, i.e., for a planar structure $\mathcal{G}$ each object $G \in \mathcal{G}$ of size n is drawn with probability $\Pr(G) = \frac{1}{g_n}$, where g_n is the number of all the objects in $\mathcal{G}$ of size n.

The well known techniques for uniform sampling are Markov chain Monte Carlo method, Boltzmann sampler and the recursive method. All the uniform sampling algorithms for planar structures presented in this thesis use recursive method. Nevertheless we first discuss briefly Markov chain Monte Carlo method and Boltzmann sampler.

Markov chain Monte Carlo method. A Markov chain defined on combinatorial objects gives a Monte Carlo method of estimating the size of the set or generating a random element according to a given probability distribution. Its efficiency depends on how fast it converges to its stationary distribution, which is called *mixing time*. Several techniques based on spectral properties, group representations, conductance and couplings are developed to obtain the upper bound of a mixing time (see [43, 52, 53, 56, 84, 109] and references therein). They have been successfully applied for many applications, such as card shuffling [3, 52], approximate volume estimations [57, 89], random generations of spanning trees [2, 114], matching [132], Hamiltonian cycles and colourings [56, 84].

Boltzmann sampler. Boltzmann sampler introduced by Duchon, Flajolet, Louchard, and Schaeffer [55], is a uniform sampler that fits well with exponential (resp. ordinary) generating functions $G(x) = \sum_{n\geq 0} g_n x^n/n!$ (resp. $\widetilde{G}(x) = \sum_{n\geq 0} \widetilde{g}_n x^n$) for a labeled (resp. unlabeled) class of graphs $\mathcal{G}$ (resp. $\widetilde{\mathcal{G}}$). It draws each object $G_n \in \mathcal{G}$ (resp. $\widetilde{G}_n \in \widetilde{\mathcal{G}}$) of size n with probability

$$\Pr(G_n) = \frac{x^n}{n!G(x)}, \quad \text{resp. } \Pr(\widetilde{G}_n) = \frac{x^n}{\widetilde{G}(x)},$$

for a fixed real value x within the radius ρ of convergence of $G(x)$ (resp. $\widetilde{G}(x)$). To evaluate the probabilities the analytic expressions for generating functions are essential. A random object generated by a Boltzmann sampler has a fluctuating size, but objects with the same size are generated with the same probability, whereas other uniform samplers generate objects of a fixed size. Once a Boltzmann sampler can be derived for any fixed real value x with $0 < x < \rho$, we can *tune* the real parameter x so that given a target-size n and a tolerance ratio $\varepsilon > 0$, the random generator outputs objects of size in $[n(1-\varepsilon), n(1+\varepsilon)]$ with high probability. If the coefficients g_n of $G(x)$ have the asymptotic behaviour $g_n \sim c\rho^{-n}n^{-3/2}$ for some constant c, then $x = \rho$ is a good choice. This sampler outputs objects of $\mathcal{G}$ of size in $[n(1-\varepsilon), n(1+\varepsilon)]$, with uniform distribution on each size $k \in [n(1-\varepsilon), n(1+\varepsilon)]$, and with average complexity bounded by cn/ε for some constant c. Hence the complexity is *linear* in the size, once the tolerance ratio is fixed. If an exact size is required, the rejection process yields an uniform sampler for objects of size n with quadratic expected time. In summary a Boltzmann sampler generates an object with an approximate size in linear time and a fixed size in quadratic time.

Recursive method. The *recursive method*, which was introduced by Nijenhuis and Wilf [108] and further developed by Flajolet, Zimmerman, and Van Cutsem [64], provides a powerful technique for sampling procedure for any class of graphs that admits a recursive decomposition. In general the graphs are decomposed *uniquely* along their connectivity structure [141], producing a decomposition tree. Along the decomposition we derive recursive formulas. As a reverse procedure of decomposition a uniform sampling algorithm is obtained, where each generation procedure branches into subroutines with the right probabilities evaluated using the counting sequences involved in recursive counting formulas. This is illustrated in Section 4.4, particularly in Figure 4.2.

One advantage of this approach is that the sampled objects via this method are *exactly* uniformly distributed, while those via the Markov chain Montel Carlo method are *approximately* uniformly distributed. Second the counting formulas are interesting in their own right. Finally, the running time for the generation improves considerably if one allows precomputation. However, the huge precomputation times may be involved when we deal with a class of graphs which does not have simple decomposition procedure, in particular unlabeled graph classes since in this case we should decompose the graphs along the symmetry as well as the connectivity.

Chapter 3

Random Graphs

In this chapter we recall facts on classical random graphs, such as evolution, the phase transition, planarity, and connectedness (Section 3.1). Further we discuss several random graph processes with degree restrictions (Section 3.2). We also present basic inequalities and facts that we will need for Chapters 12 and 13, such as Chernoff bounds, the convergence of characteristic functions, Wormald's differential equation method, and multi-type branching processes (Section 3.4).

3.1 Classical random graph models

Random graphs were first introduced by Erdős and Rényi [59, 60]. Classical models include a uniform random graph, a binomial random graph, and a random graph process (see e.g., [33, 82]).

Let n, m, M be positive integers satisfying $0 \leq m, M \leq \binom{n}{2}$, and p a real number satisfying $0 < p < 1$.

A uniform random graph $G(n, m)$ is a probability space over the set of all labeled graphs G on vertex set $[n] := \{1, \ldots, n\}$ with m edges determined by

$$\Pr(G(n, m) = G) = \binom{\binom{n}{2}}{m}^{-1}.$$

The parameter m is usually a function $m = m(n)$ of n.

A binomial random graph $G(n, p)$ is a probability space over the set of all labeled graphs G on vertex set $[n]$ determined by

$$\Pr(G(n, p) = G) = p^{e_G}(1-p)^{\binom{n}{2}-e_G}$$

where e_G denotes the number of edges in G. It can be viewed that each of the pairs $\binom{n}{2}$ of vertices is adjacent independently with probability p. The edge probability p is taken not only as a constant but also as a function $p = p(n)$ of n.

Two models are asymptotically equivalent when the number of edges M of $G(n, m)$ is about the same as the expected number of edges of $G(n, p)$, namely, $m \sim \binom{n}{2}p$. It is often the case that a binomial random graph is easy to handle since the edge occurrences are independent from each other.

A random graph process $(\mathcal{G}_M)_{0\leq M\leq\binom{n}{2}}$ is an algorithmic version of the uniform random graph model $G(n,M)$, which can be seen as a stochastic process, or more precisely, a *Markov chain*, whose state space is the set of all graphs on n vertices: The initial state $\mathcal{G}_0$ is just an empty graph. Further, $\mathcal{G}_{M+1}$ is obtained from $\mathcal{G}_M$ by adding one edge chosen uniformly at random among the $\binom{n}{2}-M$ possible edges not present in $\mathcal{G}_M$. Then $G(n,M)$ occurs as the M-th state $\mathcal{G}_M$ of the Markov chain. Therefore, we can think of $G(n,M)$ as a "living organism" that evolves over time where the time parameter is $0\leq M\leq\binom{n}{2}$. This dynamical viewpoint of the random graph process makes it natural to study how the structural properties change as the graph evolves from empty to full.

Related to a binomial random graph $G(n,p)$, there is a continuous time random graph process $(\mathcal{G}(n,p))_{0\leq p\leq 1}$, where each potential edge is equipped with a random variable, called a *birthtime*, which is uniformly distributed over the interval $[0,1]$, and the edges appeared in $\mathcal{G}(n,p)$ are those whose birthtime is at most p. However, in the thesis we will only consider random graphs processes of type $(\mathcal{G}_M)_{M\geq 0}$.

Evolution. At the beginning of the random graph process the picture of evolution is quite simple: until $M\ll n$ a.a.s. $G(n,M)$ is a forest which contains copies of all trees of order $o(\log n)$. But at $M=\Theta(n)$ the evolution changes in a quite interesting way. Let $M=cn/2$ for a positive constant. Until $c<1$ a.a.s. $G(n,M)$ still contains only trees and unicyclic components of order $O(\log n)$, and the order of components increases smoothly as new edges are added. But at some point the largest components become so large that a new edge might connect two of them, resulting in changing the order of a new component significantly. This speeds up connecting two of such new components of similar order, and merging all the largest one into a unique largest component, called a *giant component*. In fact the giant component emerges at $M=n/2+O(n^{2/3})$, called the *critical phase*. When $M-n/2\ll -n^{2/3}$, called the *subcritical phase*, a.a.s. $G(n,M)$ has no component larger than $O(\log n)$ but when $M-n/2\gg n^{2/3}$, called the *supercritical phase*, the giant component of $G(n,M)$ is of order $\Theta(n)$.

As the random graph $G(n,M)$ evolves with more edges, the giant component grows by being connected by other components, from larger ones to smaller ones. When $M=\frac{n}{2}(\log n+O(1))$, at the moment when the last isolated vertex disappears, the graph becomes connected. At the same time a perfect matching can be found. As the random graph further evolves with more edges, the minimum degree and connectivity increase. When $M=\frac{n}{2}(\log n+(k-1)\log\log n+O(1))$, for fixed $k\geq 2$, the last vertex of degree $k-1$ disappears and at the the same time a.a.s. $G(n,M)$ becomes k-connected. Especially, when $G(n,M)$ becomes 2-connected, it has a Hamiltonian cycle and furthermore it contains cycles of all lengths. As a graph $G(n,M)$ becomes denser and denser, its diameter decreases and the largest complete subgraph grows and finally $G(n,M)$ becomes complete when M reaches $\binom{n}{2}$.

Phase transition. Erdős and Rényi [59, 60] determined the period of the random graph evolution when the order of the largest component of $G(n,M)$ rapidly grows from $\Theta(\log n)$ to $\Theta(n^{2/3})$ and then to $\Theta(n)$.

Theorem 3.1.1. *Let $M=cn/2$.*

(i) *If $0 < c < 1$, then the largest component of $G(n, M)$ has $O(\log n)$ vertices and all components are a.a.s. isolated trees or unicycles.*

(ii) *If $c = 1$, then the largest component has $\Theta(n^{2/3})$ vertices.*

(iii) *If $c > 1$, then there is a giant component on $\alpha n + o(n)$ vertices where $0 < \alpha < 1$ is the unique solution of the equation*

$$e^{-\alpha c} = 1 - \alpha,$$

and all other components are of order $O(\log n)$ and are still trees and unicycles.

Bollobás [35], Łuczak [96], Łuczak, Pittel, and Wierman [98], Janson, Knuth, Łuczak, and Pittel [81], and Janson [80] found that the giant component emerges at $M = n/2 + O(n^{2/3})$, and described very precisely the overall features of the phase transition of the random graph at the critical phase. In order to describe the internal structure of the graphs we measure the difference, called *excess*, between the number of edges and the number of vertices. An isolated tree has an excess -1 and a unicyclic component has an excess 0. All other component, said to be *complex*, have more edges than vertices, in other words, they contain at least two cycles.

Theorem 3.1.2. *Let $M = n/2 - s$, $s = s(n) \to \infty$.*

(i) *The probability that $G(n, M)$ contains a complex component is at least $\frac{n^2}{4s^3}$. It implies that for $s \gg n^{2/3}$ a.a.s. $G(n, M)$ contains no complex component and thus it is planar.*

(ii) *For $n^{2/3} \ll s \ll n$ and a fixed constant r a.a.s. the r-th largest components of $G(n, M)$ are all trees of order $(1/2 + o(1))\frac{n^2}{s^2} \log \frac{s^3}{n^2}$*

Theorem 3.1.3. *Let $M = n/2 + s$, $s = s(n) \to \infty$.*

(i) *For $n^{2/3} \ll s \ll n$ and a fixed constant r a.a.s. the r-th largest component among all trees and unicyclic components of $G(n, M)$ is a tree of order $(1/2 + o(1))\frac{n^2}{s^2} \log \frac{s^3}{n^2}$.*

(ii) *For $s \gg n^{2/3}$ a.a.s. $G(n, M)$ contains exactly one complex component.*

(iii) *For $n^{2/3} \ll s \ll n$ a.a.s. the largest component of $G(n, M)$ is of order $(4 + o(1))s$ and has $(16/3 + o(1))s^3/n^2$ excess.*

Theorem 3.1.3 (i) says that for $n^{2/3} \ll s \ll n$ the evolution of the graph $\hat{G}(n, n/2+s)$, which is obtained from $G(n, n/2+s)$ by deleting all vertices of the largest component, is quite the same as that of $G(n, n/2-s)$ in the reverse order in the sense of the order of components. As the random graph evolves the larger components which have appeared later in the subcritical phase are merged into the giant component faster than the smaller ones which have appeared earlier in the subcritical phase. More precisely speaking, for $n^{2/3} \ll s \ll n$ the number of vertices and the number of edges in the giant component of $G(n, n/2 + s)$ grow roughly four times as its time. Thus for $n' = n - (4+o(1))s = (1+o(1))n$ and $M' = n/2+s-(4+o(1))s = n'/2-(1+o(1))s$, $G(n', M')$ behaves roughly the same as $G(n, M)$, which is called the *symmetry rule*.

Theorem 3.1.3 (ii) and Theorem 3.1.3 (iii) say that in the supercritical phase the giant component is the only complex component and the total number of excess of complex components is a.a.s. $(16/3 + o(1))s^3/n^2$, which again implies that the number of complex components is bounded in probability and all are of order $\Theta(n^{2/3})$. Janson, Knuth, Łuczak, and Pittel [81] showed that the probability that $G(n, M)$ has never more than one complex component throughout its evolution tends to $5\pi/18 \approx 0.8727$ as $n \to \infty$.

Planarity. Until the subcritical phase $M = n/2 - s,\ s \gg n^{2/3}$, a.a.s. $G(n, M)$ contains no complex component and so it is planar. In the critical phase a topological copy of $K_{3,3}$ appears and thus $G(n, M)$ becomes non-planar. The following is proven by Erdős and Rényi [60] and later by Łuczak, Pittel, and Wierman [98], and Janson, Knuth, Łuczak, and Pittel [81].

Theorem 3.1.4. *If $M = n/2 + c_n n^{2/3}$, then*

$$\lim_{n\to\infty} \Pr(G(n, M) \text{ is planar }) = \begin{cases} 1 & \text{if } c_n \to -\infty \\ \rho(c) & \text{if } c_n \to c \\ 0 & \text{if } c_n \to \infty\,, \end{cases}$$

where $0 < \rho(c) < 1$ and in particular

$$0.987074 \le \rho(0) \le 0.999771.$$

Connectedness As the random graph further evolves with more edges, the number of vertices, which are not in the giant component, decreases exponentially in such a way that the components are swallowed up by the giant component. Erdős and Rényi [61], and Bollobás and Thomason [37, 36] proved that when $M = n(\log n + c_n)/2$ with $c_n \to \infty$, a.a.s. $G(n, M)$ is connected. Indeed, at this time the last isolated vertex vanishes and the graph contains a perfect matching.

Theorem 3.1.5. *If $M = n(\log n + c_n)/2$, then*

$$\lim_{n\to\infty} \Pr(G(n, M) \text{ is connected}) = \begin{cases} 0 & \text{if } c_n \to -\infty \\ 1 - e^{-e^{-c}} & \text{if } c_n \to c \\ 1 & \text{if } c_n \to \infty\,. \end{cases}$$

Bender, Canfield, and McKay [17] computed the asymptotic probability that $G(n, M)$ is connected for any ratio M/n using enumerative methods, and Behrisch, Coja-Oghlan and Kang [11, 12] for $M/n = 1+\Omega(1)$ using probabilistic approaches.

3.2 Random graphs with degree constraints

Among well-studied random graph models with degree constraints are a uniformly distributed random regular graph and a uniformly distributed random graph with a given degree sequence.

The random r-regular graph $G^{(r)}(M)$ is chosen uniformly at random among all r-regular graphs with vertex set $[n]$, where we assume that rn is even. In fact, the structure of $G^{(r)}(M)$ differs significantly from $G(n, M)$, $M = rn/2$. For instance, *a.a.s.* $G^{(r)}(M)$ is connected if $r \geq 3$, while $G(n, M)$ remains disconnected until the average degree $2M/n$ exceeds $(1 - o(1)) \log n$ (Theorem 3.1.5). Furthermore, $G^{(r)}(M)$ provides an example of a sparse graph with excellent expansion properties. See [147] for more details on $G^{(r)}(M)$.

The random graph with a given degree sequence is defined as follows. For a given integer valued sequence $\mathcal{D} = \{d_0(n), d_1(n), \ldots\}$ satisfying $d_i(n) = 0$ for $i \geq n$, and $\sum_{i \geq 0} d_i(n) = n$, called the *asymptotic degree sequence*, let $\mathcal{D}_n$ be the *degree sequence* $\{a_1, a_2, \ldots, a_n\}$, where $a_j \geq a_{j+1}$ for every $j = 1, \ldots, n-1$, and $|\{j | a_j = i\}| = d_i(n)$. The value $d_i(n)$ denotes the number of vertices of degree i in a graph of order n. Let $\Omega(\mathcal{D}_n)$ be the set of graphs on n vertices with degree sequence $\mathcal{D}_n$, and let $G(\mathcal{D})$ be a random graph chosen from the set $\Omega(\mathcal{D}_n)$ uniformly at random. For $i \geq 0$, let $\lambda_i(n) = d_i(n)/n$. Furthermore, $\mathcal{D}$ is said to be *well-behaved* if $\mathcal{D}$ is such that (i) $\Omega(D_n) \neq \emptyset$ for all $n \geq 1$, (ii) $\lambda_i(n)$ converges to a constant λ_i^* as $n \to \infty$ for every $i \geq 0$, and (iii) for a polynomial$f(i)$ in i with degree at most 3, the sequence $f(i)\lambda_i(n)$ tends uniformly to $f(i)\lambda_i^*$, and the sum $\sum_{i \geq 1} f(i)\lambda_i(n)$ approaches a limit uniformly as $n \to \infty$. Consider the function $Q(\mathcal{D})$ defined by

$$Q(\mathcal{D}) = \sum_{i \geq 1} i(i-2)\lambda_i^*.$$

Molloy and Reed [102, 103] showed that if $\mathcal{D}$ is well-behaved, then the *phase transition* occurs when $Q(\mathcal{D}) = 0$. If $Q(\mathcal{D}) < 0$, then a.a.s. $G(\mathcal{D})$ consists of many small components, whereas if $Q(\mathcal{D}) > 0$, then there is a.a.s. a unique component of order $\Theta(n)$, while all other components are small.

Kang and Seierstad [87] further studied how the order of the largest component changes when $Q(\mathcal{D}) = 0$. Let τ_n be the solution of $Q_n(\tau_n) = 0$ where $Q_n(x) = \sum_{i \geq 1} i(i-2)\lambda_i(n)x^i$ is a generating function in variable x. If $\mathcal{D}$ is well-behaved and $d_i(n) = 0$ whenever $i > n^{1/4-\varepsilon}$ for some $\varepsilon > 0$, they showed that if $(1 - \tau_n)n^{1/3} \to -\infty$, then a.a.s. all components in $G(\mathcal{D})$ are of order $o(n^{2/3})$, and that there is a constant c such that if $(1 - \tau_n)n^{1/3} \geq c \log n$, then a.a.s. $G(\mathcal{D})$ has a single component of order $\gg n^{2/3}$, while all other components are of order $o(n^{2/3})$.

Further Łuczak [97] and Chung and Lu [45] studied the component structure of a random graph with a given degree sequence, and Newman, Strogatz, and Watts [107] the phase transition using generating functions.

3.3 Random graph processes with degree constraints.

There is no obvious way to define a graph process such that adding a single or a few edges in each step we could achieve a uniformly distributed random regular graph or a uniformly distributed random graph with a given degree sequence, as a final graph or at some point. There are, however, several random graph processes relevant to them, though a graph generated by such a process is usually not uniformly distributed.

The *d-process* $(G^d(M))_{0\le M\le\lfloor dn/2\rfloor}$ is defined as follows. The initial state $G^d(0)$ is an empty graph with vertex set $[n]$, and $G^d(M+1)$ is obtained from $G^d(M)$ by adding an edge e chosen uniformly at random among all edges $e \notin G^d(M)$ such that the graph $G^d(M)+e$ has maximum degree at most d. If there is no such edge e, then $G^d(M+1) = G^d(M)$. Ruciński and Wormald [125, 124] showed that *a.a.s.* the final state $G_d(dn/2)$ of the process is a d-regular graph and is connected when $d \ge 3$, provided that dn is even. Additionally, Greenhill, Ruciński, and Wormald [74] investigated a generalization of this process to hypergraphs.

In the *star d-process* suggested by Robalewska [121] several edges may be added at each step. Starting with an empty graph $G^d_*(0)$ on n vertices, we obtain $G^d_*(M+1)$ from $G^d_*(M)$ by first choosing a random vertex v of minimum degree δ, and then connecting v with $d-\delta$ other vertices of degree $< d$ chosen uniformly at random. If there are less than $d-\delta$ such vertices left, the process stops, and we let G^d_* be the final graph of the process. Robalewska and Wormald [122] proved that *a.a.s.* G^d_* is d-regular. Furthermore, Greenhill, Ruciński, and Wormald [73] showed that G^d_* is connected *a.a.s.* if $d \ge 3$. Moreover, they proved that *a.a.s.* G^d_* is d-connected if $d \ge d_0$ for a certain constant $d_0 > 0$.

A different kind of random graph processes with degree restriction is the *minimum degree multigraph process*, in which the minimum degree increases quickly. The M-th state $G_{\min}(M)$ of this process has precisely M edges, and $G_{\min}(M+1)$ is obtained from $G_{\min}(M)$ by connecting a random vertex v of minimum degree with a further vertex $w \ne v$ chosen uniformly at random from $[n] \setminus \{v\}$. Note that the process may produce multiple edges, and that the maximum degree of $G_{\min}(M)$ is not restricted.

In Chapter 12 (see also [86, 88]) we show that if $M = tn$ then *a.a.s.* $G_{\min}(M)$ becomes connected as soon as the minimum degree reaches three, which happens at $t \sim h_3 \doteq 1.7316$. Moreover, for $t \ne h_2 \doteq 1.2197$ the probability that $G_{\min}(M)$ is connected tends to a certain function $\rho(t)$ as $n \to \infty$. The function $\rho(t)$ is continuous for all $t \ne h_2$ such that $\rho(t) = 0$ for $t < h_2$, $\rho(t) = 1$ for $t \ge h_3$, and $0 < \rho(t) < 1$ for $t \in (h_2, h_3)$. Furthermore, we prove that there is a constant $h_{\mathrm{cr}} \doteq 0.8607$ such that for $M = tn$, *a.a.s.* $G_{\min}(M)$ consists of small components on $O(\log n)$ vertices if $t < h_{\mathrm{cr}}$, whereas it consists of one giant component on $\Theta(n)$ vertices and small components on $O(\log n)$ vertices if $t > h_{\mathrm{cr}}$.

Jaworski and Łuczak [83] considered a directed version $(\overrightarrow{D}(M))_{0\le M\le n(n-1)}$ of the minimum degree process. They proved that the multigraph $D(M)$ obtained from $\overrightarrow{D}(M)$ by "forgetting" the directions of the edges is connected *a.a.s.* if $M \ge (2-o(1))n$. Furthermore, if $M = (1+t)n$ for a constant $t > 0$, then the probability that $D(M)$ is connected lies strictly between 0 and 1. In addition, Jaworski and Łuczak considered the distribution of the largest component of $D(M)$. If $M = (1+o(1))n$, then the number of vertices outside the largest component has a Gamma distribution. Moreover, for $M = (1+t)n$ with $t > 0$ fixed all components of $D(M)$ except for the largest one are unicyclic, and the limiting distribution of the number of vertices outside the largest component is known precisely.

In Chapter 13 ([46]) we study the *min-min random graph process* $(G_M)_{M\ge 0}$ that "interpolates" between the random regular graphs $G^r(M)$, $r \ge 1$. The process starts with an empty graph G_0 on n vertices, where n is even. Furthermore, in each step G_{M+1} is obtained from G_M by first choosing a pair

$\{v, w\}$ of two non-adjacent distinct vertices of G_M such that *both* v, w *are of minimum degree* uniformly at random among all such pairs, and then adding the edge $\{v, w\}$ to G_M to obtain G_{M+1} (if it impossible to continue this way, which happens with probability tending to zero as $n \to \infty$, we restart the process from the empty graph). Letting $M = (1 + t)n$, we show that *a.a.s.* G_M is connected if $t > 1/2$ and is disconnected if $t \leq 0$, and that the probability that G_M is connected converges to $2\sqrt{t - t^2}\ \exp((2t - 1)^2/2)$ if $1 \leq t \leq 1/2$. Furthermore, we prove that for a constant t satisfying $1 \leq t \leq 1/2$ and for a positive integer l the order of the giant component is $n - 2l$ with probability tending to $2\exp(2t(t - 1))\sqrt{t(1 - t)/\pi l}(1 - 2t)^{2l}$ as $n \to \infty$, and that for $t = t(n) = o(1)$ and for any $x > 0$ it is at most $n - x/t$ with probability tending to $(1 + O(1/x))\ (2\pi x)^{-1/2} \exp(-2x)$ as $n \to \infty$.

3.4 Preliminaries

Notations and asymptotics. For a real number x and an integer $r \geq 0$ we let

$$(x)_r = \prod_{j=0}^{r-1}(x - j).$$

If $r = 0$, then $(x)_r = 1$. Further, if ν is even, then

$$(\nu - 1)!! = \prod_{j=1}^{\nu/2}(\nu - 2j + 1)$$

denotes the number of perfect matching of a complete graph on ν vertices. We define $(\nu - 1)!! = 1$ for $\nu = 0$. Moreover, we need Stirling's formula

$$m! \quad = \quad (1 + O(1/m))\sqrt{2\pi m}(m/\mathrm{e})^m \quad (m \to \infty), \tag{3.4.1}$$

which implies that

$$\begin{aligned}(m - 1)!! \quad &= \quad \frac{m!}{2^{m/2}(m/2)!} \\ &= \quad \sqrt{2}\left(\frac{m}{\mathrm{e}}\right)^{m/2}(1 + O(m^{-1})) \quad (m \to \infty \text{ even}).\end{aligned} \tag{3.4.2}$$

Finally, we let $\vec{i}$ denote the imaginary unit $\sqrt{-1}$.

Probability distributions. A discrete random variable X is said to have *Poisson distribution* with mean $\lambda > 0$ if

$$\Pr(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad \text{for} \quad k = 0, 1, 2, \cdots.$$

Its mean and variance are $\mathbb{E}(X) = \lambda$ and $\mathrm{Var}(X) = \lambda$. Moreover, its characteristic function is

$$\mathbb{E}(\exp(\vec{i}tY)) = \exp(\lambda(\exp(\vec{i}t) - 1)). \tag{3.4.3}$$

Let $(X_{1,n}, \ldots, X_{m,n})_{n \geq 1}$ be a family of random variables such that each of the random variables $X_{1,n}, \ldots, X_{m,n}$ are defined on the same probability space

for each n. Let $\lambda_1, \ldots, \lambda_m \geq 0$. Then $(X_{1,n}, \ldots, X_{m,n})_{n\geq 1}$ is called *a family of asymptotically independent Poisson variables* if for all $k_1, \ldots, k_m \geq 0$ we have

$$\lim_{n\to\infty} \Pr\left[X_{1,n} = k_1 \wedge \cdots \wedge X_{m,n} = k_m\right] = \prod_{j=1}^{m} \frac{\lambda_j^{k_j}}{k_j!} \exp(-\lambda_j). \tag{3.4.4}$$

To show that (3.4.4) holds, the following theorem is useful. Its proof can be found, e.g., in [33, p. 26].

Theorem 3.4.1. *Suppose that for any sequence $(r_1, \ldots, r_m)$ of integers ≥ 0,*

$$\lim_{n\to\infty} \mathbb{E}\left[\prod_{j=1}^{m} (X_{j,n})_{r_j}\right] = \prod_{j=1}^{m} \lambda_j^{r_j}.$$

Then $(X_{1,n}, \ldots, X_{m,n})_{n\geq 1}$ is a family of asymptotically independent Poisson variables.

A discrete random variable X is said to have *geometric distribution* with success probability $0 < p < 1$ if

$$\Pr(X = k) = p(1-p)^k, \quad \text{for} \quad k = 0, 1, 2, \cdots.$$

Its mean and variance are $\mathbb{E}(X) = (1-p)/p$ and $\mathrm{Var}(X) = (1-p)/p^2$.

A continuous random variable Y is said to have *Gaussian distribution* with mean μ and variance σ^2 if

$$\Pr(a \leq Y \leq b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx,$$

A continuous random variable Y is said to have *Gamma distribution* $\Gamma(k, \theta)$ with shape parameter k and scale parameter θ if

$$\Pr(a \leq Y \leq b) = \int_a^b \frac{x^{k-1} \exp(-x/\theta)}{\Gamma(k)\theta^k} dx,$$

where $\Gamma(k) = \int_0^\infty y^{k-1} \exp(-y) dy$. Its mean and variance are $\mathbb{E}(Y) = k\theta$ and $\mathrm{Var}(Y) = k\theta^2$. Moreover, its characteristic function is

$$\mathbb{E}(\exp(\vec{i}tY)) = (1 - \vec{i}t\theta)^{-k}. \tag{3.4.5}$$

Recall that a sequence $(X_m)_{m\geq 1}$ of real-valued random variables *converges to X in distribution* if $\lim_{m\to\infty} \Pr\left[X_m \leq x\right] = \Pr\left[X \leq x\right]$ for all x where the function $x \mapsto \Pr\left[X \leq x\right]$ is continuous. In order to prove convergence in distribution, we will use the following theorem (see [62, Vol. 2, p. 508]).

Theorem 3.4.2. *The characteristic functions $R \to \mathbf{C}$, $y \mapsto \mathbb{E}(\exp(\vec{i}yX_m))$, $m \geq 1$, converge pointwise to $y \mapsto \mathbb{E}(\exp(\vec{i}yX))$ if and only if $(X_m)_{m\geq 1}$ converges to X in distribution.*

Furthermore, the following lemma is helpful to analyse characteristic functions.

Lemma 3.4.3. *Let $z \in R \setminus \{0\}$, and let $0 \leq \tau, \sigma \leq 1$. Suppose that X, Y are random variables such that $\Pr[|X - Y| > \sigma/|z|] \leq \tau$. Then*

$$|\mathbb{E}(\exp(\vec{i}zX)) - \mathbb{E}(\exp(\vec{i}zY))| \leq 2\sigma + \tau.$$

Proof. If $|X - Y| \leq \sigma/|z|$, then $|\exp(\vec{i}z(X - Y)) - 1| \leq \exp(\sigma|z|) - 1 \leq 2\sigma$, so that $|\exp(\vec{i}zX) - \exp(\vec{i}zY)| \leq 2\sigma$. Hence, $|\mathbb{E}(\exp(\vec{i}zX)) - \mathbb{E}(\exp(\vec{i}zY))| \leq 2\sigma + \Pr[|X - Y| > \sigma/z] \leq 2\sigma + \tau$, as claimed. □

Basic inequalities. Recall Markov's inequality and Chebyshev's inequality: for a non-negative random variable X and any $\lambda > 0$

$$\begin{aligned} \Pr(X \geq \lambda) &\leq \frac{\mathbb{E}(X)}{\lambda}, \\ \Pr(|X - \mathbb{E}(X)| \geq \lambda) &\leq \frac{\mathrm{Var}(X)}{\lambda^2}. \end{aligned}$$

We will use the following Chernoff bounds on the tails of a binomially distributed random variable X with mean μ (see [82, pages 26–28]). If $s > 0$, then

$$\begin{aligned} \Pr(X \geq \mu + s) &\leq \exp\left(-\frac{s^2}{2(\mu + s/3)}\right), \\ \Pr(X \leq \mu - s) &\leq \exp\left(-\frac{s^2}{2\mu}\right). \end{aligned} \tag{3.4.6}$$

We need also the following generalized Chernoff bound.

Lemma 3.4.4. *Let $X_1, X_2, \ldots$ be non-negative, integral, mutually independent, and identically distributed random variables with mean μ, and let $X = \sum_{i=1}^{r} X_i$. For any $\delta > 0$ and $a > 0$, there is a $c > 0$ such that if $r \geq c \log n$, then*

$$\Pr[X \geq r(\mu + \delta)] = o(n^{-a}), \tag{3.4.7}$$

$$\Pr[X \leq r(\mu - \delta)] = o(n^{-a}). \tag{3.4.8}$$

Proof. For any $u > 0$,

$$\begin{aligned} \Pr[X \geq r(\mu + \delta)] = \Pr\left[e^{uX} \geq e^{ur(\mu+\delta)}\right] &\leq \mathbb{E}\left[e^{uX}\right] e^{-ur(\mu+\delta)} \\ &= \left(\mathbb{E}\left[e^{uX_i}\right] e^{-u(\mu+\delta)}\right)^r. \end{aligned}$$

Let $f(u) = \mathbb{E}\left[e^{uX_i}\right] e^{-u(\mu+\delta)}$, and $g(u) = \mathbb{E}\left[e^{uX_i}\right]$. To prove (3.4.7), it is sufficient to show that there is a $u > 0$ such that $f(u) < 1$. Since $f(0) = 1$, we only have to show that the derivative of $f(u)$ at $u = 0$ is negative.

Differentiating $g(u)$, we get

$$g'(u) = \sum_{k \geq 0} k e^{uk} \Pr[X_i = k],$$

so we have $g'(0) = \sum_{k \geq 0} k \Pr[X_i = k] = \mu$, and $f'(0) = g'(0) - g(0)(\mu + \delta) = \mu - \mu - \delta = -\delta < 0$.

To prove (3.4.8), we fix an m, and define random variables Y_i such that

$$Y_i = \begin{cases} m - X_i & \text{if } X_i \le m, \\ 0 & \text{otherwise.} \end{cases}$$

Let $\mu' = \mathbb{E}[Y_i]$. For every $\eta > 0$ we can choose m large enough so that

$$\sum_{k=0}^{m} k \Pr[X_i = k] \ge \mu - \eta.$$

Hence we can choose m so large that

$$\begin{aligned} \mu' = \mathbb{E}[Y_i] &= \sum_{k \ge 0} k \Pr[Y_i = k] = \sum_{k=0}^{m} (m-k) \Pr[X_i = k] \\ &= m \sum_{k=0}^{m} \Pr[X_i = k] - \sum_{k=0}^{m} k \Pr[X_i = k] \le m - \mu + \eta. \end{aligned}$$

Let $Y = \sum_{i=1}^{r} Y_i$, and note that $Y \ge rm - X$, and that $m - \mu \le \mu' \le \mu' + m\eta$. We set $\eta = \delta/2$. Then

$$\begin{aligned} \Pr[X \le r(\mu - \delta)] &= \Pr[rm - X \ge rm - r(\mu - \delta)] \\ &\le \Pr[Y \ge r(m - \mu + \delta)] \\ &\le \Pr[Y \ge r(\mu' + \delta/2)], \end{aligned}$$

by choosing m sufficiently large. By (3.4.7) this probability can be well enough bounded, by choosing a large c. □

Wormald's differential equation method. Our argument in Chapter 12 is based on the following general result proved by Wormald [145, 148]. We use a version of Wormald's theorem, which can be found in Theorem 5.1 in [145]. The note following that theorem covers the case that k is a function of n satisfying $k = O(\log n)$.

Lemma 3.4.5. *Let $\{G_M\}_{M \ge 0}$ be a random graph process whose states are graphs on n vertices. For $k = 1, 2, \ldots, k_0$, with $k_0 = O(\log n)$, let $X_k(T)$ be a random variable defined on the process up to time T, $\{G_M\}_{M=0}^{T}$, for each $T \ge 0$. Suppose also that $|X_k(T)| < Cn$ for some constant C, and that for some functions $m = m(n)$, and $f_k : R^{k_0+1} \to R$ the following conditions $(i), (ii)$ and (iii) hold.*

(i) There is a constant C' such that

$$\max_{1 \le k \le k_0} |X_k(T+1) - X_k(T)| \le C'$$

for all $T < m$.

(ii) For $k = 1, 2, \ldots, k_0$,

$$\begin{aligned} &\mathbb{E}\left[X_k(T+1) - X_k(T) \mid \{G_M\}_{M=1}^{T}\right] \\ &\qquad = f_k(T/n, X_1(T)/n, \ldots, X_{k_0}(T)/n) + o(1) \end{aligned}$$

uniformly over all $T < m$.

(iii) For each $k = 1, 2, \ldots, k_0$, the function f_k is continuous and satisfies a Lipschitz condition on D, where D is some bounded connected open set containing the intersection of $\{(s, z_1, \ldots, z_{k_0}) : s \geq 0\}$ with some neighbourhood of $\{(0, z_1, \ldots, z_{k_0}) : \Pr[X_k(0) = z_k n, 1 \leq k \leq k_0] \neq 0\}$.

Then,

(a) for $(T, \hat{z}_1, \ldots, \hat{z}_{k_0}) \in D$, the system of differential equations

$$\frac{dz_k}{ds} = f_k(s, z_1, \ldots, z_{k_0}), \qquad k = 1, 2, \ldots, k_0,$$

has a unique solution in D for $z_k : R \to R$ passing through

$$z_k(0) = \hat{z}_k, \quad k = 1, 2, \ldots, k_0,$$

and which extends to points arbitrarily close to the boundary of D, and

(b) for each $k = 1, 2, \ldots, k_0$, a.a.s.

$$X_k(T) = n z_k(T/n) + o(n)$$

uniformly for $0 \leq T \leq \min\{\sigma n, m\}$, where $z_k(T)$ is the solution in (a) with $\hat{z}_k = X_k(0)/n$, and $\sigma = \sigma(n)$ is the supremum of those s to which the solution can be extended.

Multi-type branching process. In a multi-type branching process there is a finite number of types of particles. A particle generates a number of new particles according to a probability distribution which depends only on the type of the particle. For a multi-type branching process with s types of particles, let $p^{(i)}(j_1, \ldots, j_s)$, for $i = 1 \ldots, s$, be the probability that a particle of type i produces $j_{i'}$ particles of type i' for $i' = 1, \ldots, s$. We let

$$f^{(i)}(x_1, \ldots, x_s) = \sum_{j_1, \ldots, j_s \geq 0} p^{(i)}(j_1, \ldots, j_s) x_1^{j_1} \ldots x_s^{j_s}$$

be the probability generating function associated with particles of type i for $i = 1, \ldots, s$.

We let $A = \{a_{ij}\}$ be the $s \times s$ matrix where a_{ij} is the expected number of particles of type j generated from a single vertex of type i. A is called the *transition matrix* of the branching process. Then the following holds.

Theorem 3.4.6. *Assume that A is positive regular and nonsingular, and let λ_1 be the largest eigenvalue of A. If $\lambda_1 \leq 1$, then (except in certain trivial cases) the branching process dies out after a finite number of steps with probability one. If $\lambda_1 > 1$, then the branching process continues forever with positive probability.*

Furthermore, in the latter case, the set of equations

$$\begin{aligned} y_1 &= f^{(1)}(y_1, \ldots, y_s) \\ &\vdots \\ y_s &= f^{(s)}(y_1, \ldots, y_s) \end{aligned}$$

has a unique solution satisfying $0 \leq y_i < 1$ for $i = 1, \ldots, s$. Then y_i is the probability that a branching process starting with a single particle of type i dies out after a finite number of steps.

For multi-type branching processes and relevant inhomogeneous random graphs, see [8] and [38], respectively.

Part II

Labeled Planar Structures

Chapter 4

Labeled Forests

In this chapter we present the well-known results on forests based on [63, 76], to illustrate the method that we will use for other planar structures. For labeled forests and trees, see e.g., [1, 120, 134, 144].

In Section 4.1 we derive the exponential generating functions for labeled forests and labeled trees. In Section 4.2 we determine the asymptotic numbers of labeled forests and labeled trees, e.g., the asymptotic number of labeled trees $[n] := \{1, 2, \cdots, n\}$ is $c\ n^{-5/2}\ e^n\ n!$ for a suitable constant $c > 0$ and the base of the natural logarithm $e \doteq 2.71828$. In Section 4.3 we compute the exact numbers of labeled forests and labeled trees. In Section 4.4 we present how the recursive method works. In particular, we derive a uniform sampling algorithm for labeled trees running in $\tilde{O}(n^4)$ with space $O(n^3 \log n)$.

4.1 Exponential generating functions

To obtain a well-known relation between labeled forests and labeled trees it is customary to consider labeled *rooted* forests and labeled *rooted* trees. A labeled rooted tree is a labeled tree with one distinguished vertex called the root, and a labeled rooted forest is a labeled forest, each of whose connected components is a labeled rooted tree.

Let $f(n)$ and $t(n)$ denote the number of labeled forests and the number of labeled trees on vertex set $[n]$ for a nonnegative integer n and $\hat{f}(n), \hat{t}(n)$ the rooted counterparts. As a convention we set $f(0) = \hat{f}(0) = 1$ and $t(0) = \hat{t}(0) = 0$. Let

$$F(x) = \sum_{n \geq 0} f(n) \frac{x^n}{n!} \quad \text{and} \quad T(x) = \sum_{n \geq 0} t(n) \frac{x^n}{n!}$$

be the exponential generating functions of $f(n)$ and $t(n)$. And $\hat{F}(x)$ and $\hat{T}(x)$ are defined analogously.

Note that a labeled rooted forest is a non-ordered collection of labeled rooted trees, and thus we get

$$\hat{F}(x) = \sum_{k \geq 0} \frac{(\hat{T}(x))^k}{k!} = e^{\hat{T}(x)}. \tag{4.1.1}$$

On the other hand, $\hat{f}(n)$ and $\hat{t}(n)$ satisfy

$$\hat{t}(n+1) = (n+1)\hat{f}(n). \tag{4.1.2}$$

To see this we consider a labeled rooted forest with vertex set $[n]$. We introduce a new vertex and assign to it an arbitrary label from $[n+1]$, say i with $1 \le i \le n+1$. We relabel the given forest with the vertex set $[n+1] \setminus \{i\}$ preserving the relative order of the original labels. We connect the new vertex i with the rest of the vertices to obtain a labeled rooted tree on $[n+1]$. Thus exactly $(n+1)$ labeled rooted trees with vertex set $[n+1]$ are constructed from a labeled rooted forest with vertex set $[n]$. This construction can be reversed.

From (4.1.2) we can see that

$$\hat{F}(x) = \sum_{n\ge 0} \hat{f}(n)\frac{x^n}{n!} = \sum_{n\ge 0} \frac{\hat{t}(n+1)}{n+1}\frac{x^n}{n!} = \frac{\hat{T}(x)}{x}. \tag{4.1.3}$$

Combining (4.1.1) and (4.1.3) we have

$$\hat{T}(x) = xe^{\hat{T}(x)}. \tag{4.1.4}$$

4.2 Singularity analysis

Applying Lemma 2.4.1 to (4.1.4) with $\psi(u) = ue^{-u}$ and $\phi(u) = e^u$, we see that $\psi'(u_0) = 0, \psi''(u_0) \neq 0$ at $u_0 = 1$ and hence $\hat{T}(x)$ has a dominant singularity at $x_0 = \psi(u_0) = e^{-1}$. Further the singular expansion of $\hat{T}(x)$ is of the square root type

$$\hat{T}(x) = 1 + c_1(1 - x/x_0)^{1/2} + O((1 - x/x_0)),$$

and its coefficients satisfy

$$[x^n]\hat{T}(x) \sim \frac{1}{\sqrt{2\pi}} n^{-3/2} e^n, \tag{4.2.1}$$

which implies that

$$t(n) = \frac{n![x^n]\hat{T}(x)}{n} \sim \frac{1}{\sqrt{2\pi}} n^{-5/2} e^n n!, \quad e \doteq 2.71828.$$

Furthermore (4.1.3) implies that

$$\begin{aligned} \hat{F}(x) &= e^{1+c_1(1-x/x_0)^{1/2}+O((1-x/x_0))} \\ &\sim e + ec_1(1 - x/x_0)^{1/2} + O((1 - x/x_0), \end{aligned}$$

and thus we obtain the asymptotic number of labeled rooted forests on n vertices

$$[x^n]\hat{F}(x) \sim \frac{e}{\sqrt{2\pi}} n^{-3/2} e^n.$$

4.3 Exact number of labeled trees and forests

Taking $\phi(u) = e^u$ and $f(x) = \hat{T}(x)$ in Lagrange Inversion Theorem 2.4.1 we obtain

$$\begin{aligned}[x^n]\hat{T}(x) &\overset{(2.4.2)}{=} \frac{1}{n}[u^{n-1}]\phi(u)^n \\ &= \frac{1}{n}[u^{n-1}]e^{un} = \frac{1}{n}[u^{n-1}]\sum_{k\geq 0}\frac{(un)^k}{k!} = \frac{n^{n-1}}{n!}\,, \end{aligned} \quad (4.3.1)$$

which, together with Stirling's formula (3.4.1), yields also (4.2.1). Further we have that the number of labeled rooted trees and the number of labeled trees on n vertices satisfy

$$\begin{aligned}\hat{t}(n) &= n!\,[x^n]\hat{T}(x) \overset{(4.3.1)}{=} n^{n-1}, \\ t(n) &= \frac{\hat{t}(n)}{n} = n^{n-2}, \end{aligned} \quad (4.3.2)$$

the latter of which is known as *Cayley's formula*. See [1] for four different proofs of it.

Let $\hat{f}_c(n)$ be the number of labeled rooted forests on n vertices consisting of c labeled rooted trees and $\hat{F}_c(x) = \sum_{n\geq 0}\hat{f}_c(n)\frac{x^n}{n!}$ be its exponential generating function. Then

$$\hat{F}_c(x) = \frac{\hat{T}(x)^c}{c!}.$$

Thus $\hat{f}_c(n)$ can be obtained by Lagrange Inversion Theorem 2.4.1

$$[x^n]\hat{F}_c(x) = \frac{1}{c!}[x^n]\hat{T}(x)^c \overset{(2.4.2)}{=} \frac{1}{c!}\frac{c}{n}[u^{n-c}]\phi(u)^n = \frac{n^{n-c-1}}{(c-1)!\,(n-c)!}\,,$$

and thus

$$\hat{f}_c(n) = n!\,[x^n]\hat{F}_c(x) = \frac{n!\;n^{n-c-1}}{(c-1)!\,(n-c)!} = \binom{n}{c}cn^{n-c-1}\,. \quad (4.3.3)$$

Further we have

$$\hat{f}(n) = \sum_{c=1}^{n}\hat{f}_c(n) \overset{(4.3.3)}{=} \sum_{c=1}^{n}\binom{n}{c}cn^{n-c-1} = \sum_{c=1}^{n}\binom{n-1}{c-1}n^{n-c}\,.$$

Since we may choose c roots in $\binom{n}{c}$ possible ways, the number of labeled rooted forests on n vertices consisting of c labeled rooted trees, where the vertices $\{1, 2, \cdots, c\}$ appear in different trees, can be obtained by (4.3.3) and it is cn^{n-c-1}, which is also known as Cayley's formula.

Using the following identities

$$\begin{aligned}F(x) &= e^{T(x)} \\ T(x) &= \sum_{n\geq 0} n^{n-2}\frac{x^n}{n!}\end{aligned}$$

we can obtain the first few terms of the number of trees and forests. Table 4.1 shows the exact numbers t_n and f_n of labeled trees and labeled forests on n vertices, up to $n = 15$ (see also [133, A000272, A001858]).

n	t_n	f_n
1	1	1
2	1	2
3	3	7
4	16	38
5	125	291
6	1296	2932
7	16807	36961
8	262144	561948
9	4782969	10026505
10	100000000	205608536
11	2357947691	4767440679
12	61917364224	123373203208
13	1792160394037	3525630110107
14	56693912375296	110284283006640
15	1946195068359375	3748357699560961

Table 4.1: The exact numbers t_n, f_n of labeled trees and forests on n vertices, up to $n = 15$.

4.4 Recursive counting and uniform sampling

In this section we explain how we decompose *labeled forests* and *labeled trees* recursively and obtain recursive counting formulas. We illustrate also how one can use recursive counting formulas, in order to generate a random structure uniformly at random.

As before, let $f(n)$ and $t(n)$ denote the number of forests and the number of trees on vertex set $[n]$ for a nonnegative integer n. Further let $f_c(n)$ be the number of forests consisting of c connected components. Of course, a connected component in a forest is a tree and thus $f_1(n) = t(n)$. It is clear that a forest consists of at least one tree but at most n trees. The formulas (2.2.1) and (2.2.2) holds with $g(n) = f(n)$, $g_c(n) = f_c(n)$ and $c(n) = t(n)$.

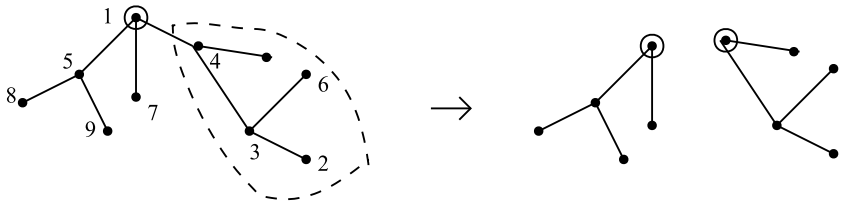


Figure 4.1: Pulling off the split subtree.

To count trees on vertex set $[n]$ we let $t_d(n)$ be the number of trees where the vertex 1 has degree d. It is clear that $t_d(n) = 0$ for $d = 0$ or $d \geq n \geq 1$, except that we define $t_1(1) = 1$, and $t(n) = \sum_{d=1}^{n-1} t_d(n)$ for $n \geq 2$. Let T be a tree on $[n]$ for $n \geq 2$. If the vertex 1 of T has degree one, we split off the vertex 1 from T and obtain a remaining tree with $n-1$ vertices, hence $t_1(n) = t(n-1)$. If the vertex 1 of T has degree $d \geq 2$, we decompose T into two subtrees: One is the subtree containing the vertex 2, which is called the *split subtree* and the

other one is the rest of the tree together with the vertex 1, which is called the *root subtree* (see Fig. 4.1). If the split subtree consists of i vertices, then there are $\binom{n-2}{i-1}$ ways to select the vertex set of the split tree, since the vertex 2 is already designated to the split tree and the vertex 1 to the root subtree. Thus $t_d(n) = \sum_{i=1}^{n-d} \binom{n-2}{i-1} t(i) t_{d-1}(n-i)$ for $d \geq 2$, $n \geq 2$.

Overall we have

$$t(n) = \begin{cases} 0 & \text{for } n = 0 \\ \sum_{d=1}^{n-1} t_d(n) & \text{for } n \geq 1\,, \end{cases} \tag{4.4.1}$$

$$t_d(n) = \begin{cases} 0 & \text{for } n = 0 \\ 1 & \text{for } d = n = 1 \\ t(n-1) & \text{for } d = 1 \text{ and } n \geq 2 \\ \sum_{i=1}^{n-d} \binom{n-2}{i-1} t(i) t_{d-1}(n-i) & \text{for } d \geq 2 \text{ and } n \geq 2\,. \end{cases} \tag{4.4.2}$$

Once the exact recursive counting formulas with appropriate parameters are known, it is easy to derive a generation procedure. See Figure 4.2, where **Generate**(n) corresponds to (4.4.1) and **Generate**(n, d) to (4.4.2).

Generate(n): returns random tree on $[n]$.
  if $n = 0$ **then return** $\emptyset$
  else
  choose the degree d of vertex 1 with probability $t_d(n)/t(n)$
    return Generate(n, d)

Generate(n, d): returns tree on $[n]$ with vertex 1 of degree d
  if $n = 0$ **then return** $\emptyset$
  if $n = 1$
    if $d = 1$ **then return** $\{1\}$
    else return $\emptyset$
  else
  if $d = 1$ **then let** $T =$ **Generate**$(n-1)$;
    relabel vertex j in T to $j + 1$;
    return $T \cup \{1\} \cup \{(1, 2)\}$
    else
    choose the size i of the split subtree with probability
      $\binom{n-2}{i-1} t(i) t_{d-1}(n-i)/t_d(n)$
      let $w_1 = 2$
      choose random subset $\{w_2, \ldots, w_i\} \subseteq \{3, \ldots, n\}$
      $T_1 =$ **Generate**(i); relabel vertex j in T_1 to w_j
      $T_2 =$ **Generate**$(n - i, d - 1)$
      return $T_1 \cup T_2 \cup \{1\} \cup \{(1, w_1)\}$

Figure 4.2: Generating labeled trees uniformly at random.

To compute the numbers $t(n)$ and $t_d(n)$ efficiently, we store all of their values in a table to avoid recomputation, which is a technique called *dynamic programming*.

The algorithm has to compute sums of a linear number of products for the entries of a two-dimensional table. Since the number of labeled trees on n vertices grows with n^{n-2} (See (4.3.2)), the entries of the table have to store $O(n \log n)$ many bits. Assuming an $O(n \log n \log \log n)$ multiplication algorithm (see e.g., [44]), the number of computation steps needed to fill the quadratic size table is therefore in $\tilde{O}(n^4)$, where $\tilde{O}(\cdot)$ denotes growth up to logarithmic factors.

If we want to generate several random trees, we have to compute this table only once, and it makes sense to analyse the computation of the table separately as the precomputation step. The actual generation of a random tree can then be done much faster: We have to make a linear number of random decisions, each involving a random number with linearly many bits. This gives a quadratic running time.

We obtain the deterministic polynomial time algorithm to generate a labeled tree uniformly at random.

Theorem 4.4.1. *Labeled trees on n vertices can be sampled uniformly at random in deterministic time $\tilde{O}(n^4)$ with space $O(n^3 \log n)$. This can also be done in deterministic time $\tilde{O}(n^2)$ if we apply a precomputation step.*

To increase the efficiency of the algorithm one can use floating point numbers instead of arbitrary precision integer arithmetic. Since the algorithm can base its decisions on $O(\log n)$ bits in most cases, it might even be possible to get an exact uniform generator if we use certified floating-point arithmetics, see e.g., [47].

Chapter 5

Labeled Outerplanar Graphs

A graph G is *outerplanar*, if there is an embedding of G in the plane that has a face containing all the vertices of G. We will always draw this face as the outer face.

Many computational problems hard in general cases become tractable for outerplanar graphs [94, 110]. But still, their structure is rich enough so that many computational tasks remain challenging when the input is restricted to outerplanar graphs [32, 72]. Outerplanar graphs also attract increasing interest in graph drawing [7, 20, 93].

Mitchell [100] gave a linear time algorithm to recognize outerplanar and maximal outerplanar graphs. A maximal outerplanar graph has a simple structure. It can be seen as a triangulation of a convex polygon and its dual graph forms a binary tree structure. Thus maximal outerplanar graphs can be counted by the *Catalan number*, and can be efficiently generated [49, 58].

It is well known that two-connected outerplanar graphs can be seen as dissections of a convex polygon: For a two-connected outerplanar graph with at least three vertices has a unique Hamiltonian cycle [94] and can therefore be embedded uniquely in the plane so that this Hamiltonian cycle lies on the outer face. This unique embedding is thus a dissection of a convex polygon. It is also well known that the number of dissections of a convex polygon can be counted by the *Schröder number* [129, 133], also called the *bracketing number*. There are, however, no such simple formulas known for general outerplanar graphs.

In this chapter we determine that the number of labeled outerplanar graphs on n vertices is asymptotically $c\, n^{-5/2}\, \gamma^n\, n!$, for a suitable positive constant c and $\gamma \doteq 7.32098$. We also study typical properties of a random outerplanar graph chosen uniformly at random among all the labeled outerplanar graphs on n vertices. We show that the *isolated vertex conjecture* (see [99]) is true for a random outerplanar graph. The expected number of edges in a random outerplanar graph is asymptotically $1.56n$ and the probability that a random outerplanar graph is connected tends to 0.861 as n goes to ∞. Furthermore, the chromatic number of a random outerplanar graph is three with probability tending to one as n goes to ∞. Finally, we derive the *first* polynomial time algorithm that samples a random outerplanar graph uniformly at random running

in $\tilde{O}(n^4)$ and with space $O(n^3 \log n)$, using the recursive method.

The rest of the chapter is organized as follows: In Sections 5.1 and 5.2 we derive the equations for generating functions for labeled outerplanar graphs with various connectivity. We also derive the asymptotics by studying the singularities of generating functions. In Section 5.3 we study the asymptotics of the probability of a random outerplanar graph being connected and the number of edges in a random outerplanar graph. In Section 5.4 we show how to decompose labeled outerplanar graphs based on connectivity and derive recurve counting formulas along the decomposition. Furthermore, we design a uniform sampling algorithm to generate a random labeled outerplanar graph.

5.1 Exponential generating functions

Let g_n, c_n, and b_n be the numbers of labeled outerplanar, connected outerplanar, and two-connected outerplanar graphs on n vertices, respectively, and let $G(x), C(x)$, and $B(x)$ be their exponential generating functions:

$$G(x) := \sum_{n\geq 0} g_n \frac{x^n}{n!}\,, \quad C(x) := \sum_{n\geq 0} c_n \frac{x^n}{n!}\,, \quad B(x) := \sum_{n\geq 0} b_n \frac{x^n}{n!}\,.$$

Then $G(x), C(x)$, and $B(x)$ satisfies Proposition 2.2.1 and the following.

Proposition 5.1.1. *Let $B(x)$ be as above. Then*

$$B'(x) = \frac{1 + 5x - \sqrt{1 - 6x + x^2}}{8}\,. \tag{5.1.1}$$

Proof. As we have seen in the introduction, the task of counting two-connected outerplanar graphs coincides with the task of counting dissections of a convex polygon. Let us assume that the vertices of the Hamiltonian cycle of a two-connected outerplanar graph with n vertices are labeled in clockwise order, and form a convex n-gon K. A two-connected outerplanar graph then corresponds to a dissection of K. Let d_n denote the number of dissections of K and $D(x) := \sum_{n\geq 0} d_n x^n$ be its ordinary generating function. Since for $n \geq 2$, a dissection of K is an edge or a sequence of k ($k \geq 2$) dissections (along the face containing an edge incident to the two smallest vertices) where $k-1$ pairs of vertices are identified, we get

$$D(x) = x^2 + \sum_{k\geq 2} \frac{D(x)^k}{x^{k-1}} = x^2 + \frac{D(x)^2}{x - D(x)}\,.$$

The only meaningful solution of $D(x)$ for this equation is

$$D(x) = \frac{x(1 + x - \sqrt{1 - 6x + x^2})}{4}\,.$$

Since $b_n = d_n = 0$ for $n \leq 1$, $b_2 = d_2 = 1$ and $b_n = \frac{(n-1)!}{2} d_n$ for $n \geq 3$, we get

$$B'(x) = \frac{D(x)/x + x}{2} = \frac{1 + 5x - \sqrt{1 - 6x + x^2}}{8}\,.$$

So we are done.

On the other hand, we can derive the same result using the well-known fact that the number of dissections of a convex n-gon equals the Schröder number s_{n-1}. Thus for $n \geq 3$, the number b_n of two-connected outerplanar graph on n vertices is

$$b_n = \frac{(n-1)!}{2} s_{n-1}\,, \tag{5.1.2}$$

and $b_0 = b_1 = 0, b_2 = 1$. The numbers s_n satisfy the recursion: $s_0 = 0, s_1 = s_2 = 1$, and

$$s_n = ((6n-9)s_{n-1} - (n-3)s_{n-2})/n.$$

It is easy to see that the corresponding ordinary generating function $S(x) = \sum_{n\geq 0} s_n x^n$ satisfies

$$S(x) = x + \frac{S(x)^2}{1 - S(x)}.$$

The meaningful solution for $S(x)$ is

$$\begin{aligned} S(x) &= \frac{1}{4}(1 + x - \sqrt{1 - 6x + x^2}) \\ &= x + x^2 + 3x^3 + 11x^4 + 45x^5 + \ldots. \end{aligned} \tag{5.1.3}$$

Therefore $B'(x) = (S(x)+x)/2$, which, together with (5.1.3), implies (5.1.1). □

5.2 Asymptotics

In this section we present the result on the asymptotic number of labeled outerplanar, connected outerplanar, and two-connected outerplanar graphs. The proof is based on a singularity analysis of the corresponding generating functions.

Theorem 5.2.1. *Let g_n, c_n, and b_n be as above. Then*

$$\begin{aligned} g_n &\sim \alpha_0\ n^{-5/2}\ \gamma^n\ n!\,, \\ c_n &\sim \alpha_1\ n^{-5/2}\ \gamma^n\ n!\,, \\ b_n &\sim \alpha_2\ n^{-5/2}\ \delta^n\ n!\,, \end{aligned}$$

where the constants $\gamma, \delta, \alpha_0, \alpha_1$ and α_2 have explicit descriptions by radicals and can be computed efficiently with arbitrary precision. In particular, the first few digits of γ and δ are $\gamma \doteq 7.32098$ and $\delta \doteq 5.828427$.

The proof of Theorem 5.2.1 is based on *singularity analysis* (see Section 2.4). The generating functions $G(x), C(x)$ are related by the equations (2.2.3) and (2.2.4), and $B'(x)$ is defined by (5.1.1).

We will analyse the singularities of the generating functions, first for labeled two-connected outerplanar graphs, next for labeled connected outerplanar graphs, and finally for labeled outerplanar graphs.

Two-connected outerplanar graphs. We now perform the singularity analysis of the exponential generating function for labeled two-connected outerplanar graphs. The radius of convergence of $B'(x)$ (5.1.1) is $3-2\sqrt{2}$. Because $B(x)$ and $B'(x)$ have the same radius of convergence, we obtain

$$[x^n]B(x) = \theta(n)\delta^n\,,$$

where $\delta = 3+2\sqrt{2}$, and $\limsup|\theta(n)|^{1/n} = 1$. To determine $\theta(n)$ using Lemma 2.4.1, we rewrite (5.1.1) as

$$B'(x) = x\phi(B'(x)),$$

where

$$\phi(u) := \frac{-1+5u-\sqrt{1-4u+u^2}}{-2+8u}\,.$$

We observe that $\phi(u)$ is analytic at the origin with non-negative Taylor coefficients satisfying $\phi(0)\neq 0$, and its radius of convergence is $1/4$. Because $\phi(u)-u\phi'(u)=0$ has a solution $u_0=(8-5\sqrt{2})/4 \doteq 0.232233 \in (0,1/4)$, $B'(x)$ has a dominant singularity at $x_0 = u_0/\phi(u_0) = 3-2\sqrt{2}$, and we get

$$\begin{aligned}[x^n]B'(x) &= (1+O(n^{-1}))\frac{1}{\sqrt{2\pi}}\sqrt{\frac{\phi(u_0)}{\phi''(u_0)}}n^{-3/2}{x_0}^{-n}\\ &= (1+O(n^{-1}))\alpha_2 n^{-3/2}\delta^n\,,\end{aligned}$$

where $\delta := {x_0}^{-1} \doteq 5.828427$ and $\alpha_2 := \frac{1}{\sqrt{2\pi}}\sqrt{\phi(u_0)/\phi''(u_0)} \doteq 0.025665$, which can be computed efficiently with arbitrary precision. Lemma 2.4.1 also yields

$$\frac{b_n}{n!} = [x^n]B(x) = \frac{1}{n}[x^{n-1}]B'(x) = (1+O(n^{-1}))\alpha_2 n^{-5/2}\delta^n\,.$$

Connected outerplanar graphs. We apply the singularity analysis to the exponential generating function for labeled connected outerplanar graphs, which is defined by the implicit equation (2.2.4):

$$xC'(x) = x\exp(B'(xC'(x)))\,.$$

If we define $F(x) := xC'(x)$, then it becomes

$$F(x) = x\exp(B'(F(x)))\,.$$

It can be reformulated as

$$F(x) = x\Phi(F(x)),$$

where

$$\Phi(u) := \exp((1+5u-\sqrt{1-6u+u^2})/8)\,.$$

The function $\Phi(u)$ satisfies the conditions in Lemma 2.4.1, and its radius of convergence is $3-2\sqrt{2} \doteq 0.171573$. Note that $\Phi(u)-u\Phi'(u)=0$ has a solution $u_1 \doteq 0.170765 \in (0, 3-2\sqrt{2})$. It follows that $F(x)$ has a dominant singularity at

$$x_1 = u_1/\Phi(u_1) = u_1\exp((-1-5u_1+\sqrt{1-6u_1+u_1^2})/8) \doteq 0.136593\,,$$

and satisfies

$$F(x) \sim F(x_1) + \alpha_F (1 - x/x_1)^{1/2} . \tag{5.2.1}$$

The constant $\alpha_F = -\sqrt{\frac{2\Psi(u_1)}{-\Psi''(u_1)}}$ can be computed according to Lemma 2.4.1. Clearly, $F(x)$ and $C(x)$ have the same radius of convergence $R = x_1$. Theorem VI.6 in [63] states that the singular expansion of $C(x)$ can be obtained from $F(x)$ through a term-by-term integration, and we obtain

$$C(x) \sim C(R) + \alpha_C (1 - x/R)^{3/2} , \tag{5.2.2}$$

where $\alpha_C = -2/3\alpha_F = 2/3\sqrt{\frac{2\Psi(u_1)}{-\Psi''(u_1)}}$. To compute $C(R)$ we integrate $C(x) = \int_0^x \frac{F(t)}{t} dt$ by part and obtain

$$C(R) = u_1 \log R - u_1 \log u_1 + u_1 + B(u_1) \doteq 0.148886 .$$

From (5.2.1) and (2.4.2) we obtain

$$\begin{aligned} [x^n]F(x) &= (1 + O(n^{-1})) \frac{1}{\sqrt{2\pi}} \sqrt{\frac{\Phi(u_1)}{\Phi''(u_1)}} n^{-3/2} {x_1}^{-n} , \\ &= (1 + O(n^{-1})) \alpha_1 n^{-3/2} \gamma^n , \end{aligned}$$

where $\gamma := x_1^{-1} \doteq 7.32098$ and $\alpha_1 := \frac{1}{\sqrt{2\pi}} \sqrt{\Phi(u_1)/\Phi''(u_1)} \doteq 0.006976$, which can be computed with arbitrary precision and

$$\frac{c_n}{n!} = [x^n]C(x) = (1 + O(n^{-1})) \alpha_1 n^{-5/2} \gamma^n .$$

Outerplanar graphs. In this section we discuss general, i.e., not necessarily connected labeled outerplanar graphs. By Proposition 2.2.1 the corresponding exponential generating function is related to the exponential generating function for labeled connected outerplanar graphs by (2.2.3):

$$G(x) = \exp(C(x)) .$$

Clearly, both generating functions $G(x)$ and $C(x)$ have the same radius of convergence $R = \gamma^{-1}$. Using the asymptotic expansion (5.2.2) of $C(x)$ near the singularity, we have the asymptotic expansion of $G(x)$:

$$\begin{aligned} G(x) = \exp(C(x)) &= \exp(C(R)) \exp(\alpha_C (1 - x/R)^{3/2} + \dots) \\ &= e^{C(R)} \sum_{i \geq 0} (C(R) + \alpha_C (1 - x/R)^{3/2} + \dots)^i / i! \\ &\sim e^{C(R)} (1 + \alpha_C (1 - x/R)^{3/2}) \\ &= e^{C(R)} + \alpha_1 e^{C(R)} (1 - x/R)^{3/2} . \end{aligned} \tag{5.2.3}$$

Thus we get

$$\frac{g_n}{n!} = [x^n]G(x) = (1 + O(n^{-1})) \alpha_0 n^{-5/2} \gamma^n ,$$

where $\alpha_0 = \alpha_1 e^{C(R)} \doteq 0.008095$.

5.3 Random outerplanar graphs

In this section we study typical properties of a random outerplanar graph G_n that is chosen uniformly at random among all labeled outerplanar graphs on n vertices. We determine the probability of a random outerplanar graph being connected, the chromatic number of a random outerplanar graph, and the distribution of the number of edges in a random outerplanar graph.

5.3.1 Connectedness

First, we show that the *isolated vertex conjecture* is true for outerplanar graphs (which was conjectured for planar graphs in [99]): Let X_n be the number of isolated vertices in G_n. Then

$$\lim_{n\to\infty} \mathbb{E}(X_n) = \lim_{n\to\infty} n\frac{g_{n-1}}{g_n} = \lim_{n\to\infty} \left(1-\frac{1}{n}\right)^{-5/2} \frac{1}{\gamma} = \gamma^{-1}.$$

Next, we can show the following theorem on the asymptotic probability of connectedness of G_n:

Theorem 5.3.1. *Let γ be as in Theorem 5.2.1. Then*

$$e^{-1} < \lim_{n\to\infty} \Pr(G_n \textit{ is connected }) < e^{-1/\gamma}\,,$$

where $e^{-1} \doteq 0.367879$, and $e^{-1/\gamma} \doteq 0.872325$.

Proof. We can get a lower bound for the probability that a random outerplanar graph is connected using the following fact from [99]. Let $\mathcal{G}$ be a non-empty set of graphs such that (i) a graph G is in $\mathcal{G}$ if and only if each component of G is in $\mathcal{G}$, and (ii) for each graph G in $\mathcal{G}$, and for u, v chosen from two distinct components of G, the graph obtained from G by adding an edge connecting u and v is also in $\mathcal{G}$. Such a graph class is said to be *addable*. Outerplanar graphs are an example of an addable class. Then for a graph G chosen uniformly at random from the graphs with n vertices in $\mathcal{G}$,

$$\Pr(G \text{ is connected}) > e^{-1} \doteq 0.367879\,.$$

Some theorems in [99] were stated under the assumption of the isolated vertex conjecture for planar graphs. A moment of thought shows that these theorems hold for outerplanar graphs as well. In fact, most of the arguments in [99] directly apply to outerplanar graphs. In particular, we can see that if H is a fixed outerplanar graph, then G_n contains linearly many vertex disjoint copies of H with probability tending to 1 as n goes to ∞, and that

$$\lim_{n\to\infty} \Pr(G_n \text{ contains an isolated vertex}) = 1 - e^{-1/\gamma}\,,$$

which gives an upper bound of the probability that a random outerplanar graph G_n is connected:

$$\lim_{n\to\infty} \Pr(G_n \text{ is connected}) < e^{-1/\gamma} \doteq 0.872325\,.$$

□

Indeed we can say more about the connectivity probability using the asymptotic results in Theorem 5.2.1:

$$\lim_{n\to\infty} \Pr(G_n \text{ is two-connected }) = \lim_{n\to\infty} \frac{b_n}{g_n} = \lim_{n\to\infty} \frac{\alpha_2}{\alpha_0} \left(\frac{\delta}{\gamma}\right)^n = 0\,,$$

$$\lim_{n\to\infty} \Pr(G_n \text{ is two-connected } \mid G_n \text{ is connected}) = \lim_{n\to\infty} \frac{b_n}{c_n} = \lim_{n\to\infty} \frac{\alpha_2}{\alpha_1} \left(\frac{\delta}{\gamma}\right)^n = 0\,.$$

In addition we can precisely compute the asymptotic probability that G_n is connected:

$$\lim_{n\to\infty} \Pr(G_n \text{ is connected}) = \lim_{n\to\infty} \frac{c_n}{g_n} = \frac{\alpha_1}{\alpha_0} = e^{-C(R)} \doteq 0.861666\,.$$

Theorem 5.3.2. *For $k = 0, 1$ let $G_n^{(k)}$ denote a random graph chosen uniformly at random among all the labeled k-connected outerplanar graphs on vertices $1, \ldots, n$. As $n \to \infty$,*

$$\Pr(G_n^{(0)} \textit{ is connected}) \to e^{-C(R)} \doteq 0.861666,$$

whereas $\Pr(G_n^{(0)}$ *is 2-connected*$) \to 0$, *and* $\Pr(G_n^{(1)}$ *is 2-connected*$) \to 0$.

5.3.2 Number of edges

Let $g_{n,m}, c_{n,m}$ and $b_{n,m}$ be the numbers of labeled outerplanar, connected outerplanar, and two-connected outerplanar graphs on n vertices and m *edges*, respectively, and let $G(x, y)$, $C(x, y)$ and $B(x, y)$ be their exponential bivariate generating functions, for example, $G(x, y) := \sum_{n,m\geq 0} g_{n,m} x^n y^m / n!$.

For any bivariate function $H(x, y)$ we denote the partial derivatives as follows.

$$H_x(x, y) := \frac{\partial H}{\partial x}\,, \quad H_y(x, y) := \frac{\partial H}{\partial y}\,, \quad H_{xx}(x, y) := \frac{\partial}{\partial x}\frac{\partial H}{\partial x}.$$

Two-connected outerplanar graphs. Let K be a convex n-gon where the vertices $\{1, , \ldots, n\}$ are labeled in clockwise order. Let $d_{n,m}$ be the number of dissections of K with m edges and $D(x, y) := \sum_{n,m\geq 0} d_{n,m} x^n y^m$ be its ordinary generating function. As we have seen in the proof of Proposition 5.1.1, for $n \geq 2$, a dissection of K is either an edge or a sequence of k ($k \geq 2$) dissections along the face containing the edge $\{1, 2\}$ where $k - 1$ pairs of vertices are identified. Hence,

$$D(x, y) = yx^2 + y \sum_{k\geq 2} \frac{D(x, y)^k}{x^{k-1}} = yx^2 + \frac{yD^2}{x - D}\,,$$

where the factor y marks the edge $\{1, 2\}$. The only meaningful solution of $D(x, y)$ for this equation is

$$D(x, y) = x\frac{1 + yx - \sqrt{1 - 2(1 + 2y)yx + y^2x^2}}{2(1 + y)}\,.$$

Since $b_{n,m} = d_{n,m} = 0$ for $n \leq 1$, and $b_{2,m} = d_{2,m} = 1$ if $m = 1$ and $b_{2,m} = d_{2,m} = 0$ otherwise, and since $b_{n,m} = \frac{(n-1)!}{2} d_{n,m}$ for $n \geq 3$, we get

$$\begin{aligned} B_x(x,y) &= (D(x,y)/x + yx)/2 \\ &= \frac{1 + (3+2y)yx - \sqrt{1 - 2(1+2y)yx + y^2x^2}}{4(1+y)} \,. \end{aligned} \tag{5.3.1}$$

This can be rewritten as

$$B_x(x,y) = x\phi(B_x(x,y), y)$$

where

$$\phi(u,y) := \frac{-y + (3y+2y^2)u - \sqrt{y^2 - 2(1+y)y^2u + y^2u^2}}{-2 + 4(1+y)u} \,.$$

Since there is $u_0(y) > 0$ within the radius of convergence satisfying

$$\phi(u_0(y), y) - u_0(y)\phi_u(u_0(y), y) = 0 \,,$$

$B_x(x,y)$ has a dominant singularity at $x_0 = u_0(y)/\phi(u_0(y), y)$. Theorem 2.4.3 implies that

$$[x^n y^m] B_x(x,y) = [y^m](1 + O(n^{-1})) \frac{1}{\sqrt{2\pi}} \sqrt{\frac{\phi(u_0(y), y)}{\phi_{uu}(u_0(y), y)}} n^{-3/2} x_0(y)^{-n} \,,$$

and it follows that

$$[x^n y^m] B(x,y) = \frac{1}{n}[x^{n-1}y^m] B_x(x,y) = [y^m](1 + O(n^{-1}))\alpha_2(y) n^{-5/2} x_0(y)^{-n} \,,$$

where $\alpha_2(y) := \frac{1}{\sqrt{2\pi}} \sqrt{\phi(u_0(y), y)/\phi_{uu}(u_0(y), y)}$ and $\alpha_2(1) = \alpha_2$.

Connected outerplanar graphs. Since the number of edges is an inherited parameter, we also have a bivariate version of Proposition 2.2.1:

$$G(x,y) = \exp(C(x,y)) \tag{5.3.2}$$

$$C_x(x,y) = \exp(B_x(xC_x(x,y), y)) \,. \tag{5.3.3}$$

Define $u := F(x,y) := xC_x(x,y)$, and combine Equation (5.3.3) and (5.3.1) as follows.

$$\begin{aligned} &u = x\Phi(u,y) \\ &\text{where } \Phi(u,y) := \exp\left(\frac{1 + (3+2y)yu - \sqrt{1 - 2(1+2y)yu + y^2u^2}}{4(1+y)} \right) . \end{aligned}$$

We can find $u_1(y) > 0$ within the radius of convergence satisfying

$$\Phi(u_1(y), y) - u_1(y)\Phi_u(u_1(y), y) = 0 \,.$$

Moreover, $F(x,y)$ has a dominant singularity at $x_1(y) = u_1(y)/\Phi(u_1(y),y)$, and

$$[x^n y^m]F(x,y) = [y^m](1+O(n^{-1}))\frac{1}{\sqrt{2\pi}}\sqrt{\frac{\Phi(u_1(y),y)}{\Phi_{uu}(u_1(y),y)}}n^{-3/2}x_1(y)^{-n}\,.$$

It follows

$$[x^n y^m]C(x,y) = \frac{1}{n}[x^n y^m]F(x,y) = [y^m](1+O(n^{-1}))\alpha_1(y)n^{-5/2}x_1(y)^{-n}\,,$$

where $\alpha_1(y) := \frac{1}{\sqrt{2\pi}}\sqrt{\Phi(u_1(y),y)/\Phi_{uu}(u_1(y),y)}$ and $\alpha_1(1) = \alpha_1$.

Not necessarily connected outerplanar graphs. Similarly we obtain

$$G(x,y) \sim e^{C(x_1(y),y)} + \alpha_1(y)e^{C(x_1(y),y)}(1-x/x_1(y)^{3/2},$$

where $x_1(1) = R$, $C(x_1(1),1) = C(R)$ and $\alpha_1(1) = \alpha_1$. This yields

$$[x^n y^m]G(x,y) = [y^m](1+o(1))\alpha_G(y)n^{-5/2}x_1(y)^{-n}\,, \tag{5.3.4}$$

for a function $\alpha_0(y) = \alpha_1(y)e^{C(x_1(y),y)}$ with $\alpha_0(1) = \alpha_1 e^{C(R)}$.

The expected number of edges. Let Y_n be the number of edges in a random outerplanar graph on n vertices and $P_n(y)$ be the probability generating function of Y_n, i.e.,

$$P_n(y) := \sum_{m\geq 0}\Pr(Y_n = m)y^m\,. \tag{5.3.5}$$

Then the expectation and variance of Y_n is

$$\mathbb{E}(Y_n) = P_n'(1)\,, \tag{5.3.6}$$
$$\mathrm{Var}(Y_n) = P_n''(1) + P_n'(1) - (P_n'(1))^2. \tag{5.3.7}$$

Define $G_n(y) := [x^n]G(x,y)$. From Definition (5.3.5) and (5.3.4), we get

$$P_n(y) = \sum_{m\geq 0}\frac{g_{n,m}}{g_n}y^m = \frac{G_n(y)}{G_n(1)} = (1+o(1))\frac{\alpha_0(y)}{\alpha_0(1)}\Big(\frac{x_1(1)}{x_1(y)}\Big)^n\,.$$

From (5.3.6) and (5.3.7) we obtain

$$\mathbb{E}(Y_n)/n \sim -\frac{x_1'(1)}{x_1(1)} \doteq 1.56251$$
$$\mathrm{Var}(Y_n)/n \sim -\frac{x_1''(1)}{x_1(1)} - \frac{x_1'(1)}{x_1(1)} + \Big(\frac{x_1'(1)}{x_1(1)}\Big)^2 \doteq 0.223992\,.$$

The same holds for a random connected outerplanar graph. For random two-connected outerplanar graphs, we analogously compute

$$\mathbb{E}(Y_n)/n \sim -\frac{x_0'(1)}{x_0(1)} \doteq 1.707106\,,$$
$$\mathrm{Var}(Y_n)/n \sim -\frac{x_0''(1)}{x_0(1)} - \frac{x_0'(1)}{x_0(1)} + \Big(\frac{x_0'(1)}{x_0(1)}\Big)^2 \doteq 0.176776\,.$$

Indeed Theorem 2.4.4 implies the following.

Theorem 5.3.3. *Let Y_n be the number of edges in a random outerplanar graph on n vertices. Then the distribution of Y_n is asymptotically Gaussian with mean $\mathbb{E}(Y_n) \sim \mu n$ and variance $\mathrm{Var}(Y_n) \sim \sigma^2 n$, where*

$$\mu \doteq 1.56251, \quad \sigma^2 \doteq 0.223992.$$

This also holds for random connected outerplanar graphs with the same μ and σ^2 and for random two-connected outerplanar graphs with $\mu \doteq 1.707106$ and $\sigma^2 \doteq 0.176776$.

5.3.3 Chromatic number

It is easy to see that every outerplanar graphs is three colourable. Indeed more is true.

Theorem 5.3.4. *Let $\chi(G_n)$ denote the chromatic number of a random outerplanar graph G_n on n vertices. Then we have*

$$\lim_{n\to\infty} \Pr(\chi(G_n) = 3) = 1\,.$$

This follows from the fact that the number of labeled bipartite outerplanar graphs on n vertices is asymptotically $c\ 4.40364^n\ n!$ for a constant $c > 0$, which was proven by Löffler [95].

5.4 Recursive counting and uniform sampling

Observe that the block structure of an outerplanar graph is a forest. Thus we can count and generate outerplanar graphs similarly as demonstrated for forests in Section 4.4. The decomposition from a (not-necessarily connected) outerplanar graphs to connected outerplanar graphs is followed by formulas (2.2.1) and (2.2.2). Thus we restrict our attention to connected outerplanar graphs.

In order to decompose a labeled connected outerplanar graph, we consider two cases. The vertex labeled with the smallest label is either a cutvertex and hence it is contained in more than one block, or it is not a cutvertex and hence it is contained in a unique block. Depending on these two cases, we apply a degree-reduction strategy as in the case of labeled trees (see Figure 5.1).

Let $c(n)$ be the number of all labeled connected outerplanar graphs with n vertices $\{1, \cdots, n\}$ and $c_d(n)$ the number of all labeled connected outerplanar graphs with n vertices, where the vertex 1 is adjacent to d blocks. Then, for $n \geq 2$,

$$c(n) \;=\; \sum_{d=1}^{n-1} c_d(n).$$

Note that $\sum_{d=2}^{n-1} c_d(n)$ counts all labeled connected outerplanar graphs with n vertices where the vertex 1 is a cutvertex, and $c_1(n)$ all labeled connected outerplanar graphs with n vertices where the vertex 1 is not a cutvertex.

Let G be a labeled connected outerplanar graph with n vertices where the vertex 1 is a cutvertex and is adjacent to d blocks, $d \geq 2$; see the upper part

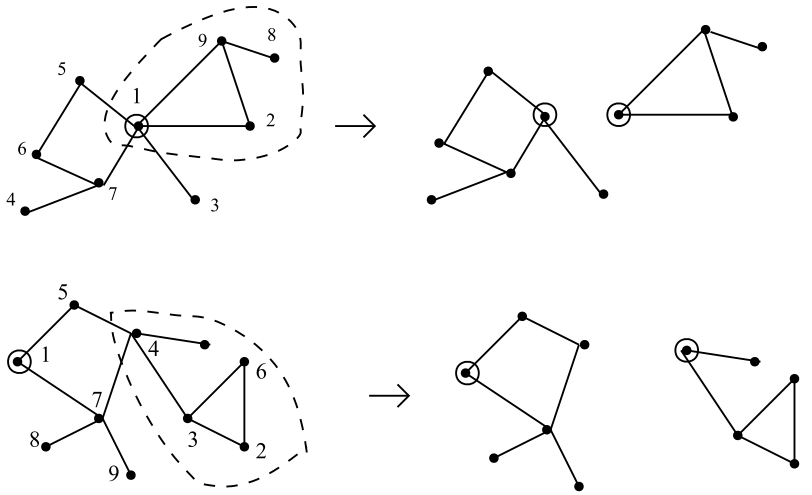


Figure 5.1: Pulling off the petals from the flower.

of Figure 5.1. At the vertex 1 we split off the connected component containing the vertex 2 from G. In the remaining graph the vertex 1 is adjacent to $d-1$ blocks. If the split subgraph has i vertices, then there are $\binom{n-2}{i-2}$ ways to choose a vertex set of the split subgraph since the two vertices 1 and 2 are already contained in the split subgraph. It follows that for $d \geq 2$, $n \geq 3$,

$$c_d(n) \;=\; \sum_{i=2}^{n-d+1} \binom{n-2}{i-2} c_1(i) c_{d-1}(n-i+1).$$

We consider the case that the vertex 1 is not a cutvertex and hence it is contained in a unique block, which we call the *root block*; see the lower part of Figure 5.1. Let $q_c(n)$ be the number of all labeled connected outerplanar graphs with n vertices, where the smallest c vertices of the root block are not cutvertices. Then clearly $c_1(n) \;=\; q_1(n)$.

From such a graph we split off a subgraph attached at the $(c+1)$-th smallest vertex, which might be any kind of outerplanar graph. Then in the remaining graph the $(c+1)$-th smallest vertex of the root block is not a cutvertex. Thus for $c \geq 1$, $n \geq 3$,

$$q_c(n) \;=\; \sum_{i=1}^{n-b+1} \binom{n-1}{i-1} c(i) q_{c+1}(n-i+1).$$

If none of the vertices in the root block of an outerplanar graph with n vertices are not cutvertices, the graph is two-connected and thus $q_n(n) = b_n$.

We have a complete set of recursive formulas that count outerplanar graphs. Table 5.1 shows the exact numbers b_n, c_n, and g_n of labeled two-connected outerplanar graphs, connected outerplanar graphs, and outerplanar graphs, on n vertices up to $n = 16$.

n	b_n	c_n	g_n
1	0	1	1
2	1	1	2
3	1	4	8
4	9	37	63
5	132	602	893
6	2700	14436	19714
7	70920	458062	597510
8	2275560	18029992	22903403
9	86264640	845360028	1056115331
10	3772681920	4593606320	56744710974
11	186972105600	2836966508216	3475626211316
12	10355595465600	196156795008384	238818544070905
13	633892275878400	15008752290350656	18183183610029003
14	42495895579737600	1258841795197091392	1519020289266947462
15	3096545573029708800	114838947237881287800	138117136134012654182
16	243680880958010496000	11319937495659268412416	1357672420635795878040 9

Table 5.1: The exact numbers b_n, c_n, g_n of labeled two-connected outerplanar graphs, connected outerplanar graphs, outerplanar graphs on n vertices, up to $n = 16$.

Uniform sampling. The decomposition and counting formulas presented above give rise to an efficient uniform random generation procedure.

Our sampling procedure first determines the number of components, and how many vertices they shall contain. Each connected component is generated independently from the others, but having the chosen numbers of vertices. To generate a connected component with given numbers of vertices, we decide for a decomposition into 2-connected subgraphs and how the vertices shall be distributed among its parts. For the generation of two-connected outerplanar graphs we use the tree structure of its dual.

Theorem 5.4.1. *Labeled outerplanar graphs on n vertices can be sampled uniformly at random in deterministic time $\tilde{O}(n^4)$ and space $O(n^3 \log n)$. If we apply a preprocessing step, this can also be done in deterministic time $\tilde{O}(n^2)$.*

Brute-force algorithms to generate random outerplanar graphs uniformly at random require exponential time, and Markov chain Monte Carlo methods have unknown mixing times and only approximate the uniform distribution. We have developed a polynomial time generation algorithm for outerplanar graphs, which can be adapted to generate and count labeled outerplanar graphs, connected outerplanar graphs and two-connected outerplanar graphs, uniformly at random. In all these cases, it is also easy to modify the counting formulas and the uniform sampling algorithm for outerplanar graphs with a given number of vertices *and* a given number of edges and also for outerplanar *multigraphs*. The recursive counting formulas and the uniform sampling algorithm are implemented by Löffler [95].

Chapter 6

Labeled Cubic Planar Graphs

In this chapter we decompose labeled cubic planar graphs along the connectivity structure, and derive the asymptotic number by interpreting the decomposition in terms of generating functions and then by applying the singularity analysis.

For the decomposition, we make use of a *rooted* cubic graph with one distinguished oriented edge, and decompose rooted connected cubic graphs into smaller parts up to rooted 3-connected cubic graphs. To complete the counting and generation procedure, it suffices to consider 3-connected cubic graphs, because no cubic graph is 4-connected. For 3-connected cubic graphs, we can use their dual, i.e., triangulations.

Based on the decomposition, we derive the equations of generating functions and apply the *resultant* method suggested by Flajolet and Sedgewick [63]. We show that the number of labeled cubic planar graphs on n vertices is asymptotically $c\, n^{-7/2}\, \rho^{-n}\, n!$, for a suitable positive constant c and $\rho^{-1} \doteq 3.132595$.

Using the asymptotic number, we also study the typical properties of a random cubic planar graph that holds when the number of vertices converges to infinity, e.g., the chromatic number. To this end, we first show that the number of isolated K_4's in a random cubic planar graph has asymptotically Poisson distribution with mean $\rho^4/4!$ and that a random cubic planar graph contains linearly many triangles with probability tending to one. As a consequence, together with Brooks' theorem, we can see that the chromatic number of a random cubic planar graph is four with probability bounded away from zero and one, and that the chromatic number of every connected component with more than four vertices in a random cubic planar graph is three with probability tending to one.

Using a complete set of recursive counting formulas, we derive a deterministic uniform generation of cubic planar graphs from the general principle. Furthermore, we can compute the exact numbers of cubic planar graphs according to the connectivity computed from the recursive enumeration.

The rest of the chapter is organized as follows: In Section 6.1, we introduce necessary terminologies and the decomposition theorem for rooted cubic planar graphs. In Section 6.2, we interpret the decomposition in terms of generating functions. In Section 6.3, we provide the relation between 3-connected cubic pla-

nar graphs and triangulations and derive counting formulas for triangulations. In Sections 6.4 and 6.5, we derive the equations of generating functions, get the asymptotic number of labeled cubic planar graphs, and study some properties of a random cubic planar graph. In Section 6.6, we derive recursive counting formulas based on the decomposition theorem, and discuss the uniform generation algorithm.

6.1 Rooted cubic planar graphs

To count labeled cubic planar (simple) graphs, we introduce 'rooted' cubic graphs. We will present a decomposition scheme for such graphs, which can then be used to count (unrooted) cubic planar (simple) graphs.

A *rooted cubic graph* $G = (V, E, st)$ consists of a connected cubic multigraph $G = (V, E)$ and an ordered pair of adjacent vertices s and t such that the *underlying graph* G^- obtained by deleting an edge between s and t is simple. Thus in G, if s and t are distinct there may be either one or two edges between them, and if $s = t$ there is a loop at this vertex, and otherwise there are no loops or parallel edges. The oriented edge st is called the *root* of G, and s and t the *poles*. Thus G^- is obtained from G by deleting the root edge. Note that a rooted cubic graph must have at least 4 vertices: we may not have a 'triple edge'.

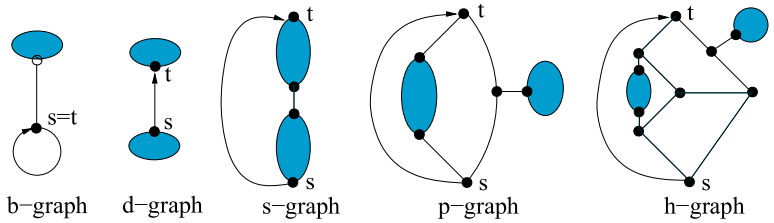


Figure 6.1: The five types of rooted cubic graphs.

The following lemma is easily checked.

Lemma 6.1.1. *A rooted cubic graph $G = (V, E, st)$ has exactly one of the following types.*

- *b: the root is a loop.*
- *d: G^- is disconnected.*
- *s: G^- is connected but there is a cut edge in G^- that separates s and t.*
- *p: G^- is connected, there is no cut edge in G^- separating s and t, and either st is an edge of G^- or $G\backslash\{s,t\}$ is disconnected.*
- *h: G^- is connected, there is no cut-edge in G^- separating s and t, G is simple and $G\backslash\{s,t\}$ is connected.*

We will make use of a *replacement* operation for rooted cubic graphs. We are often interested in rooted cubic graphs which are not d-graphs, i.e., b-, s-, p- or h-graphs: let us call these c-graphs. Let $G = (V_G, E_G, s_G t_G)$ be a rooted cubic graph, let $u_G v_G$ be obtained by orienting an edge in G^-, and let $H = (V_H, E_H, s_H t_H)$ be a c-graph. The rooted cubic graph G' obtained from G by the *replacement* of $u_G v_G$ by H has vertex set the disjoint union of V_G and V_H, edge set the disjoint union of $E_G - \{u_G v_G\}$ and $E_H - \{s_H t_H\}$ together with the edges $u_G s_H$ and $v_G t_H$, and the same root as G. When we perform a replacement by H we always insist that H is a c-graph. The following result may be compared with network decomposition results of Trakhtenbrot [136, 141].

Theorem 6.1.2. *(a) Let H be a 3-connected simple rooted cubic graph, let F be a set of oriented edges of H^-, and for each $uv \in F$ let H_{uv} be a c-graph. Let G be obtained by replacing the edges $uv \in F$ by H_{uv}. Then G is an h-graph. Further, if H is planar and each H_{uv} is planar then so is G.*

(b) Let $G = (V, E, st)$ be an h-graph. Then there is a unique 3-connected rooted cubic graph H (called the core *of G) such that we can obtain G by replacing some oriented edges e of H^- by c-graphs H_e. Further H is simple, and if G is planar then so is H and each H_e.*

Proof. (a) Note that H is an h-graph; and if G' is an h-graph and we replace an oriented edge by a c-graph then we obtain another h-graph (which is planar if both the initial and the replacing graph are). Thus part (a) follows by induction on the number of edges replaced.

(b) The main step is to identify the core H. Let W be the set of vertices $v \in V \backslash \{s, t\}$ such that there is a set of three pairwise internally vertex-disjoint (or equivalently, edge-disjoint) paths between v and $\{s, t\}$. Then W is non-empty. For, let P_1 and P_2 be internally vertex-disjoint paths between s and t in G^-. There must be a path Q between an internal vertex of P_1 and an internal vertex of P_2 (since neither P_1 nor P_2 is just a single edge, and $G \backslash \{s, t\}$ is connected), and we can insist that Q be internally vertex-disjoint from P_1 and P_2. Now the terminal vertices of Q must both be in W.

Let H be the graph with vertex set $V_H = W \cup \{s, t\}$, where for distinct vertices u and v in V_H we join u and v in H if there is a $u - v$ path in G using no other vertices in V_H. Thus in particular if vertices $u, v \in V_H$ are adjacent in G then they are adjacent also in H.

It is easy to check that H is 3-connected, and thus also is simple.

Let X be the set of vertices of G not in H. If $X = \emptyset$ then $G = H$ and we are done: suppose then that X is non-empty. Consider a component C of the subgraph of G induced by X. We claim that there are distinct vertices u and v in V_H which are adjacent in H but not in G, vertices x and y in C (possibly $x = y$) and edges ux and vy in G which are the only edges between C and V_H. Let H_{uv} be the rooted cubic graph obtained from C by adding the root edge xy. Now it is clear that we may obtain G by starting with H and replacing any edge uv of H not in G by the corresponding H_{uv}.

We have now seen that the rooted cubic graph H is simple and 3-connected, and we may obtain G by starting with H and replacing some edges e of H^- by c-graphs H_e. Finally it is easy to see that H is unique. For if H' also has these properties, then we immediately see that $V_H = V_{H'}$, and it follows easily that the graphs are the same. □

We are interested here only in planar graphs. However, all results in Sections 6.1 and 6.6 can be formulated more generally for subclasses of connected cubic graphs that are closed under replacements.

6.2 Exponential generating functions

We let b_n, d_n, s_n, p_n, h_n, and c_n be the number of b-, d-, s-, p-, h-, and c-graphs on n vertices, respectively. Thus $c_n = b_n + s_n + p_n + h_n$. Further we let $B(x), D(x), S(x), P(x), H(x)$, and $C(x)$ be the corresponding exponential generating functions. For instance, $B(x)$ is defined by

$$B(x) := \sum_{n \geq 0} \frac{b_n}{n!}\, x^n \,.$$

Note that $b_n = d_n = s_n = p_n = h_n = c_n = 0$ for all odd n, due to cubicity, also for $n = 0$ by convention, and for $n = 2$. Thus, for instance, $B(x)$ is of the form $\sum_{n \geq 2} \frac{b_{2n}}{(2n)!} x^{2n}$.

b-graphs. The structure of a b-graph is restricted by 3-regularity, and the shaded area in Figure 6.2 below together with an oriented edge between u and v is a d-, s-, p-, or h-graph. Therefore, $B(x) = x^2/2\,(D(x)+S(x)+P(x)+H(x))$, where the factor $1/2$ is due to the orientation of the edge between u and v. This can be rewritten as $B(x) = x^2\,(D(x) + C(x) - B(x))\,/\,2$.

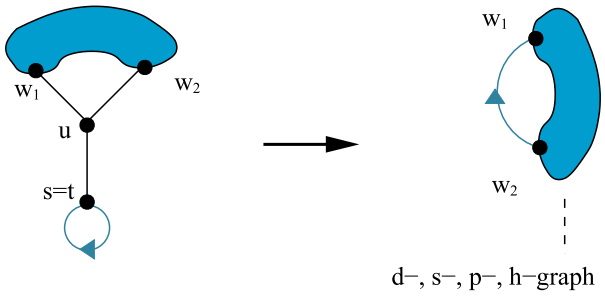


Figure 6.2: Decomposition of a b-graph.

d-graphs. A d-graph can be decomposed uniquely into two b-graphs as shown in Figure 6.3. We therefore have $D(x) = B(x)^2\,/\,x^2$.

s-graphs. For a given s-graph G, the graph G^- has a cut-edge that separates s and t and that is closest to s as in Figure 6.4. (Note that the cut edge could be a second copy of st.) We obtain $S(x) = (S(x) + P(x) + H(x) + B(x))\,(P(x) + H(x) + B(x)) = C(x)^2 - C(x)S(x)$.

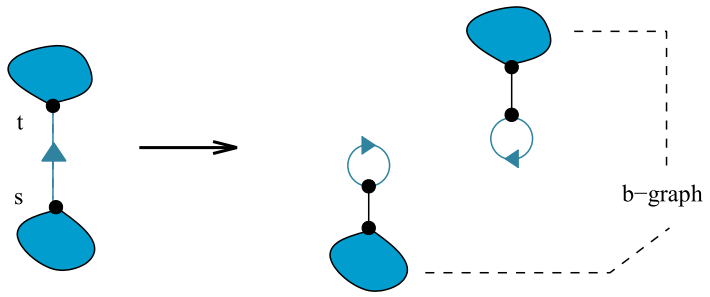


Figure 6.3: Decomposition of a d-graph.

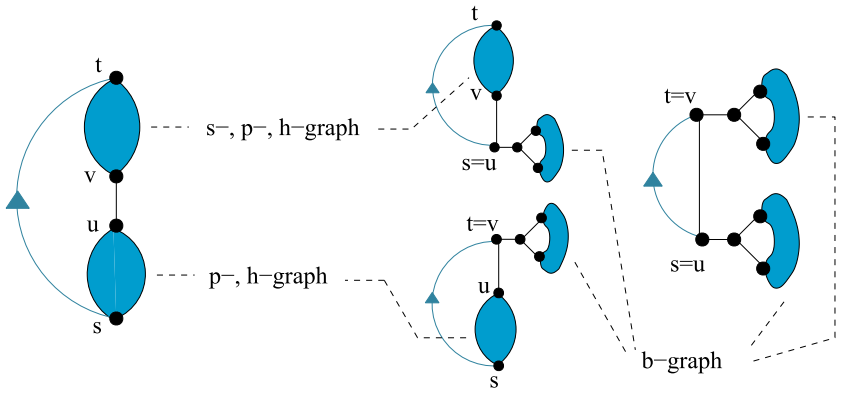


Figure 6.4: Decomposition of an s-graph.

p-graphs. For a given p-graph, we distinguish whether or not s and t are adjacent in G^-. Both situations are depicted in Figure 6.5. We obtain $P(x) = x^2\,(S(x)+P(x)+H(x)+B(x))+x^2/2\,(S(x)+P(x)+H(x)+B(x))^2 = x^2C(x)+x^2C(x)^2\,/\,2$, where the factor $1/2$ in the latter term is there because two c-graphs are not ordered.

h-graphs. From Theorem 6.1.2 we know that an h-graph is built from a rooted three-connected cubic planar graph by replacing some edges, except the root edge, by b-, s-, p-, or h-graphs, i.e., c-graphs, see Figure 6.6. Let $m_{n,l}$ be the number of labeled rooted 3-connected cubic planar graphs on n vertices and l edges and let

$$M(x,y) := \sum_{n,l\geq 0} \frac{m_{n,l}}{n!}\, x^n\, y^l$$

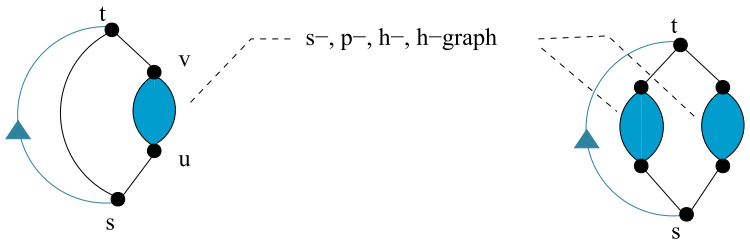


Figure 6.5: Decomposition of a p-graph.

be its exponential generating function. Clearly $m_{n,l} = 0$ for odd n, $n = 0, 2$ or $l \neq 3n/2$ since a cubic planar graph on n vertices has $3n/2$ edges. Hence

$$M(x, y) = \sum_{n \geq 2} \frac{m_{2n,3n}}{(2n)!} x^{2n} y^{3n},$$

which we will determine in Section 6.3 (see Equation (6.3.2)).

Note that the variable y in $M(x, y)$ marks the edges in rooted 3-connected cubic graphs. Thus in order to derive the exponential generating function for h-graphs, we replace the variable y in $M(x, y)$ by $C(x)+1$, (where the constant term 1 is because there is no change in a rooted 3-connected cubic graph when an edge is replaced by an edge itself,) and divide this by $C(x) + 1$, because we do not replace the root edge of a rooted 3-connected cubic planar graph. Thus we get

$$H(x) = \frac{M(x, (C(x) + 1))}{(C(x) + 1)}\,. \tag{6.2.1}$$

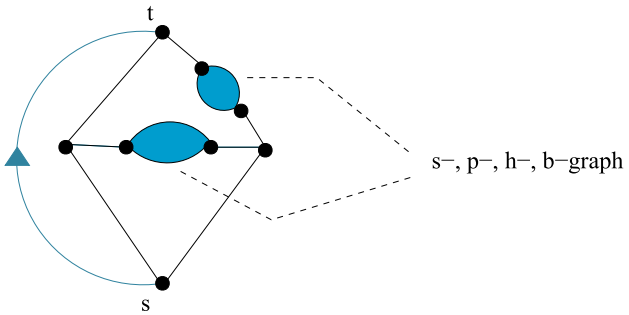


Figure 6.6: Decomposing of an h-graph along its core.

Cubic planar graphs. For $k = 0, 1, 2, 3$ let $g_n^{(k)}$ be the number of k-vertex connected cubic planar (simple) graphs on n vertices and $G^{(k)}(x)$ be the corre-

sponding exponential generating functions. Note that $g_n^{(k)} = 0$ for odd n and also for $n = 0, 2$ except that we set $g_0^{(0)} = 1$ by convention.

If we select an arbitrary edge in a connected cubic planar (simple) graph and orient this edge, we obtain a rooted cubic graph $G = (V, E, st)$ that is neither a b-graph, nor an s- or p-graph where s and t are adjacent in the underlying graph G^-, see Figure 6.7. Note that the number of connected cubic planar (simple) graphs with one distinguished oriented edge is counted by $3x\,\frac{dG^{(1)}(x)}{dx}$, and the number of s- (resp. p-)graphs $G = (V, E, st)$ where s and t are adjacent in G^- as depicted in the middle (resp. right) picture in Figure 6.7 is counted by $B(x)^2$ (resp. $x^2\,C(x)$). Therefore we get

$$3x\,\frac{dG^{(1)}(x)}{dx} = D(x) + S(x) + P(x) + H(x) - B(x)^2 - x^2\,C(x)\,. \qquad (6.2.2)$$

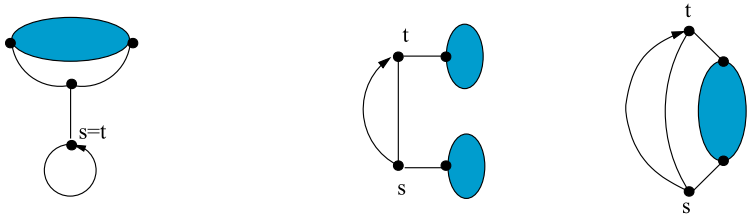


Figure 6.7: Types of rooted cubic graphs that are not simple.

As we have seen in (2.2.3), the exponential generating function for connected cubic planar graphs and that for not necessarily connected ones are related as follows.

$$G^{(0)}(x) = \exp(G^{(1)}(x))\,. \qquad (6.2.3)$$

6.3 Three-connected cubic planar graphs

The number of labeled three-connected cubic planar graphs is closely related to that of *rooted triangulations*. A rooted triangulation is an edge-maximal plane graph with a distinguished directed edge on the outer face, called the *root edge*. Tutte [137] derived exact and asymptotic formulas for the number of such objects up to isomorphisms that preserve the outer face and the root edge. Since such graphs do not have non-trivial automorphisms that fix the root edge, we can obtain the number of labeled objects from the number of unlabeled objects. Labeled three-connected planar graphs with at least four vertices have exactly two non-equivalent embeddings in the plane. Using plane duality, we can compute the number of rooted three-connected cubic planar graphs from the number of rooted triangulations.

Let t_n be the number of unlabeled rooted triangulations on $n + 2$ vertices. From the formulas Tutte computed for unlabeled rooted triangulations on $n+3$ vertices, it follows that the ordinary generating function $T(z)$ for t_n, i.e., $T(z) =$

$\sum_{n\geq 1} t_n\, z^n$, satisfies the following.

$$\begin{aligned} T(z) &= u\,(1-2u) \\ z &= u\,(1-u)^3\,. \end{aligned} \tag{6.3.1}$$

The first terms of $T(z)$ are $z+z^2+3z^3+13z^4+68z^5+399z^6+\ldots$. Further, $T(z)$ has a dominant singularity at $\xi = 27/256$ and the asymptotic growth of t_n is $\alpha_4\, n^{-5/2}\, \xi^{-n}\, n!$, where α_4 is a constant. Let $\tilde{T}(x,y)$ be the corresponding ordinary generating function, but where x marks two times the number of faces and y marks three times the number of edges. By Euler's formula, a triangulation on $n+2$ vertices has $2n$ faces and $3n$ edges. Therefore, $\tilde{T}(x,y) := \sum_{n\geq 1} t_n x^{2n} y^{3n}$ can be computed by $\tilde{T}(x,y) = T(x^2y^3)$.

We now determine the exponential generating function $M(x,y)$ for the number of labeled rooted 3-connected cubic graphs, which was needed in the decomposition of h-graphs in Section 6.1. Since the dual of a 3-connected cubic map on $2n$ vertices is a triangulation on $n+2$ vertices (and hence with $2n$ faces and $3n$ edges), we have $m_{2n,3n} = (2n)!\, t_n\,/\,2$ for $n \geq 2$. We therefore obtain

$$M(x,y) = \sum_{n\geq 2} \frac{m_{2n,3n}}{(2n)!} x^{2n} y^{3n} = \frac{1}{2}\left(\tilde{T}(x,y) - x^2y^3\right) = \frac{1}{2}\left(T(x^2y^3) - x^2y^3\right). \tag{6.3.2}$$

Thus $M(x,y) = (x^4y^6 + 3x^6y^9 + 13x^8y^{12} + 68x^{10}y^{15} + 399x^{12}y^{18} + \ldots)/2$. Furthermore the dominant singularity of $M(x) = M(x,1) = 1/2\,(T(x^2) - x^2)$ is the square-root of the dominant singularity of $T(z)$ and the asymptotic growth of m_n with n even is $\alpha_3 n^{-5/2}\theta^{-n} n!$, where $\theta = 3\sqrt{3}/16$ and α_3 is a constant.

6.4 Singularity analysis

We summarize the equations derived so far.

$$B(x) = x^2(D(x) + C(x) - B(x))/2 \tag{6.4.1}$$
$$C(x) = S(x) + P(x) + H(x) + B(x) \tag{6.4.2}$$
$$D(x) = B(x)^2/x^2 \tag{6.4.3}$$
$$S(x) = C(x)^2 - C(x)S(x) \tag{6.4.4}$$
$$P(x) = x^2C(x) + x^2C(x)^2/2\,. \tag{6.4.5}$$

We can also describe the substitution in Equation (6.2.1) for $H(x)$ algebraically, using Equations (6.3.1) and (6.3.2).

$$2(C(x)+1)H(x) = u(1-2u) - u(1-u)^3 \tag{6.4.6}$$
$$x^2(C(x)+1)^3 = u(1-u)^3\,. \tag{6.4.7}$$

Using algorithms for computing resultants and factorizations (these are standard procedures in e.g., Maple or Mathematica), we can obtain a single algebraic equation $Q(C(x),x) = 0$ from equations (6.4.1) – (6.4.7) that describes the generating function $C(x)$ uniquely, given sufficiently many initial terms of c_n.

This is in principle also possible for all other generating functions involved in the above equations; however, the computations turn out to be more tedious, whereas the computations to compute the algebraic equation for $C(x)$ are manageable.

From this equation, following the discussion in Section VII.4 in [63], one can obtain the two dominant singularities ρ and $-\rho$ of $C(x)$, where ρ is an analytic constant and the first digits are $\rho \doteq 0.319224$. We can also compute the expansion at the dominant singularity ρ. Changing the variables $Y = C(x) - C(\rho)$ and $X = x - \rho$ in $Q(C(x), x) = 0$, one can symbolically verify that the equation $Q(C(x), x) = 0$ can be written in the form

$$(aY + bX)^2 = pY^3 + qXY^2 + rX^2Y + sX^3 + \text{ higher order terms},$$

where a, b, p, q, r, s are constants that are given analytically. This implies the following expansion of $C(x)$ near the dominant singularity ρ.

$$C(x) = C(\rho) + b\rho/a\ (1 - x/\rho) + \beta_1(1 - x/\rho)^{3/2} + O((1 - x/\rho)^2), \qquad (6.4.8)$$

where $\beta_1 := \rho^{3/2}/a\ \sqrt{p(b/a)^3 - q(b/a)^2 + r(b/a) - s}$ is a positive constant. For large n, the coefficient c_n^+ of x^n on the right hand side satisfies

$$c_n^+ \ \sim\ \beta_2\ n^{-5/2}\ \rho^{-n}\ n!\,,$$

where $\beta_2 = \beta_1/\Gamma(3/2) = 2\beta_1/\sqrt{\pi}$. Similarly we get the expansion at the dominant singularity $-\rho$

$$C(x) = C(\rho) + b\rho/a\ (1 + x/\rho) + \beta_1(1 + x/\rho)^{3/2} + O((1 + x/\rho)^2),$$

and for large n, the coefficient c_n^- of x^n on the right hand side satisfies

$$c_n^- \ \sim\ \beta_2\ n^{-5/2}\ (-\rho)^{-n}\ n!\,.$$

Following Theorem VI.8 [63], the asymptotic number c_n is then the summation of these two contributions c_n^+ and c_n^-, and thus for large even n

$$c_n \ \sim\ 2\beta_2\ n^{-5/2}\ \rho^{-n}\ n!\,,$$

whereas $c_n = 0$ for odd n.

Since the generating functions for $B(x), D(x), S(x), P(x), H(x)$ are related with $C(x)$ by algebraic equations, they all have the same dominant singularities ρ and $-\rho$. The singular expansion of $G^{(1)}(x)$ can be obtained from Equation (6.2.2) through a term-by-term integration, and thus we obtain the singular expansions at ρ and $-\rho$

$$G^{(1)}(x) = G^{(1)}(\rho) + c(1 - x/\rho)^2 + \beta_3(1 - x/\rho)^{5/2} + O((1 - x/\rho)^3)\,, \qquad (6.4.9)$$
$$G^{(1)}(x) = G^{(1)}(\rho) + c(1 + x/\rho)^2 + \beta_3(1 + x/\rho)^{5/2} + O((1 + x/\rho)^3)\,, \qquad (6.4.10)$$

where c and β_3 are analytically given constants. Thus for an analytically given constant α_1 and for large even n we get

$$g_n^{(1)} \ \sim\ \alpha_1\ n^{-7/2}\ \rho^{-n}\ n!\,,$$

whereas $g_n^{(1)} = 0$ for odd n.

Because of Equation (6.2.3), the generating functions $G^{(0)}(x)$ and $G^{(1)}(x)$ have the same dominant singularities ρ and $-\rho$, and indeed we may see that $g_n^{(1)}/g_n^{(0)} \to e^{-\lambda}$ where $\lambda = G^{(1)}(\rho)$. Based on the above decomposition it is also easy to derive equations for the exponential generating function $G^{(2)}(x)$ for the number of biconnected cubic planar graphs, which has a slightly larger radius of convergence η (whose first digits are 0.319521).

We finally obtain the following.

Theorem 6.4.1. *The asymptotic number of labeled cubic planar graphs, labeled connected cubic planar graphs, labeled 2-connected cubic planar graphs, and labeled 3-connected cubic planar graphs is given by the following. For large even n*

$$\begin{aligned} g_n^{(0)} &\sim \alpha_0\ n^{-7/2}\ \rho^{-n}\ n! \\ g_n^{(1)} &\sim \alpha_1\ n^{-7/2}\ \rho^{-n}\ n! \\ g_n^{(2)} &\sim \alpha_2\ n^{-7/2}\ \eta^{-n}\ n! \\ g_n^{(3)} &\sim \alpha_3\ n^{-7/2}\ \theta^{-n}\ n!\,. \end{aligned}$$

All constants are analytically given. Also $\alpha_1/\alpha_0 = e^{-\lambda}$ where $\lambda = G^{(1)}(\rho)$. Further $\rho^{-1} \doteq 3.132595$, $\eta^{-1} \doteq 3.129684$, and $\theta^{-1} \doteq 3.079201$.

6.5 Random cubic planar graphs

In this section, we use Theorem 6.4.1 to investigate the connectedness, components and the chromatic number of a random cubic planar graph. Throughout the section, for $k = 0, 1, 2, 3$ let $G_n^{(k)}$ denote a random graph chosen uniformly at random among all the k-vertex-connected cubic planar graphs on vertices $1, \ldots, n$ for even n.

6.5.1 Connectedness

We use Theorem 6.4.1 to investigate the connectivity and the chromatic number of a random cubic planar graph. Along the way we consider components and subgraphs. Let α_0 and α_1 be as in Theorem 6.4.1.

Theorem 6.5.1. *Let $\lambda = G^{(1)}(\rho)$. As $n \to \infty$ with n even,*

$$\Pr(G_n^{(0)} \text{ is connected}) \to e^{-\lambda},$$

whereas $\Pr(G_n^{(0)}$ *is 2-connected*$) \to 0$, $\Pr(G_n^{(1)}$ *is 2-connected*$)) \to 0$ *and* $\Pr(G_n^{(2)}$ *is 3-connected*$) \to 0$.

Proof. From Theorem 6.4.1, we see that as $n \to \infty$ with n even

$$\Pr(G_n^{(0)} \text{ is connected}) = g_n^{(1)}/g_n^{(0)} \to \alpha_1/\alpha_0 = e^{-\lambda}.$$

Also,

$$\Pr(G_n^{(0)} \text{ is 2-connected}) = g_n^{(2)}/g_n^{(0)} \sim \alpha_2/\alpha_0(\eta/\rho)^{-n} \to 0,$$

with a similar proof in the other cases. □

Using the numbers in Table 6.2 we compute the probability that $G_n^{(0)}$ is connected, for even n from 10 to 20, in the following Table.

n	10	12	14	16	18	20
$g_n^{(1)}/g_n^{(0)}$	0.997837	0.997982	0.998117	0.998249	0.998368	0.998472

Table 6.1: The probability that a random cubic planar graph on n vertices is connected, from $n = 10$ to $n = 20$.

6.5.2 Components

In order to discuss colouring later we need to find the limiting probability that $G_n^{(0)}$ has a component isomorphic to K_4. Here we consider a more general problem.

Lemma 6.5.2. *Let H be a given connected cubic planar graph, and let $\lambda_H = \frac{\rho^{v_H}}{\mathrm{Aut}(H)}$, where ρ is as in Theorem 6.4.1, v_H denotes the number of vertices in H (and hence even), and $\mathrm{Aut}(H)$ denotes the size of its automorphism group. Let the random variable $X_H = X_H(n)$ be the number of components of $G_n^{(0)}$ isomorphic to H for even n. Then X_H has asymptotically the Poisson distribution $\mathrm{Po}(\lambda_H)$ with mean λ_H; that is, for $k = 0, 1, 2, \ldots$*

$$\Pr(X_H(n) = k) \to e^{-\lambda_H} \frac{\lambda_H^k}{k!} \quad \text{as } n \to \infty.$$

In particular, the probability that $G_n^{(0)}$ has at least one component isomorphic to H tends to $1 - e^{-\lambda_H}$ as $n \to \infty$ with n even.

Proof. Note first that the number of different labelings of H is counted by $v_H!/\mathrm{Aut}(H)$. Correspondingly, the number of ways to construct exactly k components isomorphic to H on kv_H vertices is equal to

$$\frac{(kv_H)!}{k!\mathrm{Aut}(H)^k}\,. \tag{6.5.1}$$

The exponential generating function for the connected cubic planar graphs that are not isomorphic to H is given by

$$G^{(1)}(x) - \frac{x^{v_H}}{\mathrm{Aut}(H)}\,.$$

Thus the number of cubic planar graphs on $n - kv_H$ vertices that do not contain a component isomorphic to H equals

$$[x^{n-kv_H}] \exp\left(G^{(1)}(x) - \frac{x^{v_H}}{\mathrm{Aut}(H)}\right).$$

As $n \to \infty$ with n even this goes to

$$\begin{aligned} & \exp\left(-\frac{\rho^{v_H}}{\mathrm{Aut}(H)}\right) [x^{n-kv_H}] \exp(G^{(1)}(x)) \\ = \ & \exp\left(-\frac{\rho^{v_H}}{\mathrm{Aut}(H)}\right) [x^{n-kv_H}] G^{(0)}(x) \\ \sim \ & \exp\left(-\frac{\rho^{v_H}}{\mathrm{Aut}(H)}\right) \alpha_0 (n - kv_H)^{-7/2} \rho^{-n+kv_H} (n - kv_H)!\,. \end{aligned} \tag{6.5.2}$$

Putting (6.5.1) and (6.5.2) together, we obtain that for $k = 0, 1, 2, \ldots$ the probability that the number of components isomorphic to H in $G_n^{(0)}$ is exactly k is asymptotically

$$\begin{aligned}
& \Pr(X_H = k) \\
\sim\ & \binom{n}{kv_H} \frac{(kv_H)!}{k!\mathrm{Aut}(H)^k} \exp\left(-\frac{\rho^{v_H}}{\mathrm{Aut}(H)}\right) \\
& \qquad \times \frac{\alpha_0 (n-kv_H)^{-7/2} \rho^{-n+kv_H} (n-kv_H)!}{[x^n]G^{(0)}(x)} \\
\sim\ & \frac{n!}{(kv_H)!(n-kv_H)!} \frac{(kv_H)!}{k!\mathrm{Aut}(H)^k} \exp\left(-\frac{\rho^{v_H}}{\mathrm{Aut}(H)}\right) \\
\times\ & \frac{\alpha_0 (n-kv_H)^{-7/2} \rho^{-n+kv_H} (n-kv_H)!}{\alpha_0 n^{-7/2} \rho^{-n} n!} \\
\sim\ & \exp\left(-\frac{\rho^{v_H}}{\mathrm{Aut}(H)}\right) \frac{\rho^{kv_H}}{k!\mathrm{Aut}(H)^k} \\
=\ & e^{-\lambda_H} \frac{\lambda_H^k}{k!}.
\end{aligned}$$

□

An alternative proof of the last result can follow the lines of the proof of Theorem 5.6 of [99]. Both proofs generalise to yield the following extension of that result.

Lemma 6.5.3. *Let $H_1, \ldots, H_m$ be given pairwise non-isomorphic connected cubic planar graphs; and as before let $\lambda_{H_i} = \rho^{v_{H_i}}/\mathrm{Aut}(H_i)$ and let the random variable $X_{H_i} = X_{H_i}(n)$ be the number of components of $G_n^{(0)}$ isomorphic to H_i, where n is even. Then $X_{H_1}, \ldots, X_{H_m}$ are asymptotically jointly distributed like independent random variables $\mathrm{Po}(\lambda_{H_1}), \ldots, \mathrm{Po}(\lambda_{H_m})$, and so the total number of components isomorphic to some H_i is asymptotically $\mathrm{Po}(\sum_i \lambda_{H_i})$.*

Let us observe here that if $H_1, H_2, \ldots$ is an enumeration of all the pairwise non-isomorphic connected cubic planar graphs, then $\sum_i \lambda_{H_i} = G^{(1)}(\rho)$. For

$$\begin{aligned}
G^{(1)}(\rho) \qquad &= \sum_n g_n^{(1)} \cdot \frac{1}{n!}\rho^n = \sum_n \sum_{i: v_{H_i} = n} \frac{n!}{\mathrm{Aut}(H_i)} \cdot \frac{1}{n!}\rho^n \\
&= \sum_i \frac{\rho^{v_{H_i}}}{\mathrm{Aut}(H_i)} = \sum_i \lambda_{H_i}. \qquad (6.5.3)
\end{aligned}$$

Next we want to show that $G_n^{(0)}$ usually has a giant component.

Lemma 6.5.4. *For any $\varepsilon > 0$ there exists t such that the probability is less than ε that each component in $G_n^{(0)}$ has order at most $n - t$.*

Proof. Let $\mathcal{C}(n)$ denote the set of labeled cubic planar (simple) graphs on the vertices $1, \ldots, n$ and so $|\mathcal{C}(n)| = g_n^{(0)}$. By Theorem 6.4.1, there are constants $\alpha > 0$ and $\beta > 1$ such that

$$g_n^{(0)} \sim \alpha n^{-\beta} \rho^{-n} n!$$

as $n \to \infty$ with n even. Thus there is an n_0 such that for all even $n \geq n_0$

$$\frac{1}{2}\alpha n^{-\beta}\rho^{-n}n! \leq g_n^{(0)} \leq 2\alpha n^{-\beta}\rho^{-n}n!.$$

Let t be a positive integer at least n_0 sufficiently large that

$$8\alpha \cdot 2^{\beta} \cdot \frac{(t-1)^{-(\beta-1)}}{\beta-1} < \varepsilon.$$

The reason for this choice will of course emerge shortly. Let $\mathcal{D}(n)$ be the set of graphs $G \in \mathcal{C}(n)$ such that each component has order at most $n-t$. Then for even $n \geq 3t$,

$$\begin{aligned}
|\mathcal{D}(n)| &\leq \sum_{j=t}^{n/2} \binom{n}{j} g_j^{(0)} g_{n-j}^{(0)} \\
&\leq 4\alpha^2 \rho^{-n} n! \sum_{j=t}^{n/2} j^{-\beta}(n-j)^{-\beta} \\
&\leq 4\alpha^2 \rho^{-n} n! \left(\frac{n}{2}\right)^{-\beta} \sum_{j=t}^{n/2} j^{-\beta} \\
&\leq 8\alpha\ g_n^{(0)}\ 2^{\beta} \sum_{j=t}^{n/2} j^{-\beta}.
\end{aligned}$$

But

$$\sum_{j=t}^{n/2} j^{-\beta} \leq \int_{t-1}^{n/2} x^{-\beta} dx < \frac{(t-1)^{-(\beta-1)}}{\beta-1}.$$

Thus our choice of t yields $|\mathcal{D}(n)|/g_n^{(0)} < \varepsilon$ as required. □

Theorem 6.5.5. *The number of components of $G_n^{(0)}$ is asymptotically $1+\mathrm{Po}(\lambda)$, where $\lambda = G^{(1)}(\rho)$.*

Observe that this theorem shows again (as in Theorem 4) that the probability that $G_n^{(0)}$ is connected tends to $e^{-\lambda}$ as $n \to \infty$.

Proof. We may use Lemmas 6.5.3 and 6.5.4, together with (15), and follow the lines of the proof of Theorem 5.5 of [99]. □

6.5.3 Triangles

In order to discuss colouring later we also need to know about triangles, in particular the unsurprising result that $G_n^{(k)}$ usually contains at least one triangle. In fact far more is true.

Lemma 6.5.6. *Let $Y_n^{(k)}$ be the number of triangles in $G_n^{(k)}$. Then for even n there exists $\delta > 0$ such that*

$$\Pr(Y_n^{(k)} \geq \delta n) = 1 - e^{-\Omega(n)}.$$

Proof. Let us consider $Y_n^{(0)}$: the other cases are very similar. Let $\delta > 0$ be sufficiently small that

$$\frac{\rho^2(1-4\delta)}{4e\delta} \geq 2.$$

By Theorem 6.4.1 there exist constants $\alpha > 0$, $\beta > 1$, and $n_0 \geq 2/\delta$ such that for all even $n \geq n_0$

$$\frac{1}{2}\alpha n^{-\beta}\rho^{-n}n! \leq g_n^{(0)} \leq 2\alpha n^{-\beta}\rho^{-n}n!. \tag{6.5.4}$$

Assume for a contradiction that for some even $n \geq n_0$

$$\Pr(Y_n^{(0)} \leq \delta n) \geq e^{-\delta n}. \tag{6.5.5}$$

We shall avoid using round-down $\lfloor x \rfloor$ and round-up $\lceil x \rceil$ in order to keep our formulae readable. Consider the following construction of cubic planar graphs on vertices $1, \cdots, n + 2\delta n$:

- pick an ordered list of $2\delta n$ *special* vertices, say $s_1, s_2, \cdots, s_{2\delta n}$; there are $\frac{(n+2\delta n)!}{n!}$ choices
- take a cubic planar graph G on the remaining n vertices with at most δn triangles; there are at least $e^{-\delta n} g_n^{(0)} \geq e^{-\delta n}\frac{1}{2}\alpha n^{-\beta}\rho^{-n}n!$ choices, by (6.5.4) and (6.5.5)
- pick a set of δn vertices in G that form an independent set and list them in increasing order, say $v_1, v_2, \cdots, v_{\delta n}$; the number of choices is at least

$$\frac{n(n-4)\cdots(n-4\delta n+4)}{(\delta n)!} \geq \frac{n^{\delta n}(1-4\delta)^{\delta n}}{(\delta n)!} \geq \left(\frac{1-4\delta}{\delta}\right)^{\delta n}$$

- construct a cubic graph G' in such a way that for each v_i we select its two largest neighbors, say m and l, and insert s_{2i-1} on the edge (v_i, m) and s_{2i} on (v_i, l) together with an edge (s_{2i-1}, s_{2i}), see Figure 6.8.

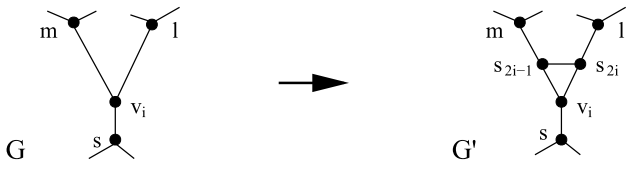


Figure 6.8: Creating a new triangle.

For a given set of δn triangles in G', there is at most one construction as above yielding G' with these as the new triangles (see Figure 6.8 and note that we can identify v_i in the triangle as the vertex adjacent to s). But G' has at most $2\delta n$ triangles. Hence the same graph G' is constructed at most $\binom{2\delta n}{\delta n} \leq 2^{2\delta n}$

times. But of course $g^{(0)}_{n+2\delta n}$ is at least the number of graphs constructed in this way. Thus

$$\begin{aligned} & g^{(0)}_{n+2\delta n} \\ \geq\ & \frac{(n+2\delta n)!}{n!} \cdot e^{-\delta n} \frac{1}{2} \alpha n^{-\beta} \rho^{-n} n! \cdot \left(\frac{1-4\delta}{\delta} \right)^{\delta n} \cdot 2^{-2\delta n} \\ >\ & \frac{1}{2} \alpha (n+2\delta n)! (n+2\delta n)^{-\beta} \rho^{-n-2\delta n} \rho^{2\delta n} e^{-\delta n} \left(\frac{1-4\delta}{\delta} \right)^{\delta n} 4^{-\delta n} \\ \geq\ & \frac{1}{4} g^{(0)}_{n+2\delta n} \left(\frac{\rho^2(1-4\delta)}{4e\delta} \right)^{\delta n} \\ >\ & g^{(0)}_{n+2\delta n}, \end{aligned}$$

a contradiction. □

6.5.4 Chromatic number

Finally we can give a full story about the chromatic number $\chi(G^{(k)}_n)$.

Theorem 6.5.7. *Let $\nu = \rho^4/4! \doteq 0.000432$. Then as $n \to \infty$*

$$\begin{aligned} \Pr(\chi(G^{(0)}_n) = 4) &\to 1 - e^{-\nu} \\ \Pr(\chi(G^{(0)}_n) = 3) &\to e^{-\nu} \doteq 0.999568. \end{aligned}$$

For $k = 1, 2, 3$ we have $\Pr(\chi(G^{(k)}_n) = 3) \to 1$ as $n \to \infty$.

Proof. By Brooks' theorem (see, e.g., [54]), for a cubic graph G with at least one triangle, $\chi(G) = 3$ unless there is a component K_4, in which case $\chi(G) = 4$. Thus the theorem follows from Lemmas 6.5.2 and 6.5.6.

□

6.6 Recursive counting and uniform sampling

Let $g^{(1)}_n$ and g^r_n be the number of labeled connected cubic planar graphs and rooted connected cubic planar simple graphs on n vertices, respectively. Since there are $3n$ ways to select an arbitrary edge and orient this edge, we get

$$3n g^{(1)}_n = g^r_n. \tag{6.6.1}$$

Moreover, a rooted connected *simple* graph G^r is neither a b-graph, nor an s- or p-graph where st are connected in the underlying graph G. Let s'_n and p'_n be the number of s- and p-graphs on n vertices without multiple root-edge. Then we get

$$g^r_n = d_n + s'_n + p'_n + h_n\,,$$

which together with (6.6.1) implies

$$3n g^{(1)}_n = d_n + s'_n + p'_n + h_n\,. \tag{6.6.2}$$

b-graphs. A b-graph's structure is restricted by 3-regularity, and the for which the undetermined part can be a d-, s-, p-, or h-graph (See Fig. 6.2).

$$b_n = \binom{n}{2}(d_{n-2} + s_{n-2} + p_{n-2} + h_{n-2}).$$

d-graphs. The d-graph is easily decomposed into two b-graphs as shown in Figure 6.3. Since the root and its adjacent vertex of such b-graphs should be selected as s, t, considering all partitions of vertices, we obtain

$$d_n = \frac{1}{(n+2)(n+1)} \sum_i \binom{n+2}{i+2} b_{i+2} b_{n-i}.$$

s-graphs. From definition s-graphs should have a *cut*-edge separating s and t in the corresponding unrooted graph. We choose the one that is nearest to vertex s for a unique decomposition. Note that the case $u = s$ or/and $v = t$ is allowed. Figure 6.4 illustrates all these cases.

If we delete the cut-edge (u, v), the unrooted s-graph falls into two separated parts, one of which containing s can be p-, h- or b-graph (due to the selection of (u, v), and the other one can be any s-, p-, h- or b-graph. Thus we have

$$\begin{aligned} s_n &= \sum_i \binom{n}{i} (p_i + h_i + b_i)\ (s_{n-i} + p_{n-i} + h_{n-i} + b_{n-i}) \\ &= \sum_i \binom{n}{i} (c_i - s_i)\ c_{n-i}\,. \end{aligned}$$

The number, s'_n, of simple s-graphs satisfies

$$s'_n = s_n - \sum_i \binom{n}{i} b_i\ b_{n-i}.$$

p-graphs. For p-graphs, we distinguish whether s and t are connected or not. Both situations are depicted in Figure 6.5.

$$p_n = n(n-1)c_{n-2} + \frac{n(n-1)}{2} \sum_i \binom{n-2}{i}\ c_i\ c_{n-i-2}$$

and

$$p'_n = \frac{n(n-1)}{2} \sum_i \binom{n-2}{i}\ c_i\ c_{n-i-2}$$

h-graphs. First we recall that an h-graph is built up from a core by *replacing* (see Section 6.1 for the definition) some edges of the core by s-, p-, h-, or b-graphs, i.e., c-graphs. We decompose an h-graph as an inverse operation of replacement.

To decompose and compute the number of h-graphs, we order the edges of cores in lexicographical order attained by the labels of the vertices of the cores. We do not count the root-edge, since the root-edge is replaced by nothing but an edge. Let $h_{n,r}$ be the number of h-graphs G^r whose first r non-root edges of the cores are also edges in G^r. Thus $h_n = h_{n,0}$.

To find a recursion formula for $h_{n,r}$ in terms of $h_{n,r+1}$, we think of an h-graph, say G^r, whose first r non-root edges of its core are also edges in G^r. Then either its $(r+1)$-th non-root edge of the core is an edge, which case is counted by $h_{n,r+1}$, or it can be replaced by a c-graph, say H^r, with $i \geq 4$ vertices, resulting a new h-graph G' (in other words, G' is decomposed into G^r and H^r, see Figure 6.6). We sum over all possible configurations and obtain

$$h_{n,r} \quad = \quad h_{n,r+1} + \frac{1}{n(n-1)} \sum_i \binom{n}{i} \, c_i \, h_{n-i,r+1} \, . \tag{6.6.3}$$

The recursion ends if all the edges of the core are replaced by edges. This means $h_{n,3n/2-1}$ is equal to the number of cores on n vertices with $3n/2-1$ edges (excluding the root-edge), which is again equal to the number of edge-rooted 3-connected cubic planar graphs on n vertices including the root-edge.

Triangulations. The dual of a 3-connected cubic planar map is a triangulation. If the 3-connected cubic planar graph is face-rooted, the triangulation is as well face-rooted, and vice versa: the root-face (incident to the root-edge) becomes the s-pole and the other face incident to the root-edge becomes the t-pole in the dual. In our drawings, the root-face incident to the root will always be the outer face.

To derive a recursion, we generalize the notion of a triangulation, as Tutte did [137]: We consider face-rooted 3-connected planar graphs where all the faces except the outer face are triangles, i.e., we do not require that the outer face is a triangle, but still assume that the graph does not contain a 2-cut. Then we distinguish between external and internal vertices and edges, where the external vertices and edges are defined to be the vertices and edges on the outer face. We call such objects *near-triangulations*. By 3-connectivity, in a near-triangulation there is no internal edge connecting two external vertices.

If a 3-connected face-rooted cubic planar graph has n vertices in total and k vertices on the outer face, the face-rooted triangulation has $\frac{n}{2}+2$ vertices in total (due to Euler's formula) and its s-pole has degree k. To count the number of such triangulations, we use the function $t_{k,l}(n)$ which denotes the number of rooted near-triangulations with n vertices, where the s-pole has degree k and there are l vertices on the outer face. When the unique internal vertex adjacent to the two poles, say u, has no internal edge connecting it to an external vertex except the poles, we remove the pole edge and move the t-pole to the unique internal vertex adjacent to the two former poles.

Otherwise we remove the edge between the poles and decompose such triangulations along the edge, say uv, connecting to the first such external vertex, say v, according to a traversal of the outer face starting from the s-pole ending at the t-pole. Then one of the two split triangulations has the new t-pole at the vertex u, and the other one has the new s-pole at v (see Figure 6.9, left) except that it has the new s-pole at u when the number of edges on the outer face is 3 (see Figure 6.9, right). All these cases can be computed inductively using the value of $t_{k,l}(n)$ for lexicographically smaller arguments.

Initially, $t_{2,3}(3) = 1$ and $t_{k,l}(n) = 0$ if $k = 2$ and $l > 3$ or $n > 3$, or if

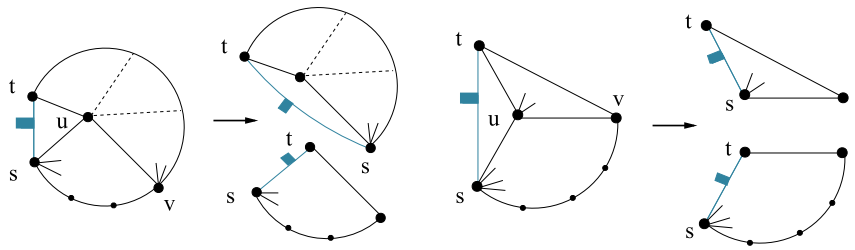


Figure 6.9: Decomposition of a rooted near-triangulation.

$l+k-2 > n$.

$$t_{k,l}(n) = t_{k-1,l+1}(n) + \sum_{k'+1,i\geq 3} t_{k-1,l}(i)t_{k',3}(n-i+2) + \sum_{k',l',i\geq 3} t_{k-1,l-l'+2}(i)t_{k',l'}(n-i+2)\,. \qquad (6.6.4)$$

The number $t(n)$ of face-rooted triangulations on n vertices is then $t(n) = \sum_{k\geq 2} t_{k,3}(n)$.

Table 6.2 shows the exact numbers $g_n^{(0)}$, $g_n^{(1)}$, $g_n^{(2)}$, and $g_n^{(3)}$ of labeled cubic planar graphs, labeled connected cubic planar graphs, labeled 2-connected cubic planar graphs, and labeled 3-connected cubic planar graphs on n vertices, up to $n = 24$. The recursive counting formulas are implemented by Löffler [95].

Uniform sampling. The decomposition and recursive counting formulas presented above yield an algorithm that samples a random cubic planar graph uniformly at random.

Theorem 6.6.1. *A labeled cubic planar graph can be sampled uniformly at random in deterministic time $O(n^6 \log^3 n)$ and in $O(n^2)$ with pre-computation, which requires $O(n^3 \log n)$ space.*

The used memory space is dominated by the tables for $t_{n,l}$ (see (6.6.4)). The number of entries in all tables is $O(n^2)$. Since each entry is bounded above by the number of all cubic planar graphs, which is $c\, n^\beta\, 3.127^n\, n!$ by Theorem 6.4.1, the binary encoding of each entry is $O(n \log n)$. Hence the total space requirement is $O(n^3 \log n)$.

The calculation of each entry involves a summation over $O(n^2)$ terms and a multiplication. All together a pre-computation of all values needs $O(n^4 \log^3 n)$ time, which dominates the running time. Using a fast multiplication algorithm, theoretically the running time can be improved to $O(n^4 \log(\log \log n))$.

The actual generation of a cubic planar graph can be done in quadratic time. The decomposition tree is of linear size and computing the probabilistic decisions at each branching step takes at most linear time, if we assume that we have access to the values in the table and their sums.

n	$g_n^{(0)}$	$g_n^{(1)}$
4	1	1
6	60	60
8	13475	13440
10	5826240	5813640
12	4124741775	4116420000
14	4379810575140	4371563196000
16	6541927990422825	6530471307360000
18	13108477865022540000	13087079865123264000
20	3398121438361359752 5425	3392927611519244 1984000
22	1107566110076203556713 93500	1105972617099522375 40320000
24	4435699914695782930344 87447675	4429662983310669492428 9280000

n	$g_n^{(2)}$	$g_n^{(3)}$
4	1	1
6	60	60
8	13440	10920
10	5700240	4112640
12	3996669600	2654467200
14	4217639025600	2625727104000
16	6272314592544000	3697449275520000
18	12526155233399808000	7034785952882688000
20	32381500604547878784000	17394357294393311232000
22	10528549715931735616128 0000	54240553998925840485920000
24	42072000093707320302838 2720000	20826458363093443003367 4240000

Table 6.2: The exact number $g_n^{(c)}$ of c-vertex-connected labeled cubic planar graphs on n vertices, up to $n = 24$, for $c = 0, 1, 2, 3$.

Chapter 7

Labeled Planar Graphs

In this chapter we derive the recursive counting formulas for labeled planar graph and design an algorithm to generate a labeled planar graph uniformly at random in deterministic time $\tilde{O}(n^7)$ with space $\tilde{O}(n^4)$.

A planar graph has in general many embeddings that are non-isomorphic as maps (i.e., graphs embedded in the plane), but some graphs have a unique embedding. A classical theorem of Whitney (see e.g., [54]) asserts that 3-connected planar graphs are *rigid* in the sense that all embeddings in the sphere are combinatorially equivalent. As *rooting* destroys any further symmetries, rooted 3-connected planar maps are closely related to *labeled* 3-connected planar graphs. Moreover, the 'degrees of freedom' of the embedding of a planar graph are governed by its connectivity structure. We exploit this fact by composing a planar graph out of 1-, 2-, and 3-connected components.

Trakhtenbrot [136] showed that every 2-connected graph with one distinguished directed edge is uniquely composed of special graphs (called *networks*) of three kinds. Such networks can be combined in series, in parallel, or using a 3-connected graph as a core (see Theorem 7.1.1 below). Using this composition we can then employ known results about counting and random sampling of 3-connected planar maps.

The concept of rooting plays an important role for the enumeration of planar maps. A *face-rooted* map is one with a distinguished edge which lies on the outer face and to which a direction is assigned. The rooting forces isomorphisms to map the outer face to the outer face, to keep the root edge incident to the outer face, and to preserve its direction. The enumeration of 3-connected face-rooted unlabeled maps with given numbers of vertices and faces, also called *c-nets*, was achieved by Mullin and Schellenberg [106]. We invoke their closed formulas in order to count 3-connected labeled planar graphs with given numbers of vertices and edges. For the generation of 3-connected labeled planar graphs with given numbers of vertices and edges we employ a recent deterministic polynomial time algorithm [28]. Alternatively, we can use a sampling procedure that runs in expected linear time that was recently presented in [67]; in this case we obtain an *expected* polynomial time sampler.

When we apply the various sampling subroutines along the stages of the connectivity decomposition, we must branch with the correct probabilities. To compute those probabilities we use recurrence formulas that can be evaluated in polynomial time using dynamic programming. Then the decomposition can

be translated immediately into a sampling procedure.

In the next section we give the graph theoretic background for the decomposition of planar graphs. In the following three sections we derive the counting formulas for planar graphs. In the last section we analyse the running time and memory requirements of the corresponding sampling procedure and discuss results from an implementation of the counting part.

7.1 Decomposition

Let us recall and fix some terminology [140, 141, 142]. A *graph* will be assumed unoriented and *simple*, i.e., having no loops or multiple (also called *parallel*) edges; if multiple edges are allowed, the term *multigraph* will be used. We consider labeled graphs whose vertex sets are initial segments of $\mathbb{N} = \{1, 2, \cdots\}$.

A *network* N is a multigraph with two distinguished vertices 1 and 2, called its *poles*, such that the multigraph N^* obtained from N by adding an edge between its poles is 2-connected. The new edge is not considered a part of the network N. We can replace an edge uv of a network M with another network X_{uv} by identifying u and v with the poles 1 and 2 of X_{uv}, and iterate the process for all edges of M. Then the resulting graph G is said to have a *decomposition* with *core* M and *components* X_e, $e \in E(M)$.

Every network can be decomposed into (or composed out of) networks of three special types. A *chain* is a network consisting of two or more edges connected in *series* with the poles as its terminal vertices. A *bond* is a network consisting of two or more edges connected in *parallel*. A *pseudo-brick* is a network N with no edge between its poles such that N^* is 3-connected. (3-connected subgraphs are sometimes called bricks.) A network N is called an *h-network* (respectively, a *p-network*, or an *s-network*) if it has a decomposition whose core is a pseudo-brick (respectively, a bond, or a chain). See Figure 7.1. Trakhtenbrot [136] (here cited from [141]) formulated a canonical decomposition theorem for networks.

Theorem 7.1.1 (Trakhtenbrot)**.** *Any network with at least 2 edges belongs to exactly one of the 3 classes: h-networks, p-networks, s-networks. An h-network has a unique decomposition and a p-network (respectively, an s-network) can be uniquely decomposed into components which are not themselves p-networks (s-networks), where uniqueness is up to orientation of the edges of the core, and also up to their order if the core is a bond.*

A network N is *simple* if N^* is a simple graph. Let $N(n, m)$ be the number of simple planar networks on n vertices and m edges. In view of Theorem 7.1.1 we introduce the functions $H(n, m)$, $P(n, m)$, and $S(n, m)$ that count the number of simple planar h-, p-, and s-networks on n vertices and m edges. Note that the components of simple networks are simple networks (or just edges). For example, K_3 (the complete graph on three vertices) is a (non-simple) p-network composed of an edge and a path of length two, which in turn is a simple s-network composed of two edges. The graph $K_4 - \{1, 2\}$ is a simple h-network, and all its components are simple edges.

Let $G^{(c)}(n, m)$ denote the number of c-connected planar graphs with n vertices and m edges. For $c = 0, 1, 2$ let us define *compose operations* for the three stages of the connectivity decomposition. Informally, for $c = 0$ the composition

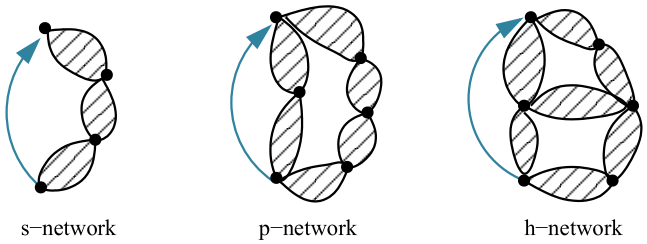


Figure 7.1: The three types of networks.

equals the disjoint union. For $c = 1$ we join the parts at a single vertex. For $c = 2$ we replace one edge of the first part by the second part. A formal definition is as follows: Assume that M and X are graphs on the vertex sets $[1 .. k]$ and $[1 .. i]$ and we want to compose them by identifying the vertices j of X with the vertices v_j of M, for $j = 1, \ldots, c$, such that the resulting graph will have $n := k + i - c$ vertices. (No vertices are identified for $c = 0$.) Moreover, let S be a set of $i - c$ vertices from $[c+1 .. n]$ which are designated for the remaining part of X. Let M' be the graph obtained by mapping the vertices of M to the set $[1 .. n] \setminus S$, retaining their relative order. Let X' be the graph obtained by mapping the vertices $[c+1 .. i]$ of X to the set S, retaining their relative order, and mapping j to the image of v_j in M' for $j = 1, \ldots, c$. Then the result of the compose operation for the arguments M, $(v_1, \ldots, v_c)$, X, and S is the graph with vertex set $[1 .. n]$ and edge set $E(M') \cup E(X')$. If $c = 2$ and M contains an edge $\{v_1, v_2\}$ it is deleted.

7.2 Planar graphs

We show how to count and generate labeled planar graphs with a given number of vertices and edges in three steps. A first simple recursive formula reduces the problem to the case of connected graphs. In the next section, we will use the block structure to reduce the problem to the 2-connected case. This may serve as an introduction to the method before we go into the more involved arguments of Section 7.4.

Let $F_k(n, m)$ denote the number of planar graphs with n vertices and m edges having k connected components. Obviously we have $F_1(n, m) = G^{(1)}(n, m)$ and $G^{(0)}(n, m) = \sum_{k=1}^{n} F_k(n, m)$. Moreover,

$$F_k(n, m) \;=\; 0 \quad \text{for} \;\; m + k < n\,.$$

We count $F_k(n, m)$ by induction on k. Every graph with $k \geq 2$ connected components can be decomposed into the connected component containing the vertex 1 and the remaining part, using the inverse of the compose operation for $c = 0$ as defined in Section 7.1. If the split-off part has i vertices, then there are $\binom{n-1}{i-1}$ ways to choose its vertex set, as the vertex 1 is always contained in it. The remaining part has $k - 1$ connected components. We obtain the recursive

formula

$$F_k(n,m) \;=\; \sum_{i=1}^{n-1}\sum_{j=0}^{m}\binom{n-1}{i-1}G^{(1)}(i,j)F_{k-1}(n-i,m-j) \quad \text{for} \;\; k\geq 2\,.$$

Thus it suffices to count connected graphs. The counting recurrence also has an analogue for generation: Assume that we want to generate a planar graph G with n vertices and m edges uniformly at random. First, we choose $k \in [1\,..\,n]$ with probability proportional to $F_k(n,m)$. Then we choose the number of vertices i of the component containing the vertex 1 and its number of edges j with a joint probability proportional to $\binom{n-1}{i-1}G^{(1)}(i,j)F_{k-1}(n-i,m-j)$. We also pick an $(i-1)$-element subset $S' \subseteq [2\,..\,n]$ uniformly at random and set $S := S' \cup \{1\}$. Then we compose G (as explained in Section 7.1) out of a random connected planar graph with parameters i and j, which is being mapped to the vertex set S, and a random planar graph with parameters $n-i$ and $m-j$ having $k-1$ connected components, which is generated in the same manner.

7.3 Connected planar graphs

In this section we reduce the counting and generation of connected labeled planar graphs to the 2-connected case. Let $M_d(n,m)$ denote the number of connected labeled planar graphs in which the vertex 1 is contained in d blocks. Here we will call them *m_d-planars*. An m_1-planar is a connected planar graph in which the vertex 1 is not a cutvertex. Clearly, $G^{(1)}(n,m) = \sum_{d=1}^{n-1} M_d(n,m)$ and

$$M_d(n,m) \;=\; 0 \quad \text{for} \;\; n<d \;\; \text{or} \;\; m<d.$$

In order to count m_d-planars by induction on d (for $d \geq 2$), we split off the largest connected subgraph containing the vertex 2 in which the vertex 1 is not a cutvertex. This is done by performing the inverse of the compose operation for $c=1$ as defined in Section 7.1. If the split off m_1-planar has i vertices, then there are $\binom{n-2}{i-2}$ possible choices for its vertex set, as the vertices 1 and 2 are always contained in it. The remaining part is an m_{d-1}-planar. Thus

$$M_d(n,m) \;=\; \sum_{i=2}^{n-d+1}\sum_{j=1}^{m-1}\binom{n-2}{i-2}M_1(i,j)M_{d-1}(n-i+1,m-j) \quad \text{for} \;\; d\geq 2\,,$$

and this immediately translates into a generation procedure.

Next we consider m_1-planars. The *root block* is the unique block containing the vertex 1. A recurrence for m_1-planars arises from splitting off the subgraphs attached to the root block at its cutvertices one at a time. Thus we consider m_1-planars such that the b least labeled vertices in the root block are not cutvertices. Let us call them *l_b-planars* and denote the number of l_b-planars with n vertices and m edges by $L_b(n,m)$. The initial cases ($b = n$) of the recurrence are connected graphs without cutvertices. We have

$$L_n(n,m) \;=\; \begin{cases} G^{(2)}(n,m) & \text{for} \;\; n\geq 3 \\ 1 & \text{for} \;\; n\in\{1,2\} \;\; \text{and} \;\; m=n-1\,. \end{cases}$$

We calculate $L_b(n,m)$ for $b = n-1,\ldots,1$, and eventually $M_1(n,m) = L_1(n,m)$, recursively as follows: To count L_b using L_{b+1}, we split off the subgraph attached

to the b-th least labeled vertex in the root block, if it is a cutvertex. This can be any connected planar graph. The remaining part is an l_{b+1}-planar. If the split off subgraph has i vertices, then there are $\binom{n-1}{i-1}$ ways to choose them, as the vertex 1 of the subgraph will be replaced with the cutvertex. We obtain the recursive formula

$$L_b(n,m) \;=\; \sum_{i=1}^{n-1}\sum_{j=0}^{m-1}\binom{n-1}{i-1}G^{(1)}(i,j)L_{b+1}(n-i+1,m-j) \quad \text{for } \; m \geq b \geq 1\,.$$

The values $G^{(1)}(i,j)$ are known since $i < n$, $j < m$. Again, the generation procedure is straightforward.

7.4 Two-connected planar graphs

In this section we show how to count and generate 2-connected planar graphs. If we take an arbitrary simple planar network with n vertices and $m-1$ edges, add an edge between the poles, then choose a pair $1 \leq x < y \leq n$, and exchange the vertex labels $(1,2)$ with (x,y), then we obtain every 2-connected labeled planar graph with n vertices and m edges in m ways. Thus

$$G^{(2)}(n,m) \;=\; \begin{cases} \dfrac{\binom{n}{2}}{m}N(n,m-1) & \text{for } \; n \geq 3\,, m \geq 3 \\ 0 & \text{otherwise.} \end{cases}$$

Now we derive recurrence formulas for the number N of simple planar networks. Trakhtenbrot's decomposition theorem implies

$$N(n,m) \;=\; \begin{cases} P(n,m) + S(n,m) + H(n,m) & \text{for } \; n \geq 3\,, m \geq 2 \\ 0 & \text{otherwise}\,. \end{cases}$$

p-Networks. Let us call a p-network with a core consisting of k parallel edges a *p_k-network*, and let $P_k(n,m)$ be the number of p_k-networks having n vertices and m edges. Clearly, $P(n,m) = \sum_{k=2}^{m} P_k(n,m)$. In order to count p_k-networks by induction on k, we split off the component containing the vertex labeled 3 by performing the inverse of the compose operation for $c = 2$ as defined in Section 7.1. Technically, it is convenient to consider the split off component as a *p_1-network*. But note that according to the canonical decomposition, a p_1-network is either an h- or an s-network. Assume that it has i vertices and j edges. Then

$$P_1(i,j) \;=\; \begin{cases} H(i,j) + S(i,j) & \text{for } i \geq 3\,, j \geq 2 \\ 0 & \text{otherwise}\,. \end{cases}$$

The remaining part is a p_{k-1}-network (even if $k = 2$). For $k \geq 2$ we have

$$P_k(n,m) \;=\; 0 \quad \text{if } \; n \leq 2 \;\text{ or }\; m < k\,.$$

There are $\binom{n-3}{i-3}$ ways how the vertex labels $[1 \,..\, n]$ can be distributed among both sides, as the labels 1, 2, and 3 are fixed. We obtain the recurrence formula

$$P_k(n,m) \;=\; \sum_{i=3}^{n-1}\sum_{j=2}^{m-1}\binom{n-3}{i-3}P_1(i,j)P_{k-1}(n-i+2,m-j) \quad \text{for } \; k \geq 2\,.$$

s-Networks. Let us call an s-network whose core is a path of k edges an *s_k-network*, and denote the number of s_k-networks which have n vertices and m edges by $S_k(n,m)$. Then $S(n,m) = \sum_{k=2}^{m} S_k(n,m)$. We use induction on k again, but for s_k-networks we split-off the component containing the vertex labeled 1. Again it can be considered as an s_1-network, and it is either an h- or a p-network, according to the canonical decomposition. Thus

$$S_1(i,j) \;=\; \begin{cases} H(i,j) + P(i,j) & \text{for } i \geq 3,\ j \geq 2 \\ 1 & \text{for } i = 2,\ j = 1 \\ 0 & \text{otherwise}\,. \end{cases}$$

The remaining part is an s_{k-1}-network (even if $k = 2$). For $k \geq 2$ we have

$$S_k(n,m) \;=\; 0 \quad \text{if} \;\; n < k+1 \;\; \text{or} \;\; m < k\,.$$

Concerning the number of ways how the labels can be distributed among both parts, note that the labels 1 and 2 are fixed, hence the new 1-root for the remaining part can be one out of $n-2$ vertices, and then the number of choices for the internal vertices of the split off s_1-network is $\binom{n-3}{i-2}$. We obtain the recurrence formula

$$S_k(n,m) \;=\; (n-2)\sum_{i=2}^{n-1}\sum_{j=1}^{m-1}\binom{n-3}{i-2} S_1(i,j) S_{k-1}(n-i+1, m-j) \quad \text{for} \;\; k \geq 2\,.$$

h-Networks. The core of an h-network is a pseudo-brick. We can order the edges of the core lexicographically using the vertex numbers. A recurrence formula similar to the p- and s-network case arises from replacing the edges of the core with components one at a time and in lexicographic order. To give names to the intermediate stages, let us call an h-network such that the components corresponding to the first k edges of the core are simple edges an *h_k-network*, and denote the number of h_k-networks with n vertices and m edges by $H_k(n,m)$. For $k \geq m$, all components must be simple edges. $H_m(n,m)$ is the number of pseudo-bricks with n vertices and m edges, the initial case of our recursion. We have

$$H_m(n,m) \;=\; \frac{(n-2)!}{2} Q(n, m+1)\,,$$

where $Q(n,m)$ denotes the number of c-nets, i.e., rooted 3-connected simple maps, with n vertices and m edges (see the next section): If we take an arbitrary c-net, assign the labels 1 and 2 to the root vertex and the other vertex of the root edge, delete the root edge, and number the remaining vertices arbitrarily, we obtain each pseudo-brick in two ways (namely, one for each face routing).

Next we derive a recurrence formula to calculate $H_k(n,m)$ for $k = m - 1, \ldots, 0$, and eventually $H(n,m) = H_0(n,m)$. To count H_k using H_{k+1}, we split off the k-th component of an h_k-network, i.e., the component replacing the k-th edge of the core. This can be a simple network of any of the three kinds or such a simple network together with an edge between its poles. Assume that it has i vertices and j edges. Then the number of choices for the component

network is

$$H'(i,j) \;=\; \begin{cases} N(i,j)+N(i,j-1) & \text{for } i\geq 3,\, j\geq 2 \\ 1 & \text{for } i=2,\, j=1 \\ 0 & \text{otherwise}\,. \end{cases}$$

The remaining part is an h_{k+1}-network. There are $\binom{n-2}{i-2}$ ways to choose the vertices of the component, as the vertices 1 and 2 are merged with the endpoints of the k-th edge of the core, respecting their relative order. We obtain the recurrence formula

$$H_k(n,m) \;=\; \sum_{i=2}^{n-2}\sum_{j=1}^{m-k+1}\binom{n-2}{i-2}H'(i,j)H_{k+1}(n-i+2,m-j+1),$$

for $m > k \geq 0$.

7.5 C-nets

In the preceding sections, we have shown how to count and sample random planar graphs assuming that we can do so for *c-nets*, i.e., 3-connected simple rooted planar maps. For this we use a formula for their number $Q(n,m)$ derived by Mullin and Schellenberg in [106]. Using Euler's formula, it asserts that

$$Q(n,m) = 0 \quad \text{for} \quad n<4 \;\; \text{or} \;\; m<n+2$$

and otherwise

$$\begin{aligned} Q(n,m) \;&=\; -\sum_{i=2}^{n}\sum_{j=n}^{m}(-1)^{i+j-n}\binom{i+j-n}{i}\binom{i}{2} \\ &\quad\times\left[\binom{2m-2n+2}{n-i}\binom{2n-2}{m-j}-4\binom{2m-2n+1}{n-i-1}\binom{2n-3}{m-j-1}\right]. \end{aligned}$$

This concludes the counting task.

The first sampling algorithm for c-nets with given numbers of vertices running in *expected* polynomial time algorithm is due to Schaeffer et al. [9, 127, 128]. For our sampling algorithm we also need to control the number of edges. A sampling procedure with this additional requirement has been described in [67]. It runs in expected time $O(n^2)$ for a fixed edge density ratio $\alpha \in \,]\frac{3}{2},3[$, where $\frac{m}{n}\to\alpha$, and in expected time $O(n^3)$ for triangulations (where $\frac{m}{n}\to 3$), which is also the worst case [67].

For a *deterministic* polynomial running time, we use an extended version of the algorithm presented in [28] with an additional parameter for the number of edges, as explained in the conclusion of [28]. The resulting algorithm runs in deterministic $\tilde{O}(n^7)$ time and $O(n^4)$ space, or, if a pre-computation is allowed, $\tilde{O}(n^3)$ time and $O(n^7)$ space.

7.6 Planar multigraphs

We have seen how to count and generate random planar graphs on a given number of vertices and edges using a recursive decomposition along the connectivity

structure. A by-product of our result is that we can also generate *connected* and *2-connected* labeled planar graphs uniformly at random. Moreover, it is easy to see that we can count and generate random planar *multigraphs* by only changing the initial values for planar networks as follows:

$$\begin{aligned} N(n,m) &= P(n,m) && \text{for } n = 2\,, m \geq 2 \\ P_k(n,m) &= 1 && \text{for } n = 2\,, m = k\,, k \geq 1\,. \end{aligned}$$

7.7 Uniform sampling and exact numbers

In this section we establish a polynomial upper bound on the running time and the memory requirement of our sampling algorithm. We also report on computational results from an implementation of the counting formulas.

Uniform sampling. Since our algorithm for sampling random labeled planar graphs is an application of the well-known 'recursive method' for sampling [47, 64, 108], we outline the essentials only.

The algorithm pre-calculates a number of dynamic programming arrays containing the values of F, M, L, N, P, S, H, Q, and G, before the actual random generation starts. Altogether these tables have $O(n^3)$ entries, and all entries are bounded by the number of planar graphs $c\, n^{-7/2}\, 27.2^n\, n!$ (see Section 2.1). Therefore the encoding length is $O(n \log n)$ [48, 112] and the total space requirement is in $O(n^4 \log n)$ bits. The computation of each entry involves a summation over $O(n^2)$ terms. Using a fast multiplication algorithm (see e.g., [44]), the tables can be filled in $\tilde{O}(n^6)$ time.

The values in the dynamic programming tables are used during the probabilistic decisions in a recursive construction of the labeled planar graph, which is essentially the inversion of the presented decomposition. For each entry, we scan over all the entries from which it was computed (there are at most nm of them) and store the partial sums in a balanced binary tree, where each internal node contains the maximum of its left-hand siblings. The total size of the resulting data structure is $\tilde{O}(n^6)$ and it can be initialized in $\tilde{O}(n^6)$ time.

We assume that we can obtain random bits at unit cost. When given a random number between 1 and the sum over all leaves, we can find the corresponding table entry in one pass through the tree of partial sums, while reading each bit of the random number only a constant number of times, and hence in $O(n \log n)$ time. Then the procedure calls itself recursively for both factors of the product. Note that the sum of the bit lengths of both factors is linear in the bit length of the entry. It follows that the total running time for traversing the decomposition tree and creating the output is in $\tilde{O}(n^2)$, and hence dominated by the generation of c-nets.

It is not necessary to create the binary trees for each entry of the tables. Instead, one can simply recompute some of the values from the preprocessing step and stop if the partial sum exceeds the random number. In this way, the recursive decomposition uses $\tilde{O}(n^6)$ time and $O(n^4 \log n)$ space. Now the following theorem follows by combining the results of this and the preceding section.

Theorem 7.7.1. *Labeled planar graphs on n vertices and m edges can be sampled uniformly at random in deterministic time $\tilde{O}(n^7)$ and space $O(n^4 \log n)$.*

If we apply a preprocessing step, this can also be done in deterministic time $\tilde{O}(n^3)$ *and space* $O(n^7)$.

To increase the efficiency of the algorithm one might want to apply a technique where the generated combinatorial objects only have approximately the correct size; this can then be turned into an exact generation procedure by rejection sampling. A general framework to tune and analyse such procedures has been developed in [9, 55] and applied to structures derived by e.g., disjoint unions, products, sequences and sets. To deal with planar graphs it needs to be extended to the compose operation used in this chapter. Fusy [66] developed a Boltzmann sampler that sample a labeled planar graph of an exact size and runs in quadratic time, based on the general frame work of Boltzmann sampler [55], a bijection between trees and c-nets [67], and analytic expressions of generating functions for planar graphs with various connectivity [71].

Exact numbers. Table 7.1 shows the exact numbers $g_n^{(0)}$, $g_n^{(1)}$, $g_n^{(2)}$, and $g_n^{(3)}$ of labeled planar graphs, labeled connected planar graphs, labeled 2-connected planar graphs, and labeled 3-connected planar graphs on n vertices, up to $n = 18$, which are computed from the recursive counting formulas.

Using the computed numbers we can study several basic questions about a random labeled planar graph.

Denise, Vasconcellos, and Welsh [48] introduced a Markov chain whose stationary distribution is the uniform distribution on all labeled planar graphs. However, its mixing time is unknown and seems hard to analyse, and is perhaps not polynomial. Moreover, the corresponding sampling algorithm only *approximates* the uniform distribution.

Denise, Vasconcellos and Welsh [48] proved that $g_n^{(0)} \leq n!75.8^{n+o(n)}$. Bender, Gao, and Wormald [13] proved that $g_n^{(0)} \leq n!26.1^{n+o(n)}$, which is the number of labeled 2-connected planar graph. Furthermore, Osthus, Prömel, and Taraz [112] improved the upperbound to be $g_n^{(0)} \leq n!37.3^{n+o(n)}$. McDiarmid, Steger, and Welsh [99] proved that the quantity $(g_n^{(0)}/n!)^{1/n}$ converges to a limit γ, the *labeled planar graph growth constant* [99], as $n \to \infty$, and Gerke, McDiarmid, Steger, and Weißl [70] proved that the quantity $(G^{(0)}(n, m)/n!)^{1/n}$ with $m = qn$ $(1 < q < 3)$ converges to a limit γ_q as $n \to \infty$.

To see the speed of convergence, we compute the value of $g_n^{(c)}/(n \cdot g_{n-1}^{(c)})$ for various connectivity c, e.g., $g_{50}^{(0)}/(50 \cdot g_{49}^{(0)}) \doteq 25.2$. We compute also the value of $g_n^{(c)}/g_n^{(0)}$ for several ranges of the connectivity c, in particular $g_{50}^{(1)}/g_{50}^{(0)} \doteq 0.96$.

Gerke and McDiarmid [69] proved that the limit μ of the expected edge density of general (no connectivity requirement) labeled planar graphs is at least $13/6 \doteq 1.86$, and Bonichon, Gavoille and Hanusse [39] proved that μ is smaller than 2.54. Our computation shows that this value for $n = 50$ is 2.12. Giménez and Noy [71] determined the labeled planar graph growth constant $\gamma \doteq 27.2$, the asymptotic fraction of connected graphs $p_{\text{con}} \doteq 0.96$, and the limit of expected edge density $\mu \doteq 2.21$.

n	$g_n^{(0)}$	$g_n^{(1)}$
1	1	1
2	2	1
3	8	4
4	64	38
5	1023	727
6	32071	26013
7	1823707	1597690
8	163947848	149248656
9	20402420291	18919743219
10	3209997749284	3005354096360
11	604611323732576	569226803220234
12	131861300077834966	124594074249852576
13	32577569614176693919	30861014504270954737
14	8977083127683999891824	852044383864683323 1236
15	2726955513946123452637877	259215068456593597715 2860
16	904755724004585279250537376	861079753184429687852978432
17	325403988657293080813790670641	310008316267496041749182487881
18	126073204858661604803062210068760	120210565158574034465039064701904

n	$g_n^{(2)}$	$g_n^{(3)}$
1	1	0
2	1	0
3	3	0
4	28	1
5	490	25
6	15306	1227
7	822766	84672
8	70546120	7635120
9	8646554043	850626360
10	1374022343240	112876089480
11	304206135619160	17381709797760
12	65030138045062272	3046480841900160
13	15659855107404275280	598731545755324800
14	4191800375194003211360	130389773403373545600
15	1234179902360142341550240	3116361648643483806 7200
16	396280329098426228719121280	810921300929658613094 4000
17	137779269467538258010671193472	228201401065777376416 0588800
18	515599688357675820346930550429 76	690521215428258768326957184000

Table 7.1: The exact number $g_n^{(c)}$ of c-vertex connected labeled planar graphs on n vertices, up to $n = 18$, for $c = 0, 1, 2, 3$.

Part III

Unlabeled Planar Structures

Chapter 8

Unlabeled Trees

In this chapter we recall enumeration results on unlabeled trees, based on the results by Otter [76, 113], to illustrate how to use the cycle index sums to compute the number of unlabeled structures.

First we determine that the number of unlabeled trees is asymptotically $c\ n^{-5/2}\ \xi^{-n}\ n!$ for constants $c \doteq 0.5349485$ and $\xi \approx 2.95577$. Then we derive the recursive counting formulas for rooted trees, which yield a uniform sampling algorithm for *unrooted* trees running in deterministic time $\tilde{O}(n^4)$ with space $\tilde{O}(n^3)$.

In Sections 8.1 and 8.2 we derive the cycle index sum and the ordinary generating function for unlabeled rooted trees. We also derive the identity relating the ordinary generating function for unlabeled *rooted* trees and that for unlabeled *unrooted* trees. In Section 8.3 we determine the asymptotic numbers of unlabeled (rooted and unrooted) trees. In Section 8.4 we present the recursive counting formulas, which can be turned into a uniform sampling algorithm that samples an unlabeled unrooted tree.

8.1 Unlabeled rooted trees

We will begin with enumerating unlabeled *rooted* trees and then apply the dissimilarity characteristic theorem to enumerate unlabeled *unrooted* trees.

Let $\hat{t}(n)$ be the number of unlabeled rooted trees on n vertices and let

$$\hat{T}(x) = \sum_{n \geq 0} \hat{t}(n) x^n$$

be its corresponding ordinary generating function. Pólya showed that

$$\hat{T}(x) = x \exp\left(\sum_{k \geq 1} \frac{\hat{T}(x^k)}{k} \right). \tag{8.1.1}$$

To see this, we consider the number $\hat{t}_d(n)$ of the unlabeled rooted trees where the root has degree d. As we have seen in Section 4.4, given such a tree, deleting the root and the edges adjacent to the root yields a collection of d unlabeled rooted trees. Conversely, given a collection of d unlabeled rooted trees, a new

unlabeled rooted tree is constructed by adding one new vertex and connecting it with each of the roots of d given rooted trees.

To compute $\hat{t}_d(n)$ we consider the power group I^{S_d} with object set Y^X where $X = \{1, 2, \cdots, d\}$, Y is the set of all unlabeled rooted trees, S_d is the symmetry group on X, and I is the identity group on Y. We define the weight of each unlabeled rooted tree in Y to be the number of vertices in the tree, that is, $\omega : Y \to \{0, 1, \cdots, \}$ is defined by $\omega(y) = k$ for $|y| = k$. Then $\hat{T}(x)$ enumerates the elements of Y by weight, i.e., it is the figure counting seires for Y. Thus the weight of each function f in Y^X defined by (2.3.3) is the total number of vertices in the d rooted trees to which the function f corresponds.

Since S_d consists of all permutations of X, the orbits of the power group I^{S_d} correspond precisely to rooted trees whose root has degree d. The weight of each orbit is one less than the total number of vertices in the rooted tree to which the orbit corresponds due to its root vertex.

From Pólya's enumeration theorem (Theorem 2.3.1) with $A = S_d$ and $c(x) = \hat{T}(x)$ we know that $Z(S_d, \hat{T}(x))$ is the function counting series and the coefficient of x^n in $Z(S_d, \hat{T}(x))$ is the number of rooted trees on $n+1$ vertices, whose root have degree d. Thus the coefficient of x^n in $xZ(S_d, \hat{T}(x))$ is the number of rooted trees on n vertices, whose root have degree d. Thus summing over all possible values of d we obtain $\hat{T}(x)$, that is,

$$\begin{aligned}\hat{T}(x) = x\sum_{d\geq 0} xZ(S_d, \hat{T}(x)) &= x\sum_{n\geq 0} xZ(S_d; \hat{T}(x), \hat{T}(x^2), \hat{T}(x^3), \cdots) \\ &\overset{(2.3.1)}{=} x\exp\left(\sum_{k\geq 1}\frac{\hat{T}(x^k)}{k}\right).\end{aligned}$$

The following identity enables us to determine the coefficient of $\hat{T}(x)$ recursively: If $\sum_{n\geq 0} A_n x^n = \exp(\sum_{n\geq 0} a_n x^n)$, then $a_n = A_n - \frac{1}{n}\left(\sum_{k=1}^{n-1} k a_k A_{n-k}\right)$. The first few terms computed by Riordan [120] are

$$\hat{T}(x) = x + x^2 + 2x^3 + 4x^4 + 9x^5 + 20x^6 + 48x^7 + 115x^8 + \cdots.$$

8.2 Unlabeled trees

The dissimilarity characteristic theorem (see e.g., Theorem 8.2.1) expresses the number of dissimilar vertices of a graph in terms of the numbers of dissimilar blocks and the number of dissimilar vertices of each block in the graph [76]. In the case of trees, the blocks of the graph are the edges.

For any tree T, let v^* be the number of dissimilar vertices, that is, the number of orbits of vertices determined by the automorphism group $\Gamma(T)$. Further $\Gamma(T)$ determines the similarity classes of edges. An edge is said to be *symmetry* if its end vertices are similar. Let e^* the number of dissimilar edges and s the number of symmetry edges. Note first that s equals 0 or 1. In particular, $s = 1$ if and only if the tree has two central points that are similar. Furthermore we have the following identity, known as the dissimilarity characteristic theorem for trees.

Theorem 8.2.1. *Let v^*, e^* and s be as above. Then*

$$1 = v^* - e^* + s. \tag{8.2.1}$$

Proof. We prove this by induction on the number of classes of dissimilar edges. If T has just one class of dissimilar edges, then clearly $v^* = 2, e^* = 1, s = 0$. Thus (8.2.1) holds. Otherwise, consider a dissimilar edge e of T, one of whose end vertices is a leaf, and delete all the leaves in the same class as e from T. Then the resulting graph T' has $v^* - 1$ dissimilar vertices, and $e^* - 1$ dissimilar edges and s symmetry edges. Applying the induction hyperthesis to T', we have $1 = (v^* - 1) - (e^* - 1) + s = v^* - e^* + s$. □

Let $t(n)$ be the number of unlabeled (unrooted) trees on n vertices and let $T(x) = \sum_{n\geq 0} t(n)x^n$ be its corresponding ordinary generating function. The identity (8.2.1) can be used to derive $T(x)$ from the generating function for trees rooted at a vertex $\hat{T}(x)$. Summing (8.2.1) over all unlabeled (unrooted) trees on n vertices we obtain

$$\sum_{T_n} 1 = \sum_{T_n} v^* - \sum_{T_n} (e^* - s). \tag{8.2.2}$$

Note that $\sum_{T_n} 1 = t(n)$ and $\sum_{T_n} v^* = \hat{t}(n)$. Furthermore $\sum_{T_n}(e^* - s)$ counts the number of unlabeled trees on n vertices rooted at an asymmetric edge. Such a tree is determined by any two different trees rooted at a vertex by joining two roots by a distinguished edge. Thus $\sum_{T_n}(e^* - s) = ((\hat{T}(x))^2 - \hat{T}(x^2))/2$. From (8.2.2) we have the following identity.

Theorem 8.2.2. *Let $T(x)$ and $\hat{T}(x)$ be as above. Then*

$$T(x) = \hat{T}(x) - \frac{1}{2}\left((\hat{T}(x))^2 - \hat{T}(x^2)\right). \tag{8.2.3}$$

8.3 Asymptotics

From (8.1.1) one can show that $\hat{T}(x)$ has a singular expansion of the form

$$\hat{T}(x) = 1 + a(1 - x/\xi)^{1/2} + b(1 - x/\xi) + c(1 - x/\xi)^{3/2} + \cdots,$$

for some nonzero constants ξ, a, b, c. Applying Lemma 2.4.2 and the rescaling rule (2.4.3) we obtain

$$\hat{t}_n \sim \hat{c} n^{-3/2} \xi^{-n},$$

where $\hat{c} \doteq 0.5349485$ and $\xi \approx 2.95577$.

From Theorem 8.2.2 we may get $T(x) = d + e(1 - x/\xi)^{1/2} + f(1 - x/\xi) + g(1 - x/\xi)^{3/2} + \cdots$, for some constants d, e, f, g. However by differentiating (8.2.3) we get

$$T'(x) = \hat{T}'(x)\left(1 - \hat{T}(x)\right) - \hat{T}'(x^2)x,$$

which implies that $e = 0$. Moreover one can show that $g \neq 0$ using the second derivative of $T(x)$. Thus $T(x)$ has a singular expansion of the form

$$T(x) = d + f(1 - x/\xi) + g(1 - x/\xi)^{3/2} + \cdots,$$

for some nonzero constants d, f, g. Again applying Lemma 2.4.2 and the rescaling rule (2.4.3) we obtain

$$t_n \sim c n^{-5/2} \xi^{-n},$$

where $c \doteq 0.5349485$ and $\xi \approx 2.95577$.

8.4 Recursive counting and uniform sampling

In this section we derive the recursive counting formulas for unlabeled but *rooted* trees by the decomposition strategy, and then derive the formula for the unlabeled (unrooted) trees using (8.2.3).

As before let $\hat{t}(n)$ be the number of all unlabeled rooted trees with n vertices. Let $\hat{t}_l(n)$ be the number of all trees with n vertices where all the rooted subtrees have size $\leq l$. Clearly $\hat{t}_1(n) = 1$ (which counts a so-called *star* graph). Then summing over the number k of rooted subtrees of size l we get, for $n \geq 2$ and $1 \leq l \leq n-1$,

$$\begin{aligned} \hat{t}(n) &= \hat{t}_{n-1}(n), \\ \hat{t}_l(n) &= \sum_{k=0}^{\lfloor (n-1)/l \rfloor} \binom{\hat{t}(l)+k-1}{k} \hat{t}_{l-1}(n-kl). \end{aligned}$$

From (8.2.3) we get that for n even,

$$t_n = \hat{t}(n) - \sum_{i=0}^{\lfloor n/2 \rfloor} \hat{t}(i)\hat{t}(n-i) + \frac{1}{2}\hat{t}(n/2)\,,$$

and for n odd,

$$t_n = \hat{t}(n) - \sum_{i=0}^{n/2-1} \hat{t}(i)\hat{t}(n-i) - \frac{1}{2}\hat{t}(n/2)^2\,.$$

All the counting numbers can be computed in polynomial time and their values can be stored in linear space. We have a complete set of counting formulas for unrooted trees, which yields a polynomial time algorithm that samples an unrooted tree uniformly at random.

Theorem 8.4.1. *Unlabeled trees on n vertices can be sampled uniformly at random in deterministic time $\tilde{O}(n^4)$ with space $\tilde{O}(n^3)$.*

There are known algorithms that sample a random (unrooted) tree uniformly at random. To name a few, [143] uses the centroid of trees and [27] uses a Boltzmann sampler that runs in linear time with no pre-processing.

Chapter 9

Unlabeled Outerplanar Graphs

In this chapter we determine the asymptotic number of *unlabeled* outerplanar graphs using cycle indices, and derive recursive counting formulas and a uniform sampling algorithm.

We have seen in Chapter 5 that outerplanar graphs can be decomposed according to the degree of connectivity. An outerplanar graph is a set of connected outerplanar graphs, and a connected outerplanar graph can be decomposed into two-connected outerplanar graphs. In the labeled case this decomposition yields the equations that link the exponential generating functions of two-connected, connected, and general outerplanar graphs.

For unlabeled outerplanar graphs we use *cycle index sums* to obtain exact and asymptotic results. Using the cycle index sums we obtain implicit information about the ordinary generating functions of unlabeled outerplanar graphs. The difficulty is that the generating function for unlabeled connected outerplanar graphs is defined implicitly by a multiset of connected outerplanar graphs. However, we overcome this by applying the singular implicit function theorem.

We show that the number of unlabeled outerplanar graphs on n vertices is asymptotically $c\ n^{-5/2}\ \eta^{-n}\ n!$, for a suitable positive constant c and $\eta^{-1} \approx 7.503597$. We also study typical properties of a random unlabeled outerplanar graph chosen uniformly at random among all the unlabeled outerplanar graphs on n vertices. The expected number of edges in a random unlabeled outerplanar graph is asymptotically $1.54894n$ and the probability that a random unlabeled outerplanar graph is connected tends to 0.845721 as n goes to ∞. Furthermore, the chromatic number of a random unlabeled outerplanar graph is three with probability tending to one as n goes to ∞.

Using the decomposition along the connectivity we derive recursive counting formulas for unlabeled *rooted* connected outerplanar graphs, and present a uniform generation procedure for unlabeled *unrooted* connected outerplanar graphs running in expected time $\tilde{O}(n^7)$ with space $\tilde{O}(n^3)$, using *rejection sampling*.

The rest of the chapter is organized as follows. In Section 9.1 we derive cycle indices for unlabeled two-connected, connected, and general outerplanar graphs. In Section 9.3 we estimate the asymptotic numbers for unlabeled outerplanar graphs with various connectivity. In Section 9.4 we investigate typical

properties of a random outerplanar graph on n vertices, such as the probability of connectedness, the expected number of components, the expected number of isolated vertices, the chromatic number, and the number of edges. In Section 9.5 we present recursive counting formulas and the uniform sampling algorithm.

9.1 Cycle indices

In this section we discuss the cycle indices of unlabeled two-connected, connected, and general outerplanar graphs.

Two-connected outerplanar graphs. The generating functions for unlabeled two-connected outerplanar graphs were derived in [26, 119, 139] as follows.

Theorem 9.1.1. *The cycle index sum for unlabeled two-connected outerplanar graphs (i.e., dissections) is given by*

$$\begin{aligned} Z(\mathcal{B}) = &-\frac{1}{2}\sum_{d\geq 1}\frac{\varphi(d)}{d}\log\left(\frac{3}{4}-\frac{1}{4}s_d+\frac{1}{4}\sqrt{s_d^2-6s_d+1}\right)+\frac{s_2+s_1^2-4s_1-2}{16}\\ &+\frac{s_1^2-3s_1^2s_2+2s_1s_2}{16s_2^2}+\frac{3-s_1}{16}\sqrt{s_1^2-6s_1+1}\\ &-\frac{1}{16}\left(1+\frac{s_1^2}{s_2^2}+\frac{2s_1}{s_2}\right)\sqrt{s_2^2-6s_2+1}. \end{aligned}$$

Using Formula (2.3.6) and Theorem 9.1.1 we derive the cycle index sum for vertex rooted dissections, which we will need later.

Corollary 9.1.2. *The cycle index sum for vertex rooted dissections is given by*

$$\begin{aligned} Z(\mathcal{V};s_1,s_2) = &\frac{s_1}{8}\left(1+s_1-\sqrt{s_1^2-6s_1+1}\right)\\ &+\frac{s_1}{8s_2^2}(s_1+s_2)\left(1-3s_2-\sqrt{s_2^2-6s_2+1}\right). \end{aligned} \tag{9.1.1}$$

Connected outerplanar graphs. We denote the set of unrooted connected outerplanar graphs by $\mathcal{C}$, and the set of vertex rooted connected outerplanar graphs by $\hat{\mathcal{C}}$. All rooted graphs considered in this section are rooted at a vertex. Consider the corresponding generating functions $C(x)=\sum_n c_n x^n$ and $\hat{C}(x)=\sum_n \hat{c}_n x^n$.

The cycle index sum for rooted connected outerplanar graphs is derived by decomposing the graphs into rooted two-connected outerplanar graphs, i.e., vertex rooted dissections.

Lemma 9.1.3. *The cycle index sum for vertex rooted connected outerplanar graphs is implicitly determined by the equation*

$$Z(\hat{\mathcal{C}}) = s_1\exp\left(\sum_{k\geq 1}\frac{Z(\mathcal{V};Z(\hat{\mathcal{C}};s_k,s_{2k},\cdots),Z(\hat{\mathcal{C}};s_{2k},s_{4k},\cdots))}{k\,Z(\hat{\mathcal{C}};s_k,s_{2k},\cdots)}\right). \tag{9.1.2}$$

Proof. Graphs in $\hat{\mathcal{C}}$ rooted at a vertex that is not a cut-vertex can be constructed by taking a rooted dissection and attaching a rooted connected outerplanar graph at each vertex of the dissection other than the root vertex. By (2.3.5) we obtain that

$$s_1\left(\frac{Z\left(\mathcal{V}\right)}{s_1}\right)\left[Z(\hat{\mathcal{C}})\right] \tag{9.1.3}$$

is the cycle index sum for connected outerplanar graphs rooted at a non-cut-vertex. The division (resp. multiplication) by s_1 is due to the removal (resp. addition) of the root vertex before (resp. after) application of (2.3.5).

The cycle index sum for rooted connected outerplanar graphs where the root vertex is incident to exactly n blocks, $n \geq 2$, can be obtained by another application of the composition theorem. We join n connected outerplanar graphs that are rooted at a vertex other than a cut-vertex at their root vertex. Applying (2.3.5) with the symmetric group S_n and (9.1.3) (divided by s_1) for the cycle index sum for non-cut-vertex rooted connected outerplanar graphs (excluding the root) yields

$$s_1 Z\left(S_n\right)\left[\left(\frac{Z(\mathcal{V})}{s_1}\right)\left[Z(\hat{\mathcal{C}})\right]\right].$$

Summing over $n \geq 0$, we get $\left(Z\left(S_0\right) := 1\right)$

$$Z(\hat{\mathcal{C}}) = s_1 \sum_{n\geq 0} Z\left(S_n\right)\left[\left(\frac{Z\left(\mathcal{V}\right)}{s_1}\right)\left[Z(\hat{\mathcal{C}})\right]\right].$$

With the well-known formula (2.3.1), the statement follows. □

Theorem 9.1.4. *The cycle index sum for connected outerplanar graphs is given by*

$$Z\left(\mathcal{C}\right) = Z(\hat{\mathcal{C}}) + Z(\mathcal{B}; Z(\hat{\mathcal{C}})) - Z(\mathcal{V}; Z(\hat{\mathcal{C}})). \tag{9.1.4}$$

Proof. To derive the cycle index sum for unrooted connected outerplanar graphs, one can use (2.3.7). We obtain

$$Z\left(\mathcal{C}\right) = \int_0^{s_1} \frac{1}{s_1} Z(\hat{\mathcal{C}}) ds_1 + Z\left(\mathcal{C}\right)|_{s_1=0}\,. \tag{9.1.5}$$

The term $Z\left(\mathcal{C}\right)|_{s_1=0}$ can be further replaced by $Z\left(\mathcal{B}\right)|_{s_1=0}\left[Z(\hat{\mathcal{C}})\right]$ because each fixed-point free permutation in a connected graph G has a unique block whose vertices are setwise fixed by the automorphisms of G. Using the special structure (9.1.2) of $Z(\hat{\mathcal{C}})$, a closed solution for the integral in (9.1.5) can be found [139]. We put these facts together and obtain (9.1.4). □

Outerplanar graphs. We denote the set of outerplanar graphs by $\mathcal{G}$, its ordinary generating function by $G\left(x\right)$ and the number of outerplanar graphs with n vertices by g_n. As an outerplanar graph is a collection of connected outerplanar graphs, it is now easy to obtain the cycle index sum for outerplanar graphs. An application of the composition formula (2.3.5) with the symmetric group S_l and object set $\mathcal{C}$ yields that $Z\left(S_l\right)\left[Z\left(\mathcal{C}\right)\right]$ is the cycle index sum for outerplanar graphs with l connected components. Thus, by summation over all $l \geq 0$ (we include here also the empty graph into $\mathcal{G}$ for convenience), we obtain the following theorem.

Theorem 9.1.5. *The cycle index sum for outerplanar graphs is given by*

$$Z\left(\mathcal{G}\right)=\exp\left(\sum_{k\geq 1}\frac{1}{k}Z\left(\mathcal{C};s_k,s_{2k},\cdots\right)\right).$$

9.2 Ordinary generating functions and exact numbers

As discussed in Section 2.3, the ordinary generating functions can be derived from their cycle index sums by replacing the formal variables s_i by x^i. Exact numbers can then be computed from the ordinary generating functions by Taylor series expansion around $x=0$.

Ordinary generating functions. Replacing s_1 by x, s_2 by x^2, ... we obtain the generating functions, $V(x)$ and $B(x)$ of vertex rooted dissections and dissections:

$$\begin{aligned}V\left(x\right)=&\frac{1}{8x^2}\left(1+x-3x^2-2x^3+x^4\right)\\&-\frac{1}{8x^2}\left(\left(1+x\right)\sqrt{x^4-6x^2+1}-x^3\sqrt{x^2-6x+1}\right),\end{aligned}\tag{9.2.1}$$

$$\begin{aligned}B\left(x\right)=&-\frac{1}{2}\sum_{d\geq 1}\frac{\varphi\left(d\right)}{d}\log\left(\frac{1}{4}\left(3-x^d+\sqrt{x^{2d}-6x^d+1}\right)\right)\\&+\frac{x^2}{8}-\frac{1}{4}x-\frac{5}{16}+\frac{1}{8x}+\frac{1}{16x^2}+\frac{3-x}{16}\sqrt{x^2-6x+1}\\&-\frac{1+2x+x^2}{16x^2}\sqrt{x^4-6x^2+1}.\end{aligned}\tag{9.2.2}$$

The coefficients of $B(x)$, counting unlabeled dissections, can be extracted in polynomial time,

$$B(x)=x^2+x^3+2x^4+3x^5+9x^6+20x^7+75x^8+262x^9+\ldots,$$

matching the values computed by Read, see [133, A001004].

Replacing s_i by x^i in $Z(\hat{\mathcal{C}})$, we obtain that the generating function $\hat{C}(x)$ counting vertex rooted connected outerplanar graphs satisfies

$$\hat{C}\left(x\right)=x\exp\left(\sum_{k\geq 1}\frac{Z(\mathcal{V};\hat{C}\left(x^k\right))}{k\,\hat{C}\left(x^k\right)}\right),\tag{9.2.3}$$

from which the coefficients $\hat{C}_n$ counting vertex rooted connected outerplanar graphs can be extracted in polynomial time:

$$\hat{C}\left(x\right)=x+x^2+3x^3+10x^4+40x^5+181x^6+918x^7+\ldots.$$

See [135, 139] for more entries. The numbers in [135] verify the correctness of our result and were computed by the polynomial algorithm proposed in [21].

In addition, it follows from (9.1.4) that the generating function $C(x)$ counting connected outerplanar graphs satisfies:

$$C\left(x\right) = \hat{C}\left(x\right) + Z(\mathcal{B}; \hat{C}\left(x\right)) - Z(\mathcal{V}; \hat{C}\left(x\right)), \tag{9.2.4}$$

from which the coefficients c_n counting connected outerplanar graphs can be extracted in polynomial time:

$$C(x) = x + x^2 + 2x^3 + 5x^4 + 13x^5 + 46x^6 + 172x^7 + \ldots.$$

See [133, A111563] for more entries.

The generating functions $G(x)$ and $C(x)$ of outerplanar and connected outerplanar graphs are related by

$$G\left(x\right) = \exp\left(\sum_{k\geq 1} \frac{1}{k} C\left(x^k\right)\right). \tag{9.2.5}$$

From this, we can extract in polynomial time the coefficients counting outerplanar graphs,

$$G(x) = 1 + x + 2x^2 + 4x^3 + 10x^4 + 25x^5 + 80x^6 + 277x^7 + \ldots.$$

See [133, A111564] for more entries.

Exact numbers. We compute the exact numbers using Taylor series expansion, around $x = 0$, of the ordinary generating functions presented above. We can also derive recursion formulas to compute the exact numbers for vertex rooted unlabeled outerplanar graphs in polynomial time (see Section 9.5). All computations necessary to compute b_n, c_n, and g_n can be executed in polynomial time. Table 9.1 shows the exact numbers that are computed in this way and implemented by Vigerske [139]. The exact numbers $\hat{c}_n$ for vertex rooted unlabeled outerplanar graphs were verified by Tomii [135] using the counting formulas in Section 9.5.

9.3 Asymptotics

To determine the asymptotic number of unlabeled two-connected, connected, and general outerplanar graphs, we use singularity analysis introduced in Section 2.4. To compute the growth constants and subexponential factors we expand the generating functions for outerplanar graphs around their dominant singularities. For unlabeled two-connected outerplanar graphs we present an analytic expression of the growth constant. For the connected and the general case we give numerical approximations of the growth constants.

Two-connected outerplanar graphs. We first derive the asymptotic number of unlabeled two-connected outerplanar graphs.

Theorem 9.3.1. *Let b_n be the number of unlabeled two-connected outerplanar graphs on n vertices. Then*

$$b_n \sim b\, n^{-\frac{5}{2}} \delta^{-n}$$

with growth rate $\delta^{-1} = 3 + 2\sqrt{2} \approx 5.82843$ and constant $b \doteq 0.00596026$.

n	b_n	c_n	g_n
1	0	1	1
2	1	1	2
3	1	2	4
4	2	5	10
5	3	13	25
6	9	46	80
7	20	172	277
8	75	777	1150
9	262	3783	5291
10	1117	20074	26918
11	4783	111604	145744
12	21971	646409	828856
13	102249	3846640	4872771
14	489077	23410035	29395784
15	2370142	144965988	180857382
16	11654465	910898943	1130700488
17	57916324	5794179218	7163245811
18	290693391	37248630398	45895629266
19	1471341341	241676806702	296937363511
20	7504177738	1580880366039	1937625709854
21	38532692207	10416314047854	12739784808937
22	199076194985	69080674190341	84331837321404
23	1034236705992	460841447382976	561647630439975
24	5400337050086	3090747326749823	3761221057579892
25	28329240333758	20829976038652612	25314597326376883

Table 9.1: The exact numbers b_n, c_n, g_n of unlabeled two-connected outerplanar graphs, connected outerplanar graphs and outerplanar graphs on n vertices, up to $n = 25$.

Proof. Let δ be the smallest root of $x^2 - 6x + 1$, $\delta = 3 - 2\sqrt{2}$. Equation (9.2.2) implies that $B(x)$ can be written as

$$B(x) = -\frac{1}{2}\log\left(1 - \frac{\sqrt{x^2-6x+1}}{x-3}\right) + \frac{3-x}{16}\sqrt{x^2-6x+1} + A(x),$$

where $A(x)$ is analytic at 0 with radius of convergence $> \delta$. Since the logarithmic term is analytic for $|x| < \delta$, we can expand it and collect ascending powers of $\sqrt{x^2-6x+1}$ in $B(x)$. Thus,

$$B(x) = \left(-\frac{1}{16(x-3)} + \frac{1}{6(x-3)^3}\right)\left(\sqrt{x^2-6x+1}\right)^3 + \sum_{k\geq 4}\frac{1}{2k}\left(\frac{\sqrt{x^2-6x+1}}{x-3}\right)^k + \tilde{A}(x),$$

where $\tilde{A}(x)$ is again analytic at 0 with radius of convergence $> \delta$. Finally, using $\sqrt{x^2-6x+1} = \sqrt{1-x/\delta}\sqrt{1-\delta x}$ for $x \leq \delta$ and applying Lemma 2.4.2 we

obtain

$$b_n = \left(-\frac{1}{16\,(\delta-3)} + \frac{1}{6\,(\delta-3)^3}\right)\left(\sqrt{1-\delta^2}\right)^3 \times \frac{n^{-5/2}\delta^{-n}}{\Gamma\,(-3/2)}\left(1+O\left(\frac{1}{n}\right)\right)$$
$$\sim \frac{\left(3\sqrt{2}-4\right)^{3/2}}{8\sqrt{2\pi}} n^{-5/2}\left(3+2\sqrt{2}\right)^n. \qquad \square$$

Connected outerplanar graphs. We now turn to the problem of asymptotic enumeration of connected outerplanar graphs. First we have to establish the singular development of the generating function for vertex rooted connected outerplanar graphs $\hat{C}\,(x)$. Let η be the radius of convergence of $\hat{C}\,(x)$. Observe that the coefficients $\hat{c}_n$ are bounded from below by the number of unlabeled vertex rooted dissections v_n, which has exponential growth > 1. Furthermore, these coefficients are bounded from above by the number of embedded outerplanar graphs with a root edge,which also has exponential growth > 1 (this follows from classical enumerative results on planar maps; see [138]). Hence η is in $(0,1)$.

To apply Theorem 2.4.3 for rooted connected outerplanar graphs, we consider the function

$$H\,(x,y) := x\exp\left(\frac{Z(\mathcal{V};y,\hat{C}(x^2))}{y} + \sum_{k\geq 2}\frac{Z(\mathcal{V};\hat{C}(x^k),\hat{C}(x^{2k}))}{k\,\hat{C}(x^k)}\right) - y.$$

Observe that (9.2.3) implies that $H(x,\hat{C}(x)) = 0$. The difficulty in the application of the singular implicit functions theorem (Thm. 2.4.3) is the verification of the requirements of this theorem. Hence, to apply Theorem 2.4.3, we have to check that the dominant singularity of the generating functions for the connected components is determined by its implicit definition (like (9.2.3)) and not by a singularity of $H(x,y)$. This analysis is the main purpose of the next proposition. Observe that it can also be easily generalized to other classes of connected unlabeled graphs with known blocks.

Lemma 9.3.2. *The generating function $\hat{C}(x)$ satisfies the conditions of Theorem 2.4.3 with the function $H\,(x,y)$ from (9.3) and $(r,s) = (\eta,\tau)$, where η is the dominant singularity of $\hat{C}(x)$ and $\tau := \lim_{x\to\eta^-}\hat{C}(x)$.*

As a consequence, Theorem 2.4.3 ensures that $\hat{C}(x)$ has a singular expansion

$$\hat{C}\,(x) = \sum_{k\geq 0}\hat{C}_k X^k, \tag{9.3.1}$$

where

$$X := \sqrt{1-\frac{x}{\eta}}, \quad \hat{C}_0 = \tau, \quad \hat{C}_1 = -\sqrt{\frac{2\eta\frac{\partial}{\partial x}H\,(\eta,\tau)}{\frac{\partial^2}{\partial y^2}H\,(\eta,\tau)}},$$

and constants $\hat{C}_k$, $k \geq 2$, are computable from the derivatives of $H\,(x,y)$ at (η,τ).

Proof. The conditions $H(0,0) = 0$ and $\frac{\partial}{\partial y}H(0,0) = -1$ can easily be verified. The positivity conditions on the coefficients of $H(x,y)$ follow from the positivity

of the coefficients of $Z(\mathcal{V})$. The analyticity domain of $H(x,y)$ is determined by the dominant singularities of $Z(\mathcal{V})$; that is, $H(x,y)$ is analytic for x and y such that $|y| < \delta$ and $|x^l| < \eta$ and $|\hat{C}(x^l)| < \delta$ for each $l \geq 2$. Since $\hat{C}(x)$ is strictly increasing for positive x and since $\eta < 1$, we have $|\hat{C}(x^l)| \leq |\hat{C}(x^2)|$ for all $l \geq 2$ and $|x| < \eta$. Therefore, $H(x,y)$ is analytic for $|x| < R := \min(\sqrt{\eta}, \sqrt{\hat{C}^{-1}(\delta)})$ and $|y| < S := \delta$.

We show next that $\eta < R$ and $\tau < S$. Let $\tilde{H}(x,y) := H(x,y) + y$. $\tilde{H}(x,y)$ satisfies $\tilde{H}(x, \hat{C}(x)) = \hat{C}(x)$ and has the same domain of analyticity as $H(x,y)$. Assume $\tau > \delta$. Then there exists $x_0 < \eta$ such that $\hat{C}(x_0) = \delta$. Observe that, if $|x| < x_0$ then $|\hat{C}(x^2)| \leq |\hat{C}(x)| < \hat{C}(x_0) = \delta$. Thus $(x, \hat{C}(x))$ is in the analyticity domain of $\tilde{H}(x,y)$, so that $\tilde{H}(x, \hat{C}(x)) = \hat{C}(x)$. By continuity we obtain $\tilde{H}(x_0, \hat{C}(x_0)) = \hat{C}(x_0)$. We have now the contradiction that $\hat{C}(x)$ is analytic at x_0 since $x_0 < \eta$, whereas $\tilde{H}(x, \hat{C}(x))$ is singular at x_0 because $\hat{C}(x_0) = \delta$. Hence $\tau \leq S = \delta$. Since we know that $\tau \leq \delta$, i.e., $\eta \leq \hat{C}^{-1}(\delta)$, we get $R = \sqrt{\eta} > \eta$. Now it remains to prove that $\tau < \delta$. Assume $\tau = \delta$. Observe from (9.1.1) and (9.2.3) that

$$\hat{C}(x) = x \exp(\Psi(\hat{C}(x)) + A(x))$$

where $\Psi(y) = 1/8(1+y-\sqrt{1-6y+y^2})$ has a dominant singularity at $y = \delta$, and where $A(x)$ is a generating function analytic for $|x| < \eta$ and having nonnegative coefficients. (This follows from the fact that $2A(x)$ is the generating function for reflective vertex rooted dissections [139].) Hence, for $0 < x < \eta$,

$$\hat{C}'(x) \geq \hat{C}'(x)\Psi'(\hat{C}(x))\hat{C}(x),$$

so that $\Psi'(\hat{C}(x)) \leq 1/\hat{C}(x)$. Thus, $\Psi'(\hat{C}(x))$ is bounded when $x \to \eta^-$, which contradicts the fact that $\lim_{y \to \delta^-} \Psi'(y) = +\infty$.

Thus, $H(x,y)$ is analytic at (η, τ) and $H(\eta, \tau) = 0$ is satisfied. As pointed out before, the dominant singularity η of $\hat{C}(x)$ is determined either by a singularity in a component of (9.2.3), or by a non-uniqueness in the definition of $\hat{C}(x)$ by (9.2.3). The relation $\tau < \delta$ excludes the first case, so that the singularity is caused by a non-uniqueness of the inversion. Hence, the derivative of $H(x,y)$ with respect to y has to vanish at $(x,y) = (\eta, \tau)$, since otherwise the implicit function theorem ensures a (unique) analytic continuation of $\hat{C}(x)$ at $x = \eta$. Therefore, the equations from (2.4.7) are satisfied for $(r,s) = (\eta, \tau)$.

Furthermore, one can see that

$$\begin{aligned} \frac{\partial^2}{\partial y^2} H(x,y)\Big|_{(x,y)=(\eta,\tau)} &= \frac{1}{\tau} + \frac{\partial^2}{\partial s_1^2} \frac{Z(\mathcal{V}; s_1, \hat{C}(\eta^2))}{s_1}\Big|_{s_1=\tau} \\ &= \frac{1}{\tau} + \frac{\tau}{(\tau^2 - 6\tau + 1)^{3/2}}, \end{aligned}$$

and

$$\begin{aligned} \frac{\partial}{\partial x} H(x,y)\Big|_{(x,y)=(\eta,\tau)} = \tau &\left(\frac{1}{\eta} + \frac{\partial}{\partial x} \frac{Z(\mathcal{V}; \tau, \hat{C}(x^2))}{\tau} \right) \\ &+ \tau \sum_{k \geq 2} \frac{Z(\mathcal{V}; \hat{C}(x^k), \hat{C}(x^{2k}))}{k\, \hat{C}(x^k)}\Big|_{x=\eta}. \end{aligned}$$

From $0 < \tau < \delta$ and the fact that the derivative in $\frac{\partial}{\partial x} H(\eta, \tau)$ is a derivative of a formal power series with positive coefficients evaluated at $\eta > 0$, it follows that both derivatives are strictly positive and hence do not vanish.

Finally, the aperiodicity of $\hat{C}(x)$ follows from the fact that $\hat{c}_1 \neq 0$ and $\hat{c}_2 \neq 0$. □

Theorem 9.3.3. *The function $C(x)$ has a singular expansion of the form*

$$C(x) = C(\eta) + \sum_{k \geq 2} C_k X^k, \qquad X := \sqrt{1 - \frac{x}{\eta}}, \tag{9.3.2}$$

with constants C_k, $k \geq 2$, which can be computed from the constants $\hat{C}_k$, and with η as in Lemma 9.3.2. Hence,

$$c_n \sim \frac{3C_3}{4\sqrt{\pi}}\ n^{-5/2}\ \eta^{-n},$$

where $C_3 \approx 0.017972$ and $\eta^{-1} \approx 7.503597$.

Proof. Recall (9.2.4) for the ordinary generating function for connected outerplanar graphs,

$$C(x) = \hat{C}(x) + Z(\mathcal{B}; \hat{C}(x)) - Z(\mathcal{V}; \hat{C}(x)).$$

Since $\tau < \delta$, it is clear that the dominant singularity of $C(x)$ is the same as $\hat{C}(x)$ [63, Chap. VI.6]. The singular expansion of $C(x)$ around η can then be obtained by injecting the singular expansion of $\hat{C}(x)$ into (9.2.4):

$$\begin{aligned} C(x) =& \sum_{k \geq 0} \hat{C}_k X^k \\ &+ Z(\mathcal{B}; \sum_{k \geq 0} \hat{C}_k X^k, \hat{C}(\eta^2(1-X^2)^2), \hat{C}(\eta^3(1-X^2)^3), \cdots) \\ &- Z(\mathcal{V}; \sum_{k \geq 0} \hat{C}_k X^k, \hat{C}(\eta^2(1-X^2)^2), \hat{C}(\eta^3(1-X^2)^3), \cdots). \end{aligned} \tag{9.3.3}$$

Developing in terms of X (around $X = 0$) gives a singular expansion

$$C(x) = \sum_{k \geq 0} C_k X^k.$$

It remains to check that $C_1 = 0$ and $C_3 \neq 0$. From (9.3.3) it is clear that

$$C_1 = \hat{C}_1 + \hat{C}_1\ \frac{\partial}{\partial s_1} Z(\mathcal{B})\Big|_{(s_1,s_2)=(\tau,\hat{C}(\eta^2))} - \hat{C}_1\ \frac{\partial}{\partial s_1} Z(\mathcal{V})\Big|_{(s_1,s_2)=(\tau,\hat{C}(\eta^2))}.$$

From (2.3.6) we know $s_1 \frac{\partial}{\partial s_1} Z(\mathcal{B}) = Z(\mathcal{V})$, so that

$$C_1 = \hat{C}_1 \left(1 + \frac{Z(\mathcal{V})}{s_1} - \frac{\partial}{\partial s_1} Z(\mathcal{V})\right)\Big|_{(s_1,s_2)=(\tau,\hat{C}(\eta^2))}.$$

On the other hand, (9.3) implies that

$$\frac{\partial}{\partial y}H\left(x,y\right)=-1+$$
$$\left(H\left(x,y\right)+y\right)\left(\frac{1}{y}\,\frac{\partial}{\partial s_1}Z(\mathcal{V};y,\hat{C}(x^2))-\frac{1}{y^2}\,Z(\mathcal{V};y,\hat{C}(x^2))\right). \quad (9.3.4)$$

By (2.4.7) and Lemma 9.3.2,

$$0=\frac{\partial}{\partial y}H\left(\eta,\tau\right)=\left.\frac{\partial}{\partial s_1}Z\left(\mathcal{V}\right)-\frac{1}{s_1}Z\left(\mathcal{V}\right)-1\right|_{(s_1,s_2)=(\tau,\hat{C}(\eta^2))}.$$

Thus, $C_1=0$. Assume $C_3=0$. Then the expansion (9.3.2) yields

$$c_n\sim O(n^{-k/2-1})\eta^{-n}$$

for some odd number $k\geq 5$. This contradicts $nc_n\geq\hat{c}_n\sim-\hat{C}_1/\left(2\sqrt{\pi}\right)n^{-3/2}\eta^{-n}$ (by Lemma 9.3.2).

Next, we approximate the value of the dominant singularity η. As far as we know, the computation of analytic expressions for growth constants has not been possible for some classes of unlabeled structures that are even simpler than outerplanar graphs, for example, for trees, see [63, Sec. VII.2.3] and [113]. Nevertheless, we can simplify the problem by reducing it to one variable, and provide numerical estimates of the growth constants.

With (9.3.4) for $\frac{\partial}{\partial y}H\left(\eta,\tau\right)$ and the explicit formula for $Z\left(\mathcal{V}\right)$ from Corollary 9.1.2 the equation $\frac{\partial}{\partial y}H\left(\eta,\tau\right)=0$ becomes

$$\tau\left(1+\hat{C}(\eta^2)(\hat{C}(\eta^2)-3)-\frac{\hat{C}(\eta^2)^2(\tau-3)}{\sqrt{\tau^2-6\tau+1}}-\sqrt{\hat{C}(\eta^2)^2-6\hat{C}(\eta^2)+1}\right)$$
$$=8\hat{C}\left(\eta^2\right)^2. \quad (9.3.5)$$

With algebraic elimination [63, App. B.1], (9.3.5) can be reformulated as a system of polynomial equations, regarding $\hat{C}\left(\eta^2\right)$ as a fixed value. We obtain a polynomial equation of degree 8 in τ with coefficients $p_i\left(\eta\right)$ (depending on $\hat{C}\left(\eta^2\right)$), $i=0,\cdots,8$,

$$p_0\left(\eta\right)+p_1\left(\eta\right)\tau+p_2\left(\eta\right)\tau^2+p_3\left(\eta\right)\tau^3+p_4\left(\eta\right)\tau^4$$
$$+p_5\left(\eta\right)\tau^5+p_6\left(\eta\right)\tau^6+p_7\left(\eta\right)\tau^7+p_8\left(\eta\right)\tau^8$$
$$=0. \quad (9.3.6)$$

The solutions of (9.3.6) do not need to satisfy (9.3.5), but every τ that is a solution of (9.3.5) is also a solution of (9.3.6). We denote the solutions of (9.3.6) by $\tau_1\left(\eta\right),\cdots,\tau_8\left(\eta\right)$. It remains to solve the equations

$$H\left(\eta,\tau_i\left(\eta\right)\right)=0,\qquad i=1,\cdots,8,$$

and to pick the correct solution η. Since $H\left(x,y\right)$ depends on $\hat{C}\left(x\right)$, which we do not know explicitly, and since it includes also an infinite sum that we were not able to simplify, we can only approximate the solutions of $H\left(\eta,\tau_i\left(\eta\right)\right)=0$ by truncating the infinite sum in $H\left(x,y\right)$ at some index m and replacing $\hat{C}\left(x\right)$

with $\hat{C}^{[m]}(x) := \sum_{n=1}^{m} \hat{c}_n x^n$ for known coefficients $\hat{c}_1, \cdots, \hat{c}_m$. That is, we search for roots of the functions

$$\tilde{H}_i^{[m]}(\eta) := \eta \exp\left(\frac{Z(\mathcal{V}; \tau_i(\eta), \hat{C}^{[m]}(\eta^2))}{\tau_i(\eta)} + \sum_{k=2}^{m} \frac{Z(\mathcal{V}; \hat{C}^{[m]}(\eta^k), \hat{C}^{[m]}(\eta^{2k}))}{k\, \hat{C}^{[m]}(\eta^k)} \right) - \tau_i(\eta),$$

$i = 1, \cdots, 8$, in the interval $(0, 1)$. We solve the equation $\tilde{H}_i^{[m]}(\eta) = 0$ for $m = 25$ numerically, select the correct root, and obtain the estimates

$$\eta \approx 0.1332694 \quad \text{and} \quad \tau \approx 0.1707560.$$

The residuals in the equations $\tilde{H}_i^{[m]}(\eta, \tau) = 0$ and $\frac{\partial}{\partial y} \tilde{H}_i^{[m]}(\eta, \tau) = 0$ have an order of 10^{-58}.

□

Outerplanar graphs. The singular expansion and the asymptotic number of outerplanar graphs are as follows.

Theorem 9.3.4. *The function $G(x)$ has a singular expansion of the form*

$$G(x) = G(\eta) + \sum_{k \geq 2} G_k X^k, \qquad X := \sqrt{1 - \frac{x}{\eta}},$$

where η is as in Lemma 9.3.2, and where the constants G_k, $k \geq 2$, can be computed from the constants C_k, in particular $G_3 = G(\eta) C_3$. Furthermore, g_n satisfies

$$g_n = \sum_{k \geq 1} \binom{n + k - \frac{1}{2}}{n} G_{2k+1} \eta^{-n},$$

and in particular

$$g_n \sim \frac{3G_3}{4\sqrt{\pi}}\ n^{-5/2}\ \eta^{-n},$$

where $G_3 \approx 0.021504$ and $\eta^{-1} \approx 7.503597$.

Proof. Recall (9.2.5) for the ordinary generating function for outerplanar graphs,

$$G(x) = \exp\left(\sum_{k \geq 1} \frac{1}{k} C\left(x^k\right) \right).$$

As the exponential function $\exp(\cdot)$ is regular, the dominant singularity of $G(x)$ is the same as $C(x)$. Replacing $C(x)$ by its singular expansion (9.3.2) and x^k by $\eta^k \left(1 - X^2\right)^k$ for $k \geq 2$, we get

$$G(x) = \exp\left(C(\eta) + \sum_{k \geq 2} C_k X^k + \sum_{k \geq 2} \frac{1}{k} C(\eta^k (1 - X^2)^k) \right),$$

from which the singular expansion of $G(x)$ can be computed. Then, by Lemma 2.4.2 we derive the asymptotic estimate of g_n. □

Bipartite outerplanar graphs. Finally, using the same techniques as for the general case, we can compute the asymptotic estimate of bipartite outerplanar graphs.

Theorem 9.3.5. *Let $(g_b)_n$ be the number of bipartite outerplanar graphs on n vertices. Then*

$$(g_b)_n \sim b\ n^{-5/2}\ \eta_b^{-n}$$

with $\eta_b^{-1} \approx 4.57717$.

9.4 Random outerplanar graphs

This section investigates typical properties of a random (unlabeled) outerplanar graph chosen uniformly at random among all the unlabeled outerplanar graph on n vertices. We first discuss the probability of a random outerplanar graph being connected, and the number and type of components, and then proceed with the distribution of the number of edges.

9.4.1 Connectedness

We start with the probability that a random outerplanar graph is connected.

Theorem 9.4.1. *The probability that a random outerplanar graph is connected is asymptotically $c/g \approx 0.845721$.*

Proof. The probability that a random outerplanar graph on n vertices is connected is exactly c_n/g_n. The asymptotic estimates for c_n and g_n from Theorem 9.3.3 and Theorem 9.3.4 yield $c_n/g_n \sim C_3/G_3 \approx 0.845721$. □

9.4.2 Components

The number of components can be studied by considering a bivariate generating function for outerplanar graphs with a variable that counts the number of components.

Theorem 9.4.2. *The expected number of components in a random outerplanar graph is asymptotically equal to a constant ≈ 1.17847.*

Proof. Let κ_n denote the number of components in a random outerplanar graph on n vertices and let

$$G(x,u) := \exp\left(\sum_{k\geq 1} \frac{1}{k} u^k C\left(x^k\right)\right)$$

be the generating function for outerplanar graphs, where the additional variable u marks the number of components. Thus, the probability that an outerplanar graph has k components is

$$\Pr(\kappa_n = k) = \left[x^n u^k\right] G(x,u)\,/g_n,$$

and the expected number of components is

$$\begin{aligned}\mathbb{E}(\kappa_n) = \frac{1}{g_n}\sum_{k\geq 1} k\left[x^n u^k\right] G(x,u) &= \frac{1}{g_n}[x^n]\frac{\partial}{\partial u}G(x,1)\\ &= \frac{1}{g_n}[x^n]\,G(x)\sum_{k\geq 1} C\left(x^k\right).\end{aligned}$$

By asymptotic expansion around $x=\eta$, we obtain

$$[x^n]\,G(x)\sum_{k\geq 1} C\left(x^k\right) \sim G(\eta)\,C_3(1+\sum_{r\geq 1} C(\eta^r))\frac{n^{-5/2}\eta^{-n}}{\Gamma(-3/2)},$$

which together with Theorem 9.3.4, more precisely

$$g_n \sim G(\eta)\,C_3\frac{1}{\Gamma(-3/2)}n^{-5/2}\eta^{-n},$$

yields

$$\mathbb{E}(\kappa_n) \sim 1+\sum_{r\geq 1} C(\eta^r) \approx 1.17847. \qquad \square$$

Given a family $\mathcal{A}$ of unlabeled connected outerplanar graphs, we can make the following statements about the probability that a random outerplanar graph has exactly k components in $\mathcal{A}$. Denote the number of graphs in $\mathcal{A}$ that have exactly n vertices by a_n, and let $A(x) := \sum_n a_n x^n$.

Theorem 9.4.3. *Given an unlabeled outerplanar graph G with n vertices, let $\kappa_n^{\mathcal{A}}$ be the number of connected components of G belonging to $\mathcal{A}$. If the radius of convergence α of $A(x)$ is strictly larger than η, that is, a_n is exponentially smaller than c_n, then the probability that a random outerplanar graph with n vertices has exactly $k \geq 0$ components belonging to $\mathcal{A}$ converges to a discrete law:*

$$\Pr(\kappa_n^{\mathcal{A}} = k) \sim Z(S_k; A(\eta))\exp\left(-\sum_{r\geq 1}\frac{1}{r}A(\eta^r)\right),$$

and the expected number of components belonging to $\mathcal{A}$ in a random outerplanar graph with n vertices is

$$\mathbb{E}(\kappa_n^{\mathcal{A}}) \sim \sum_{r\geq 1} A(\eta^r).$$

Proof. Let $G^{\mathcal{A}}(x,u)$ be the bivariate generating function for unlabeled outerplanar graphs, where the additional variable u marks the number of components belonging to $\mathcal{A}$,

$$\begin{aligned}G^{\mathcal{A}}(x,u) &:= \exp\left(\sum_{k\geq 1}\frac{1}{k}\left(u^k A\left(x^k\right) + \left(C\left(x^k\right) - A\left(x^k\right)\right)\right)\right)\\ &= G(x)\exp\left(\sum_{k\geq 1}\frac{u^k-1}{k}A\left(x^k\right)\right).\end{aligned}$$

Then we have

$$\Pr(\kappa_n^{\mathcal{A}} = k) = \left[x^n u^k\right] G^{\mathcal{A}}(x,u)/g_n.$$

Since $A(x)$ is analytic at η, the dominant singularity of $G^{\mathcal{A}}(x,u)$ for fixed u is determined by $G(x)$. Thus we get

$$[x^n u^k] G^{\mathcal{A}}(x,u) \underset{n\to\infty}{\sim} [u^k] \exp\left(\sum_{k\geq 1} \frac{u^k - 1}{k} A(\eta^k)\right) [x^n] G(x),$$

i.e.,

$$\begin{aligned}\Pr(\kappa_n^{\mathcal{A}} = k) &\underset{n\to\infty}{\sim} \left[u^k\right] \exp\left(\sum_{k\geq 1} \frac{u^k - 1}{k} A\left(\eta^k\right)\right) \\ &= Z\left(S_k; A(\eta)\right) \exp\left(-\sum_{k\geq 1} \frac{1}{k} A\left(\eta^k\right)\right).\end{aligned}$$

For the expectation of $\kappa_n^{\mathcal{A}}$ we again use

$$\mathbb{E}(\kappa_n^{\mathcal{A}}) = \frac{1}{g_n}[x^n] \frac{\partial}{\partial u} G^{\mathcal{A}}(x,1) = \frac{1}{g_n}[x^n] G(x) \sum_{k\geq 1} A\left(x^k\right).$$

The statement follows from the analyticity of $A(x)$ at η and Theorem 9.3.4. □

9.4.3 Isolated vertices

The asymptotic distribution of the number of isolated vertices in a random outerplanar graph can now be easily computed.

Theorem 9.4.4. *The number of isolated vertices in a random outerplanar graph has asymptotically* geometric law *with parameter η. In particular, the expected number of isolated vertices in a random outerplanar graph is asymptotically $\eta/(1-\eta) \approx 0.153761$.*

Proof. Let $\mathcal{A}$ be the family consisting of the graph that is a single vertex, i.e., $A(x) = x$. By Theorem 9.4.3, we have

$$\Pr(\kappa_n^{\mathcal{A}} = k) \sim \eta^k/(1-\eta),$$

since $Z\left(S_k; A(\eta)\right) = \eta^k$ and $\sum_r \frac{1}{r} A(\eta^r) = \log(1-\eta)$. In other words, the distribution of the number of isolated vertices $\kappa_n^{\mathcal{A}}$ is asymptotically a geometric law with parameter η. □

Other consequences of Theorem 9.4.3 concern the number of two-connected components and the number of bipartite components in a random outerplanar graph.

Corollary 9.4.5. *In a random outerplanar graph, the expected number of connected components that are two-connected is asymptotically $\sum_{k\geq 1} B\left(\eta^k\right) \approx 0.175054$.*

Proof. Let $\mathcal{A} := \mathcal{B}$ be the family of dissections, $A(x) = B(x)$. The radius of convergence of $B(x)$ is $\delta > \eta$ (Lemma 9.3.2). Hence, by Theorem 9.4.3, $\mathbb{E}(\kappa_n^{\mathcal{B}}) = \sum_{k\geq 1} B\left(\eta^k\right)$. □

Corollary 9.4.6. *In a random outerplanar graph, the expected number of connected components that are bipartite is asymptotically*

$$\sum_{k\geq 1} C_b\left(\eta^k\right) \approx 0.175427,$$

where $C_b(x)$ denotes the generating function for bipartite connected outerplanar graphs.

Proof. We apply Theorem 9.4.3 with $\mathcal{A} = \mathcal{C}_b$. □

9.4.4 Number of edges

In this section, we analyse the distribution of the number of edges in a random outerplanar graph. To do this, we add a variable y whose power (in the cycle index sums and generating functions) indicates the number of edges. For a graph G on n vertices and m edges, and with the automorphism group $\Gamma(G)$ (acting on the vertices), we define

$$Z(G; s_1, s_2, \cdots; y) := Z(\Gamma(G); s_1, s_2, \cdots; y) := y^m \frac{1}{|\Gamma(G)|} \sum_{\alpha\in\Gamma(G)} \prod_{k=1}^{n} s_k^{j_k(\alpha)}.$$

Taking the number of edges into account we can derive the cycle index sums for all encountered families of outerplanar graphs with the additional parameter marking the edges.

$$Z(\hat{\mathcal{C}}) = s_1 \exp\left(\sum_{k\geq 1} \frac{1}{k} \frac{Z(\mathcal{V}; Z(\hat{\mathcal{C}}; s_k, s_{2k}; y^k), Z(\hat{\mathcal{C}}; s_{2k}, s_{4k}; y^{2k}), \cdots; y^k)}{Z(\hat{\mathcal{C}}; s_k, s_{2k}; y^k)}\right),$$

$$Z(\mathcal{C}) = Z(\hat{\mathcal{C}}) + Z(\mathcal{B}; Z(\hat{\mathcal{C}})) - Z(\mathcal{V}; Z(\hat{\mathcal{C}})),$$

$$Z(\mathcal{G}) = \exp\left(\sum_{k\geq 1} \frac{Z\left(\mathcal{C}; s_k, s_{2k}, \cdots; y^k\right)}{k}\right).$$

Similarly as in Section 9.2, the coefficients counting outerplanar graphs with respect to the number of vertices and the number of edges can be extracted in polynomial time from the expressions of the cycle index sums.

With the help of Theorem 2.4.4, we can study the limit distributions of the number of edges in a random dissection and in a random outerplanar graph, respectively.

Theorem 9.4.7. *The distribution of the number of edges in a random outerplanar graph on n vertices is asymptotically* Gaussian *with mean μn and variance $\sigma^2 n$, where*

$$\mu \approx 1.54894, \quad \sigma^2 \approx 0.227504.$$

The same holds for random connected outerplanar graphs with the same mean and variance.

Proof. We first determine the distribution of the number of edges in a rooted connected outerplanar graph. The generating function $\hat{C}(x,y)$ is implicitly defined by

$$\hat{C}(x,y) = x \exp\left(\sum_{k\geq 1} \frac{Z(\mathcal{V}; \hat{C}(x^k,y^k); y^k)}{k\,\hat{C}(x^k,y^k)} \right).$$

In order to apply the singular implicit functions theorem 2.4.3 for the function $x \mapsto \hat{C}(x,y)$ with a fixed y close to 1, we define

$$H(x,y,z) := \\ x \exp\left(\frac{Z(\mathcal{V}; z, \hat{C}(x^2,y^2); y)}{z} + \sum_{k\geq 2} \frac{Z(\mathcal{V}; \hat{C}(x^k,y^k), \hat{C}(x^{2k},y^{2k}); y^k)}{k\,\hat{C}(x^k,y^k)} \right) - z.$$

We search for a solution $(x,z) = (\eta(y), \tau(y))$ of the system

$$H(x,y,z) = 0, \qquad \frac{\partial}{\partial z} H(x,y,z) = 0, \tag{9.4.1}$$

such that $(\eta(y), \tau(y))$ is in the analyticity domain of $(x,z) \mapsto H(x,y,z)$.

For $y = 1$, the solution is at $x = \eta$, $z = \tau$ by Lemma 9.3.2. Then the classical implicit functions theorem, applied to the system (9.4.1), ensures that the solution $(\eta, 1, \tau)$ can be extended into solutions $(\eta(y), y, \tau(y))$ for y close to 1, where the functions $\eta(y)$ and $\tau(y)$ are analytic in a neighbourhood of 1. To apply the classical implicit function theorem on system (9.4.1), it remains to check that the determinant of the Jacobian of system (9.4.1), with respect to x and z,

$$\begin{pmatrix} \dfrac{\partial}{\partial x} H(x,y,z) & \dfrac{\partial}{\partial z} H(x,y,z) \\ \dfrac{\partial}{\partial x}\dfrac{\partial}{\partial z} H(x,y,z) & \dfrac{\partial}{\partial z}\dfrac{\partial}{\partial z} H(x,y,z) \end{pmatrix},$$

does not vanish at $(x,y,z) = (\eta(1), 1, \tau(1))$. This is clear, since from Lemma 9.3.2 we have

$$\begin{aligned} &\frac{\partial}{\partial z} H(\eta(1), 1, \tau(1)) = 0, \\ &\frac{\partial}{\partial x} H(\eta(1), 1, \tau(1)) \neq 0, \\ &\frac{\partial^2}{\partial z^2} H(\eta(1), 1, \tau(1)) \neq 0. \end{aligned}$$

Hence, there exist analytic functions $\eta(y)$ and $\tau(y)$ such that

$$\begin{aligned} &H(\eta(y), y, \tau(y)) = 0, \qquad \frac{\partial}{\partial z} H(\eta(y), y, \tau(y)) = 0, \\ &\frac{\partial^2}{\partial z^2} H(\eta(y), y, \tau(y)) \neq 0, \qquad \frac{\partial}{\partial x} H(\eta(y), y, \tau(y)) \neq 0 \end{aligned} \tag{9.4.2}$$

for y close to one. In addition, these solutions are in the analyticity domain of $(x,z) \mapsto H(x,y,z)$ for y close to 1, by analyticity of $(x,y,z) \mapsto H(x,y,z)$ at

$(\eta, 1, \tau)$. Next, the singular implicit functions theorem 2.4.3 yields a singular expansion

$$\hat{C}(x,y) = \sum_{k\geq 0} \hat{C}_k(y) \left(\sqrt{1 - x/\eta(y)}\right)^k$$

with coefficients $\hat{C}_k(y)$ analytic at $y = 1$ and verifying $\hat{C}_1(y) \neq 0$ for y close to 1.

To find $\eta'(1)$ and $\eta''(1)$ we compute the first and second derivatives of the equations in (9.4.2) with respect to y, and express $\eta'(y)$ and $\eta''(y)$ in terms of $\eta(y)$, $\tau(y)$, and the partial derivatives of $H(x,y,z)$ at $(x,z) = (\eta(y), \tau(y))$. Using the approximated values we obtain

$$\begin{aligned} \eta'(1) &\approx -0.206426, \\ \eta''(1) &\approx 0.495849, \\ -\frac{\eta''(1)}{\eta(1)} - \frac{\eta'(1)}{\eta(1)} + \left(\frac{\eta'(1)}{\eta(1)}\right)^2 &\approx 0.227504 \neq 0. \end{aligned}$$

Theorem 2.4.4 implies that the distribution of the number of edges in a random rooted connected outerplanar graph with n vertices asymptotically follows a Gaussian law with mean μn and variance $\sigma^2 n$, where $\mu = -\frac{\eta'(1)}{\eta(1)} \approx 1.54894$ and $\sigma^2 \approx 0.227504$. The same holds for unrooted connected outerplanar graphs and for outerplanar graphs, since their generating functions have the same dominant singularity. □

9.4.5 Chromatic number

Using Theorems 9.3.4 and 9.3.5 we prove that the chromatic number of a random outerplanar graph is asymptotically almost surely 3.

Theorem 9.4.8. *Let G_n be a random outerplanar graph on n vertices. Then*

$$\lim_{n\to\infty} \Pr[\chi(G_n) = 3] = 1.$$

Proof. It is easy to see that $\chi(G_n) \leq 3$ for all outerplanar graphs on n vertices. Hence, it remains to show that $\Pr[\chi(G_n) \leq 2] \to 0$ for $n \to \infty$, that is, we have to show that almost all outerplanar graphs are not bipartite. Since $\eta_b > \eta$, we obtain that for some positive constant c

$$\Pr[\chi(G_n) \leq 2] = \frac{(g_b)_n}{g_n} \sim c\frac{\eta_b^{-n}}{\eta^{-n}} \to 0 \qquad (n \to \infty). \qquad \square$$

9.5 Recursive counting and uniform sampling

The exact number of unlabeled but *rooted connected* outerplanar graphs can be found by the same decomposition strategy.

Let $p(n)$ be the number of all unlabeled connected *rooted* outerplanar graphs with n vertices. We distinguish two cases: The case where the root is a cutvertex, and the case with a unique root block. In the first case, the graph without the root has several connected components. Let $p_l(n)$ be the number of all unlabeled connected rooted outerplanar graphs with n vertices where all these

components have size $\leq l$; thus clearly $p_1(n) = 1$ (which counts a so-called *star* graph). Let $q(n)$ be the number of all unlabeled connected outerplanar graphs with n vertices and a unique root block. Then summing over the number k of components of size l we get, for $n \geq 2$ and $1 \leq l \leq n-1$,

$$p(n) = p_{n-1}(n), \qquad p_l(n) = \sum_{k=0}^{\lfloor (n-1)/l \rfloor} \binom{q(l+1)+k-1}{k} p_{l-1}(n-kl).$$

In the second case, where the outerplanar graph has a unique root block, it is not so easy to avoid double counting. The number of unlabeled connected outerplanar graphs where the root block consists of a single edge is $p(n-1)$. But otherwise we have to distinguish between two cases: The outerplanar graph might or might not have a symmetry, in other words, it might or might not have a nontrivial automorphism that fixes the root vertex, and exchanges the two vertices on the unique Hamiltonian cycle of the root block that have the same distance to the root. According to that, let $a(n)$ denote the number of unlabeled connected *asymmetric* outerplanar graphs with n vertices. Correspondingly $b(n)$ denotes the number of unlabeled connected *symmetric* such graphs. Then for $n \geq 3$,

$$q(n) = p(n-1) + a(n) + b(n).$$

To compute the symmetric (resp. asymmetric) outerplanar graphs with a unique root block let $b_l(n)$ (resp. $a_l(n)$) be the number of unlabeled connected symmetric (resp. asymmetric) outerplanar graphs with n vertices that do not have a cutvertex at distance $\leq l$ from the root on the unique Hamiltonian cycle of the root block. For $n \geq 3$,

$$b_l(n) = \sum_{i=1} p(i+1) b_{l+1}(n-2i).$$

An outerplanar graph with a unique root block can be asymmetric for three disjoint reasons: either the graph formed by deleting a graph attached to the cut vertex with distance l to the root is already asymmetric, or two graphs of different size are attached to the cutvertices with distance l to the root, or the two graphs attached to the cutvertices with distance l to the root are distinct for another reason.

$$\begin{aligned} a_l(n) = &\sum_{i,j=0} p(i+1)p(j+1)a_{l+1}(n-i-j) \\ &+ \sum_{i \neq j \geq 0} p(i+1)p(j+1)b_{l+1}(n-i-j) + \sum_{i=0} \binom{p(i+1)}{2} b_{l+1}(n-2i). \end{aligned}$$

It remains to clarify the initial cases for $a_l(n)$ and $b_l(n)$, where $l = n$. In the sense of above, we also distinguish between symmetric and asymmetric (with respect to the root vertex) *biconnected* outerplanar graphs. So, let $c(n)$ be the number of unlabeled connected symmetric biconnected graphs with n vertices. Clearly $b_n(n) = c(n)$ and $a_n(n) = (s(n-1) - c(n))/2$, where $s(n)$ is the n-th Schröder number that was already mentioned in Section 5.4. Using $s(n)$

we can also calculate $c(n)$ as follows: $c(1) = c(2) = c(3) = 1$ and for $n \geq 2$, $0 \leq l \leq \lceil n/2 \rceil$,

$$c(2n) = c(2n+1) \;=\; c(2n-1) + 2\sum_{i=2}^{n} s(i)c(2n+1-2i).$$

Thus we complete the exact counting formulas for unlabeled connected rooted outerplanar graphs. It is easy to see that the number of unlabeled connected outerplanar graphs only grows exponentially, and thus the values of this section can be stored in linear space. To compute the numbers using dynamical programming, we use at most two-dimensional table, each of whose entry involves the summation running over at most two parameters and a multiplication with large numbers. Thus the running time for the computation of the values is within $\tilde{O}(n^6)$ with space $O(n^3 \log n)$.

The decomposition and the recursive counting formulas derived in this section can be used to generate an unlabeled rooted connected outerplanar graph uniformly at random, with one exception: We did not decompose asymmetric biconnected outerplanar graphs, and counted them by counting all biconnected outerplanar graphs and *subtracting* the symmetric objects. To generate a random asymmetric biconnected outerplanar graph, we use *rejection sampling*: First we generate an arbitrary biconnected rooted outerplanar graph, and accept it only in the case that it is asymmetric. Otherwise we restart the procedure. It is easy to see that the graph will be asymmetric with high probability, and therefore we obtain an expected polynomial time sampling procedure.

Using the uniform generation procedure for unlabeled *rooted* connected outerplanar graph derived above we can derive a uniform generation procedure for unlabeled *unrooted* connected outerplanar graphs, again using rejection sampling. The algorithm first generates a random *rooted* connected outerplanar graph. Then it outputs the graph with probability $\frac{1}{o}$, where o is the number of orbits in the automorphism group of the unrooted connected outerplanar graph. Otherwise we say that the graph is *rejected*, and the algorithm is restarted. Since the probability that the graph is rejected is at most $1 - \frac{1}{n}$, the expected number of restarts is at most

$$\sum_{i=1}^{\infty} i \left(1 - \frac{1}{n}\right)^{i-1} \frac{1}{n} = n.$$

It is easy to compute the orbit of the root using an efficient linear time procedure for isomorphism testing of outerplanar graphs (one could even use the linear time isomorphism testing algorithm for planar graphs [78, 79], checking for each position of the root in the graph whether there is an automorphism that maps the vertex of the root to this vertex).

Theorem 9.5.1. *An unlabeled outerplanar graph on n vertices can be sampled uniformly at random in expected time $\tilde{O}(n^7)$ with space $O(n^3 \log n)$.*

Using generating functions for outerplanar graphs derived in Section 9.2 one can derive a *Boltzmann sampler*. Note however that the derived generating functions involve subtraction, which requires an extra cost for rejection sampling. To overcome this problem a new operator, called *cycle-pointing*, is introduced in [27], which yields generating functions for a Boltzmann sampler that runs in linear time.

Chapter 10

Unlabeled 2-connected Planar Graphs

One of the challenging open problems concerning planar structures is the enumeration and uniform sampling of unlabeled planar graphs. As a step toward this task we study unlabeled *two-connected* planar graphs and present an algorithm that generates unlabeled two-connected planar graphs on m edges uniformly at random in expected time $O(m^9)$.

Unlabeled planar graphs have in general many automorphisms and also might have many embeddings on the sphere. A standard way to destroy such an automorphism is to mark a vertex, an edge, or a face of a graph. To count and sample unlabeled 2-connected planar graphs, we first *root* them by marking a directed edge. Then rooted unlabeled 2-connected planar graphs are counted up to isomorphisms that map the root of one graph to the root of the other graph. We also call such a rooted 2-connected planar graph a *(planar) network*.

We decompose networks along their connectivity structure and apply the recursive method for uniform generation. Clearly, generating a random *rooted* 2-connected planar graph and then simply *ignoring* the root edge does not yield the uniform distribution, since unlabeled graphs might correspond to different numbers of rooted graphs. But this imbalance can be compensated by *rejection sampling*, i.e., the sampling procedure is restarted with a probability that is inverse proportional to the size of the orbit of the root. In this way we can sample unlabeled 2-connected planar graphs in expected polynomial time, uniformly at random.

In principle, our approach here is similar to the one described for labeled planar graphs, but for unlabeled structures several new techniques are necessary. A classical theorem of Whitney (see e. g., [54]) says that a rooted 3-connected planar graph, i.e., 3-connected networks, can have either one or two embeddings in the plane where the root edge is embedded on the outer face. Such embedded three-connected networks are called *c-nets*. In the case that both embeddings of the 3-connected graph are isomorphic, we say that it has a *sense-reversing automorphism* or it is *symmetric*.

In order to count symmetric c-nets we prove a new bijective correspondence to *coloured networks* (defined below), and a decomposition of these objects. We also need to consider rooted graphs with an automorphism that reverts the

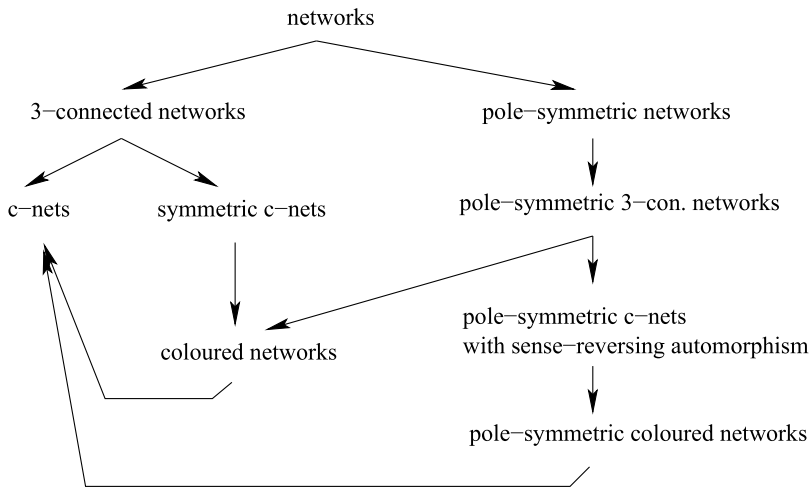


Figure 10.1: Dependencies of the concepts.

direction of the root. We say that such a graph is *pole-symmetric*. We present a decomposition of pole-symmetric networks, and finally also a decomposition of pole-symmetric c-nets with a sense-reversing automorphism. It will be useful to note that the dual of a pole-symmetric c-net is a c-net with a sense-reversing automorphism.

As a final step we use a *deterministic* polynomial time generation algorithm for c-nets of [28]. Note that the generation algorithm of [9] for such objects has *expected* polynomial running time. We need the algorithm for c-nets of [28], since it can easily be adapted to generate c-nets with a certain specified number of edges on the outer face, which we need in the generation algorithm for unlabeled 2-connected planar graphs.

The rest of the chapter is organized as follows. In Section 10.1 we decompose networks and derive recurrence formulas. In Sections 10.2 and 10.3 we prove a bijection between symmetry 3-connected planar graphs and coloured networks, and derive recurrences formulas of coloured networks. In Section 10.4 we decompose pole-symmetric networks and derive recurrence formulas. In Section 10.5 we compute the number of pole-symmetric networks with a sense-reversing automorphism, using coloured networks with a pole-exchanging automorphism. In Section 10.6 we derive a uniform sampling algorithm for unlabeled unrooted 2-connected planar graphs.

10.1 Enumeration of networks

In this section we present a decomposition of networks and derive recurrence formulas to count them. Let $n(m)$ be the number of networks with m edges. According to Theorem 7.1.1 we have $n(m) = s(m) + p(m) + h(m)$, where the functions $s(m), p(m)$, and $h(m)$ count the number of s-, p-, and h-networks with m edges, respectively.

s-networks. Note that each s-network has a unique cut vertex v that is closest to the pole s (here, *closest* is meant with respect to the length of the shortest connecting path).

$$s(m) \;=\; \sum_{j=1}(p(j)+h(j))n(m-j)\,.$$

p-networks. Let $p_l(m)$ denote the number of p-networks where the number of edges of the largest network that replaces an edge of the core is bounded by l. The index k in the formula below denotes the number of networks of order l that replace an edge in the core.

$$\begin{aligned} p(m) &= p_m(m) \\ p_l(m) &= \sum_{k=0}^{\lfloor m/l \rfloor} \binom{s(l)+h(l)+k-1}{k} p_{l-1}(m-kl)\,. \end{aligned}$$

h-networks. Let N be an h-network. Theorem 7.1.1 asserts that there is a unique rooted 3-connected network H, such that we can derive N from H by replacing edges of H with subnetworks. We call H the *core* of N and denote $H = \mathrm{core}(N)$. We call N *symmetric* if it has a *sense-reversing automorphism* φ, i. e., $\varphi \neq \mathrm{id}$, but $\varphi(s) = s$ and $\varphi(t) = t$, and *asymmetric* otherwise.

If $H = \mathrm{core}(N)$ is asymmetric, one can uniquely order its edges. The idea is to label the vertices of the core according to their occurrence in a depth first search traversal of the core, beginning with the root edge and visiting the neighbors of a vertex in clockwise order with respect to one of the (at most two) possible embeddings of the core. The edges are then labeled by the vertex labels obtained from the depth first search traversal. Then we lexicographically compare the sequence of these edge labels in the order they were visited by the depth first search. If the core is asymmetric, one of the sequences is smaller than the other; We can distinguish between the two embeddings. If the network has a symmetric core, both edge sequences are the same unless we have inserted two different subnetworks into a pair of core edges.

If $H = \mathrm{core}(N)$ is symmetric, we order its edges in the following way. We start with the edges uv where $u = \varphi(u)$ and $v = \varphi(v)$ according to the traversal; We colour such edges *blue*. Then we list the edges uv where $u = \varphi(v)$ and $v = \varphi(u)$ according to the traversal; We colour such edges *red*. We continue with the edges that are not fixed by the nontrivial automorphism φ, and order them according to the above traversal. Edges and their images, which we call *corresponding edges*, are ordered arbitrarily.

To count the number of symmetric and asymmetric h-networks we repeatedly replace subnetworks in the above order. Let $b_{b,r,l}(m)$ be the number of symmetric h-networks B with m edges where $\mathrm{core}(B)$ has b blue and r red edges, and the first l edges of $\mathrm{core}(B)$ are also edges of B. To generate such a symmetric h-network, we have to start from a symmetric core. The l+1-st edge of $\mathrm{core}(B)$, say e_{l+1}, is either blue (for $l+1 \leq b$), or red (for $b < l+1 \leq b+r$), or uncoloured (for $l+1 > b+r$). Note that e_{l+1} might be either an edge of B or might be replaced by a certain network H as follows. If it is blue, we could replace H by an arbitrary network: if it is red, we could replace H by a network with a pole-exchanging automorphism (counted by $\tilde{n}$, see Section 10.4): if it is

uncoloured, we could replace the corresponding uncoloured l+1-st and l+2-nd edges in pairs by the same copy of a network. When l equals the number of edges of B, we have a symmetric *3-connected* h-network, which will be counted in Section 10.2.

$$b_{b,r,l}(m) \;=\; \begin{cases} \sum_{j\geq 0} n(j)\, b_{b,r,l+1}(m-j) & \text{for } l+1 \leq b \\ \sum_{j\geq 0} \tilde{n}(j)\, b_{b,r,l+1}(m-j) & \text{for } b < l+1 \leq b+r \\ \sum_{j\geq 0} n(j)\, b_{b,r,l+2}(m-2j) & \text{for } l+1 > b+r\,. \end{cases}$$

Let $a_l(m)$ count the number of asymmetric h-networks A with m edges, where the first l edges of core(A) are also edges of A. To generate an asymmetric h-network we could first take an asymmetric h-network and replace the l+1-st core edge by an arbitrary network. Or we could take a symmetric h-network and replace either the l+1-st core edge (which is red for $b < l+1 \leq b+r$) by a network with no pole-exchanging automorphism, or the corresponding l+1-st and l+2-nd edges (which are uncoloured for $l+1 > b+r$) by two different subnetworks. When l is equal to the number of the edges of A, we have to count the number of asymmetric 3-connected h-networks, which we consider in Section 10.2. For the enumeration of networks with and without a pole-exchanging automorphism we refer to Section 10.4.

$$\begin{aligned} a_l(m) \;&=\; \sum_{j\geq 0} n(j)\, a_{l+1}(m-j) \;+\; \sum_{I} (n(j)-\tilde{n}(j))\, b_{b,r,l+1}(m-j) \\ &\quad+\; \sum_{II} (n(i)n(j)\, b_{b,r,l+2}(m-i-j) - n(j)\, b_{b,r,l+2}(m-2j))/2 \end{aligned}$$

where I stands for the indices b, r, j satisfying $0 \leq b \leq l, \max\{0,l\} < b+r, j \geq 0$ and II the indices b, r, i, j satisfying $0 < b+r \leq l, i, j \geq 0$. With these numbers we can compute $h(m) = a_0(m) + \sum_{b,r\geq 0, 0<b+r} b_{b,r,0}(m)$.

We finally end up with the problems (i) to count and sample networks *with a pole-exchanging automorphism* – see Section 10.4, (ii) to count and sample 3-connected *symmetric* networks – see Section 10.2, and (iii) to sample 3-connected *asymmetric* networks.

For the last task, we apply rejection sampling. That is, we first generate an arbitrary 3-connected network. We then check whether it has such a symmetry, which can be done in linear time [78, 79]. If yes, we restart the algorithm. If no, we output the asymmetric network. Since almost all 3-connected networks do not have a sense-reversing automorphism (see [146] for a much stronger result), the expected number of restarts is constant, and we obtain an expected polynomial time algorithm.

10.2 Symmetric c-nets and coloured networks

This section contains one of the main ideas to deal with symmetries when counting unlabeled planar graphs. We want to count 3-connected planar networks with a distinguished directed edge, up to isomorphisms that fix this edge. There might be one or two embeddings where the root lies at the outer face.

As mentioned in the introduction, embedded 3-connected networks are called *c-nets* and counting formulas and sampling procedures for c-nets are known. If a network has a nontrivial automorphism that fixes the root edge, we call this

automorphism *sense-reversing*, and say that the network is *symmetric*. In this case we only have one corresponding c-net. Hence, if we can compute the number of symmetric 3-connected networks, then we can also compute the number of asymmetric 3-connected networks.

Let H be a symmetric 3-connected planar network, and φ its nontrivial sense-reversing automorphism. A vertex v of H is coloured *blue* if $\varphi(v) = v$, and *red* if v is connected to $\varphi(v)$ by an edge. The edge $v\varphi(v)$ is also coloured red. An edge uv of a coloured network is blue if both u and v are blue. (Red and blue edges were already defined in Section 10.1.) Thus a vertex or an edge is either blue, red, or uncoloured, and the poles and the root are blue.

We can think of H as being embedded in the plane in such a way that φ corresponds to a reflection, the blue vertices being aligned on the reflection axis, and the red vertices having an edge crossing this axis perpendicularly (see Fig. 10.2, left part). Our arguments, however, do not rely on such a representation.

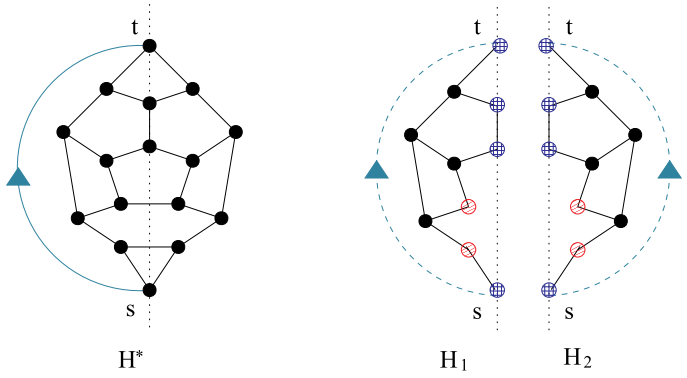


Figure 10.2: Decomposition of a symmetric rooted 3-connected graph.

If we remove from H the blue vertices and their incident edges, and also remove the red edges (that is, we cut H along the symmetry axis), then the resulting graph has exactly two connected components (see Fig. 10.2). The graphs induced by these components and the blue vertices are isomorphic and will be called H_1 and H_2. We claim that H_1^* is 2-connected, and hence H_1 is a network rooted at s and t: Suppose there is a cut-vertex in H_1^*. Then this cut-vertex together with the corresponding cut-vertex in H_2^* is a 2-cut in H^*, contradicting the 3-connectivity of H^*.

Now we extract some more properties of the graphs H_1 and H_2 and define *coloured networks*. They are defined in such a way that we can recursively decompose them, and that we can establish a bijection between symmetric h-networks and certain coloured networks.

Definition 10.2.1. *A* coloured network *is a network N, where some vertices are coloured* red *and* blue*, satisfying:*

(**P1**) *N^* has an embedding such that all coloured vertices and the poles lie on the outer face.*

(**P2**) *N and every proper subnetwork of N contain at least one coloured vertex.*

(**P3**) *No subnetwork of N has two blue poles.*

Then the bijection to symmetric 3-connected networks is as follows.

Theorem 10.2.2. *For all m, b, r there is a bijection between the following two sets of objects:*

(i) coloured networks with $(m + b - r)/2$ edges and blue poles, where b is the number of blue edges and r the number of red vertices, and

(ii) 3-connected networks with m edges having a nontrivial automorphism that fixes $b + r$ edges, and point-wise fixes the root and b other edges.

Proof. Given a symmetric 3-connected network H we first check that both H_1 and H_2, constructed as described above, are networks and satisfy properties (**P1**) – (**P3**). First, H_1 and H_2 are 2-connected: if there were a cut-vertex in H_1, we also would have a cut-vertex in H_2, and together they would form a 2-cut in H^*, contradicting the assumption that H is 3-connected. (**P1**) is immediate from the definition of H_1 and H_2. (**P2**): Every subnetwork contains a coloured vertex, since otherwise its poles would be a 2-cut in H^*. (**P3**): No subnetwork has two blue pole vertices, since these blue pole vertices would be a 2-cut in H^*.

Conversely, we construct for every coloured network H_1 with blue poles a corresponding symmetric 3-connected network H. First make an isomorphic copy H_2 of H_1. Identify corresponding blue vertices in H_1 and H_2, and add edges between corresponding red vertices in H_1 and H_2. The constructed graph H is clearly a symmetric and by (**P1**) a planar network. We finally prove that it is also 3-connected. Suppose for contradiction that H^* was not 3-connected. Then there is a split pair $\{k_1, k_2\}$ in H^* that determines at least two proper subnetworks N_1 and N_2. We distinguish four cases:

(1) Both of k_1, k_2 are blue. This is impossible because then H_1 or H_2 would contain a subnetwork with two blue poles k_1, k_2 contradicting (**P3**).

(2) Exactly one of k_1, k_2 is blue. Wlog. k_2 is blue and k_1 is in $H_1 - H_2$. Let N_1' and N_2' be those (non-empty) parts of N_1 and N_2 that also lie in H_1. By (**P2**) there are coloured vertices $v_1 \in N_1'$ and $v_2 \in N_2'$. Since H_2 is 2-connected, there is a path from v_1 to v_2 passing through H_2 and avoiding k_2 (and k_1), which contradicts the assumption that k_1, k_2 is a split pair in H.

(3) None of k_1, k_2 is blue, and either both lie in H_1 or both lie in H_2. Suppose wlog. both vertices lie in H_1. Then k_1 and k_2 define a nontrivial subnetwork in H_1. But since every such subnetwork contains a coloured vertex, this contradicts that k_1, k_2 is a 2-cut in H^*.

(4) Again none of k_1, k_2 is blue, but this time k_1 is in $H_1 - H_2$, and k_2 is in $H_2 - H_1$. It can not be that H_1 contains vertices from both N_1 and N_2, because of 2-connectivity; the same for H_2. Thus H_1 either equals N_1 or N_2, which is impossible by (**P2**), because H_1 contains a coloured vertex.

□

By Theorem 10.2.2 we can express the number of symmetric 3-connected networks as $b_{b,r,m-1}(m) = n'_{b,r}((m + b - r)/2)$.

10.3 Counting coloured networks

To derive recurrences of the number of coloured networks we observe that by Theorem 7.1.1 they are either s-, p-, or h-networks. To count s-, p-, or h-networks with m edges, b blue edges and r red vertices we introduce the corresponding functions ${}^{y}_{x}s_{b,r}(m)$, ${}^{y}_{x}p_{b,r}(m)$, and ${}^{y}_{x}h_{b,r}(m)$, where x and y are subsets of $\{\mathbf{b},\mathbf{r},\mathbf{u}\}$ and specify the possible colours for the poles s and t, respectively. For example, if $x = \{\mathbf{b}\}$ and $y = \{\mathbf{b},\mathbf{r},\mathbf{u}\}$ then the s-pole has to be blue, whereas we do not impose any constraint on the t-pole – it might be blue, red, or uncoloured.

The recurrences then follow very much the decomposition that we had in Section 10.1, but we have to control the possible colours of the poles. Another difficulty is that in the recursive decomposition we might or might not have a blue cut vertex in the coloured network without the root edge. However, we can handle this with the help of appropriately chosen counting functions. Define functions ${}^{y}_{x}n'_{b,r}(m)$, ${}^{y}_{x}s'_{b,r}(m)$, ${}^{y}_{x}p'_{b,r}(m)$, and ${}^{y}_{x}h'_{b,r}(m)$ to count the number of coloured networks, s-, p-, or h-networks with m edges, b blue edges and r red vertices that have no blue cutvertex, respectively. It is clear that coloured networks with blue poles can have no blue cutvertex. Otherwise the blue cutvertex and a blue pole form a 2-cut of a network. Thus ${}^{\{\mathbf{b}\}}_{\{\mathbf{b}\}}n_{b,r}(m) = {}^{\{\mathbf{b}\}}_{\{\mathbf{b}\}}n'_{b,r}(m)$. Moreover ${}^{y}_{x}p_{b,r}(m) = {}^{y}_{x}p'_{b,r}(m)$, ${}^{y}_{x}h_{b,r}(m) = {}^{y}_{x}h'_{b,r}(m)$ since coloured p-, or h-networks do not have any blue (in deed any) cutvertex. For convenience let ${}^{y}_{x}ph_{b,r}(m) = {}^{y}_{x}p_{b,r}(m) + {}^{y}_{x}h_{b,r}(m)$.

Coloured s-networks. A coloured s-network S has either blue poles, one blue and one non-blue poles, or non-blue poles. Let u be the cutvertex in S which is closest to s. If at least one of the poles s,t is blue S can not have any blue cutvertex (in particular u is not blue). Thus ${}^{\{\mathbf{b}\}}_{\{\mathbf{bru}\}}s_{b,r}(m) = {}^{\{\mathbf{b}\}}_{\{\mathbf{bru}\}}s'_{b,r}(m)$ and ${}^{\{\mathbf{ru}\}}_{\{\mathbf{b}\}}s_{b,r}(m) = {}^{\{\mathbf{ru}\}}_{\{\mathbf{b}\}}s'_{b,r}(m)$. The cutvertex u induces a split coloured p- or h-network with poles s,u (counted by ${}^{\{\mathbf{ru}\}}_{\{\mathbf{bru}\}}ph_{b,r}$), and a remaining part with poles u,t, which is an arbitrary coloured network that has no blue cut-vertex (counted by ${}^{\{\mathbf{bru}\}}_{\{\mathbf{ru}\}}n'_{b,r}$). If S has non-blue poles u might be blue. If u is blue the remaining coloured network has no blue cutvertex. If u is not blue the remaining coloured network might have blue cutvertex.

$$
{}^{\{\mathbf{bru}\}}_{\{\mathbf{bru}\}}s'_{b,r}(m) = \sum_{j,b',r'} {}^{\{\mathbf{ru}\}}_{\{\mathbf{bru}\}}ph_{b',r'}(j)\ {}^{\{\mathbf{bru}\}}_{\{\mathbf{ru}\}}n'_{b-b',r-r'}(m-j),
$$

$$
\begin{aligned}
{}^{\{\mathbf{ru}\}}_{\{\mathbf{ru}\}}s_{b,r}(m) &= \sum_{j,b',r'} {}^{\{\mathbf{b}\}}_{\{\mathbf{ru}\}}ph_{b',r'}(j)\ {}^{\{\mathbf{ru}\}}_{\{\mathbf{b}\}}n'_{b-b',r-r'}(m-j) \\
&+ \sum_{j,b',r'} {}^{\{\mathbf{ru}\}}_{\{\mathbf{ru}\}}ph_{b',r'}(j)\ {}^{\{\mathbf{ru}\}}_{\{\mathbf{ru}\}}n_{b-b',r-r'}(m-j)\,.
\end{aligned}
$$

Coloured p-networks. Due to property $(\mathbf{P1}-\mathbf{P2})$ all the coloured vertices of a coloured p-network must lie in one of its parts, whereas the rest part must be a single edge. If at least one of the poles is blue the coloured part has no

blue cutvertex. If not the coloured part might have blue cutvertex.

$$\begin{aligned}
{}^{\{\mathbf{b}\}}_{\{\mathbf{bru}\}}p_{b,r}(m) &= {}^{\{\mathbf{b}\}}_{\{\mathbf{bru}\}}n'_{b,r}(m-1)\\
{}^{\{\mathbf{ru}\}}_{\{\mathbf{b}\}}p_{b,r}(m) &= {}^{\{\mathbf{ru}\}}_{\{\mathbf{b}\}}n'_{b,r}(m-1)\\
{}^{\{\mathbf{ru}\}}_{\{\mathbf{ru}\}}p_{b,r}(m) &= {}^{\{\mathbf{ru}\}}_{\{\mathbf{ru}\}}n_{b,r}(m-1)\,.
\end{aligned}$$

Coloured h-networks. There is a unique embedding of the core of a coloured h-network H into the plane where the root edge and the core edges replaced by coloured networks lie on the outer face. We can again find a recurrence if we use the function ${}^{y}_{x}h_{b,r,l}(m)$ counting the number of coloured h-networks where we additionally require that the l closest core edges (on the outer face) to vertex s are also edges in H.

We look at the l+1-st closest edge uv on the outer face of the core. It might also be an edge in H. If not, uv is a split pair in H and determines a subnetwork S. Due to property (**P3**) it is not possible that both u, v are blue. If either u or v is blue $\{u, v\}$ induces a coloured network with no blue cutvertex. If neither u nor v is blue $\{u, v\}$ induces a coloured network that might have blue cutvertex. It might be the case that all coloured vertices lie in S. Then the remaining network after the replacement of S is 3-connected with at least l+1 vertices on the outer face. Let $c_l(m)$ denote the number of such c-nets. These numbers are computed in [28].

$$\begin{aligned}
&{}^{\{\mathbf{bru}\}}_{\{\mathbf{bru}\}}h_{b,r,l}(m)\\
&= \sum_j ({}^{\{\mathbf{ru}\}}_{\{\mathbf{b}\}}n'_{b,r}(j) + {}^{\{\mathbf{b}\}}_{\{\mathbf{ru}\}}n'_{b,r}(j) + {}^{\{\mathbf{ru}\}}_{\{\mathbf{ru}\}}n_{b,r}(j)) \sum_{k\geq l+1} c_k(m-j+1)\\
&\quad + \sum_{j,b',r'} ({}^{\{\mathbf{ru}\}}_{\{\mathbf{b}\}}n'_{b',r'}(j) + {}^{\{\mathbf{b}\}}_{\{\mathbf{ru}\}}n'_{b',r'}(j)\\
&\quad + {}^{\{\mathbf{ru}\}}_{\{\mathbf{ru}\}}n_{b',r'}(j)) {}^{\{\mathbf{bru}\}}_{\{\mathbf{bru}\}}h_{b-b',r-r',l+1}(m-j+1)\,.
\end{aligned}$$

10.4 Pole-symmetric networks

We saw in Section 10.1 that in a symmetric h-network with a sense-reversing automorphism φ a red edge uv of the core (i.e., $\varphi(u) = v$ and $\varphi(v) = u$) can only be replaced by a *pole-symmetric* subnetwork, that is, a subnetwork with an automorphism ψ that exchanges s and t. Such networks are further decomposed in this section.

Let $\tilde{n}(m)$ be the number of pole-symmetric networks with m edges. According to Theorem 7.1.1 we have $\tilde{n}(m) = \tilde{s}(m) + \tilde{p}(m) + \tilde{h}(m)$, where the functions $\tilde{s}(m), \tilde{p}(m), \tilde{h}(m)$ count the number of pole-symmetric s-, p-, and h-networks with m edges, respectively.

Pole-symmetric s-networks. Here we split off the same p- or h-network at both poles simultaneously. The remaining is either again a pole-symmetric network, or an edge, or a vertex.

$$\tilde{s}(m) = \sum_j (p(j) + h(j))\tilde{n}(m-2j) + 2p(\lfloor\frac{m}{2}\rfloor) + 2h(\lfloor\frac{m}{2}\rfloor)\,.$$

Pole-symmetric p-networks. Given a pole-symmetric p-network, s and t might or might not be adjacent, and there might be several pole-symmetric s- or h-networks between s and t. Similar to the enumeration of p-networks in Section 10.1 we use the function $\tilde{p}_l(m)$ to count the number of pole symmetric p-networks where the number of edges of the largest subnetwork is bounded by l.

$$\begin{aligned}\tilde{p}(m) &= \tilde{p}_m(m)\\ \tilde{p}_l(m) &= \sum_{k=0}^{\lfloor m/l\rfloor}\binom{\tilde{s}(l)+\tilde{h}(l)+k-1}{k}\tilde{p}_{l-1}(m-kl)\,.\end{aligned}$$

Pole-symmetric h-networks. Here we want to control the number of pole-symmetric h-networks with and without a sense-reversing automorphism φ satisfying $\varphi(s)=s$ and $\varphi(t)=t$. In the case where we do not have a sense-reversing automorphism, we order the edges of the core of H in such a way that blue edges uv where $\psi(u)=u$ and $\psi(v)=v$ come first, followed by the red edges uv where $\psi(u)=v$ and $\psi(v)=u$. Finally we have the uncoloured edges, ordered in such a way that corresponding uncoloured edges with respect to the pole-symmetry are consecutive – but we do not care about their order.

In the case that we have a sense-reversing automorphism φ, we order the edges of the core in such a way that we start with the blue edges with respect to φ, and then the blue edges with respect to ψ. Next we list the red edges with respect to φ and then the red edges with respect to ψ. Finally we list corresponding edges with respect to φ consecutively, which are followed by the two corresponding edges with respect to ψ, respectively.

Let $\tilde{b}_{b,\tilde{b},r,\tilde{r},l}(m)$ be the number of pole symmetric networks $\tilde{B}$ with a sense-reversing automorphism φ with m edges, where core$(\tilde{B})$ has $b,\tilde{b}$ blue and $r,\tilde{r}$ red edges with respect to φ and ψ, respectively, and the first l edges of core$(\tilde{B})$ are also edges of $\tilde{B}$. Analogously we introduce $\tilde{a}_{\tilde{b},\tilde{r},l}(m)$ for the number of pole-symmetric networks $\tilde{A}$ without a sense-reversing automorphism. Similarly as in Section 10.1 it is now possible to formulate recurrences for these functions.

To produce $\tilde{B}$, we take a network with both a sense-reversing automorphism φ and a pole-exchanging automorphism ψ. The $l{+}1$-st edge of core$(\tilde{B})$ might be an edge of $\tilde{B}$. If not, we could replace corresponding blue (with respect to φ and ψ) $l{+}1$ and $l{+}2$ edges by an arbitrary network (for $l+1\le b+\tilde{b}$), or we could replace corresponding red (with respect to φ and ψ) $l{+}1$ and $l{+}2$ edges by a network with a pole-exchanging automorphism (for $b{+}\tilde{b}<l{+}1\le b{+}\tilde{b}{+}r{+}\tilde{r}$). Finally we could replace corresponding uncoloured edges $l{+}1$ till $l{+}4$ by the same network (for $l+1>b+\tilde{b}+r+\tilde{r}$). The initial case is that all edges of core$(\tilde{B})$ are also edges of $\tilde{B}$, where we have a pole-symmetric 3-connected network, i.e., a pole-symmetric c-net, with a sense-reversing automorphism - see Section 10.5.

$$\tilde{b}_{b,\tilde{b},r,\tilde{r},l}(m) = \begin{cases}\sum_j n(j)b_{b,\tilde{b},r,\tilde{r},l+2}(m{-}2j)\,, & l{+}1\le b{+}\tilde{b}\\ \sum_j \tilde{n}(j)\tilde{b}_{b,\tilde{b},r,\tilde{r},l+2}(m{-}2j)\,, & b{+}\tilde{b}<l{+}1\le b{+}\tilde{b}{+}r{+}\tilde{r}\\ \sum_j n(j)b_{b,\tilde{b},r,\tilde{r},l+4}(m{-}4j)\,, & l{+}1>b{+}\tilde{b}{+}r{+}\tilde{r}\,.\end{cases}$$

To produce $\tilde{A}$, we could take either a pole-symmetric network with no sense-reversing automorphism or a network with both a pole-exchanging and a sense-

reversing automorphism. In the first case we might replace the l+1st edge of core($\tilde{A}$) by a network. In the second case, we could take a network with both a pole-exchanging automorphism ψ and a sense-reversing automorphism φ, whose core has additional b blue and r red edges with respect to φ. Then we might replace two corresponding blue (with respect to ψ) l+1st and l+2-nd edges by two different network (for $b < l+1 \leq b+\tilde{b}$), or two corresponding red edges with respect to φ by a network without a pole-symmetry (for $b+\tilde{b} < l+1 \leq b+\tilde{b}+r$), and then red edges with respect to ψ by two different pole-symmetric networks (for $b+\tilde{b}+r < l+1 \leq b+\tilde{b}+r+\tilde{r}$): If there is no red nor blue edge left (for $l+1 > b+\tilde{b}+r+\tilde{r}$), we could replace the l+1st and l+3-rd core edges with a different network than the l+2-nd and l+4-th core edge.

The initial case is that all edges of core($\tilde{A}$) are also edges of $\tilde{A}$, which is counted by pole-symmetric c-nets with no sense-reversing automorphism. Such graphs can be counted by subtracting the number of pole-symmetric c-nets with a sense-reversing automorphism from the number of all pole-symmetric c-nets. But to sample asymmetric pole-symmetric c-nets, we need another rejection sampling step. This works analogously as at the end of Section 10.1. To show that we have only a linear number of expected restarts of the rejection sampling procedure, it suffices to show that there are more asymmetric and pole-symmetric c-nets than symmetric pole-symmetric, which can be done.

$$\begin{aligned}
\tilde{a}_{\tilde{b},\tilde{r},l}(m) \;=\; & \sum_{j} n(j)\tilde{a}_{\tilde{b},\tilde{r},l+1}(m-j) \\
& + \sum_{I} n(i)n(j)\tilde{b}_{b,\tilde{b},r,\tilde{r},l+2}(m-i-j) - n(j)\tilde{b}_{b,\tilde{b},r,\tilde{r},l+2}(m-2j) \\
& + \sum_{II} (n(j)-\tilde{n}(j))\tilde{b}_{b,\tilde{b},r,\tilde{r},l+2}(m-2j) \\
& + \sum_{III} (\tilde{n}(i)\tilde{n}(j)\tilde{b}_{b,\tilde{b},r,\tilde{r},l+2}(m-i-j) - \tilde{n}(j)\tilde{b}_{b,\tilde{b},r,\tilde{r},l+2}(m-2j))/2 \\
& \sum_{IV} (n(i)n(j)\tilde{b}_{b,\tilde{b},r,\tilde{r},l+4}(m-2i-2j) - n(j)\tilde{b}_{b,\tilde{b},r,\tilde{r},l+4}(m-4j))/2
\end{aligned}$$

where I stands for the indices b, r, i, j satisfying $l-\tilde{b} < b \leq l, 0 < b+r, i, j \geq 0$, II the indices b, r, j satisfying $0 \leq b \leq l-\tilde{b}, \max\{0, l-\tilde{b}\} < b+r, j \geq 0$, III the indices b, r, i, j satisfying $\max\{0, l-\tilde{b}-\tilde{r}\} < b+r \leq l-\tilde{b}, i, j \geq 0$, and IV the indices b, r, i, j satisfying $0 < b+r \leq l-\tilde{b}-\tilde{r}, i, j \geq 0$.

With these numbers we can compute

$$\tilde{h}(m) = \sum_{b,\tilde{b},r,\tilde{r}\geq 0, 0<b+r,\tilde{b}+\tilde{r}} \tilde{a}_{\tilde{b},\tilde{r},0}(m) + \tilde{b}_{b,\tilde{b},r,\tilde{r},0}(m).$$

10.5 Pole-symmetric coloured networks

To compute the number of pole-symmetric networks with a sense-reversing automorphism, we again use coloured networks, but impose the additional constraint that the coloured network has a pole-exchanging automorphism. Along the lines of Theorem 10.2.2 we have a bijection between these pole-symmetric coloured networks and pole-symmetric networks with a sense-reversing automorphism.

The decomposition of pole-symmetric coloured networks is a straightforward combination of the ideas in Section 10.2 and 10.4.

Pole-symmetric coloured s-networks. Given a pole-symmetric coloured s-network, we split off a p-, or h- graph at both poles simultaneously and count the remaining graphs, which are pole-symmetric coloured s-, p-, or h-graphs.

$$\tilde{s}'_{b,r}(m) \;=\; \sum_j (p'_{b,r}(j) + h'_{b,r}(j))\tilde{n}'_{b,r}(m-2j)\,.$$

Pole-symmetric coloured p-networks. Because we have a coloured network, one of the two components induced by the poles is a single edge.

$$\tilde{p}'_{b,r}(m) \;=\; \tilde{n}'_{b,r}(m-3)\,.$$

Pole-symmetric coloured h-networks. Let $\tilde{h}'_{b,r,l}(m)$ be the number of coloured h-networks such that the l closest core edges (on the outer face) to vertex s are also edges in H.

$$\begin{aligned}\tilde{h}'_{b,r,l}(m) \;&=\; \sum_{j\geq 0} \tilde{p}'_{b,r}(j) \sum_{k\geq l+1} \tilde{t}_k(m-2j) \\ &\quad + \sum_{b',r',i\geq 0} p'_{b',r'}(j) h'_{b-2b',r-2r',l+1}(m-2j)\,.\end{aligned}$$

When we remove the last coloured subnetwork in a pole-symmetric coloured h-network, we have an embedded 3-connected pole-symmetric network with l edges on the outer face. The dual of such an object is an embedded 3-connected network with a sense-reversing automorphism where the s-pole has degree l (Blue edges correspond to red edges and vice versa). It is possible to modify the decomposition of coloured networks in Section 10.2 to control also this parameter. Finally for coloured h-networks we have to use c-nets where the degree of the s-pole is specified; it is also possible, as we will see in the next section.

Due to Theorem 10.2.2 we have

$$\tilde{b}_{r,\tilde{r},m-1}(m) = \sum_b \tilde{n}'_{b,r,\lceil \tilde{r}/2\rceil}((m+b-r)/2).$$

To count pole-symmetric-nets we its dual, i.e., c-nets with sense-reversing automorphism and obtain $\tilde{c}_l(m) = \sum_r \tilde{c}_{l,r}(m)$ and

$$\tilde{c}_{l,r}(m) = \sum_b \sum_{k\geq \lceil l/2\rceil} n'_{b,r,k}((m+b-r)/2).$$

10.6 Uniform sampling

The decomposition together with the counting formulas immediately gives a polynomial time generation procedure for planar networks. Using rejection-sampling we obtain the main result:

Theorem 10.6.1. *There is an algorithm that samples an unlabeled 2-connected planar graph with m edges uniformly at random, in expected time $O(m^9)$. If the algorithm has direct access to the values of appropriately chosen counting formulas, the algorithm can generate such an object in expected cubic time.*

Proof. The algorithm first generates a planar network N with m edges, using the above decomposition and the values of the counting formulas that can be computed efficiently using dynamic programming. Note that the representation size of all the numbers is linear, since we deal with unlabeled structures. We use at most six-dimensional tables (in Section 10.4). The summation there runs over one parameter, and within the sum we have to perform a multiplication with large numbers, which can be done in quadratic time. Hence, the overall running time for the computation of the values is within $O(m^9)$.

Using these values we can make the correct probabilistic decisions in a recursive construction of a planar network according to the presented decomposition – this method is standard and known as the *recursive method* for sampling [47, 64, 108]. Then the algorithm computes the size o of the orbit of the root in the automorphism group of the graph of N, which can be done in linear time using e.g. well-known correspondingly adapted graph isomorphism algorithms for planar graphs [78, 79], and outputs the graph of N with probability $1/o$. Since the number of edges in a planar graph is linear, the expected number of restarts is also linear. Thus the overall expected running time is in $O(m^9)$.

If we do not charge for the costs for computing the values in the table and the partial sums of the formulas, e.g. because we performed a precomputation step, the generation can be done in cubic time. □

The counting formulas resulting from the ideas presented here can easily be extended to count objects where also the number of vertices is specified. It is also easy to allow parallel edges, and count multigraphs. In principle, the decomposition and the recursive formulas have a form that allows to formulate equations between the corresponding *generating functions*. It is sometimes possible to solve these equations and obtain closed formulas or asymptotic estimates from the solutions. However, due to the large number of parameters needed in the decomposition, it will not be easy to handle these equations.

Combining the methods in this chapter and Chapter 6 a uniform sampling algorithm for unlabeled cubic planar graphs was developed in [23].

Chapter 11

Gaussian matrix integral method

A seminal technique of theoretical physics called Wick's theorem interprets the Gauss-ian matrix integral of the product of the traces of powers of Hermitian matrices as the number of labeled maps with a given degree sequence, sorted by their Euler characteristics (these maps are the *Feynman diagrams* for the matrix integral). This approach leads to the formulas for the numbers of labeled maps with a given degree sequence and the genus, which are analogous to the formulas obtained by combinatorial methods [14]. Furthermore, the method has been successfully used in enumeration of combinatorial structures and in other fields of mathematics (see e.g., [19, 40, 41, 42, 50, 51, 77, 91, 92, 104, 105, 111, 115, 116, 149]).

In this chapter we first provide a complete proof of the map enumeration obtained from the Gaussian matrix integral. We then apply the method of the Gaussian matrix integral to other problems. We show that the enumeration of the graphs *embeddable* on a given 2-dimensional surface, in particular of planar graphs with a given degree sequence, can also be formulated by the Gaussian matrix integral.

In Section 11.1 we recall the definition of Gaussian matrix integral and Wick's theorem. In Section 11.2 we discuss the pictorial interpretation of the Gaussian matrix integral. In Section 11.3 we prove that the number of maps with a given degree sequence can be formulated as a Gaussian matrix integral of powers of the traces of Hermitian matrices. In Sections 11.4 and 11.6 we replace the trace by a *flow* defined on a directed graph induced by a Hermitian matrix, and show that the number of planar graphs with a given degree sequence can be formulated as the Gaussian matrix integral.

11.1 Gaussian matrix integral

Let $M = (M_{ij})$ be an $N \times N$ Hermitian matrix, i.e., $M_{ij} = \overline{M}_{ji}$ for every $1 \leq i, j \leq N$, where $\overline{M}_{ji}$ denotes the complex conjugate of M_{ji}. Let $f(M) = \sum_I a_I \prod_{(ij)\in I} M_{ij}$ be a polynomial in its entries, where I ranges over a finite system of multisets of elements of $N \times N$ and a_I's are complex numbers. We start with the following definition (we will see later in this section how Wick's

theorem, Theorem 11.1.1, yields this definition).

Definition 11.1.1. *We let*

$$< f >= \sum_I a_I < \prod_{(ij)\in I} M_{ij} >= \sum_I a_I \sum_P \prod_{(p,q)\in P} < M_p M_q >,$$

where P ranges over all partitions of I into pairs, and for $p = (p_1, p_2), q = (q_1, q_2)$ we have $< M_p M_q >$ is non-zero only if $p_1 = q_2$ and $p_2 = q_1$ and in that case $< M_p M_q >= 1/N$.

Next we recall Gaussian integral and Gaussian matrix integral.

Gaussian integral. We first consider the case $N = 1$. For an arbitrary real function f, the standard Gaussian integral is defined as

$$< f >= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} f(x) dx, \tag{11.1.1}$$

where we abuse notation by a multiple use of the symbol $<>$. Note that $< 1 > = 1$. We are in particular interested in a function of the form $f(x) = x^{2n}$, where n is an integer. In order to compute $< x^{2n} >$, we introduce the so-called *source integral* $< e^{xs} >$ for a given real s. The source integral can be computed as follows.

$$\begin{aligned} < e^{xs} > &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} e^{xs} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-s)^2}{2} + \frac{s^2}{2}} dx \\ &= e^{\frac{s^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-s)^2}{2}} dx \\ &= e^{\frac{s^2}{2}}. \end{aligned} \tag{11.1.2}$$

On the other hand, taking the k-th derivatives of $< e^{xs} >$ with respect to s and taking $s = 0$, we get

$$\begin{aligned} \frac{\partial^k}{\partial s^k} < e^{xs} > \Big|_{s=0} &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \frac{\partial^k}{\partial s^k} e^{xs} \Big|_{s=0} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} x^k dx \\ &= < x^k >, \end{aligned} \tag{11.1.3}$$

where the first equality is due to the Leibniz integral rule. To interpret $< x^k >$, we use (11.1.3) and (11.1.2) to get

$$< x^k > \overset{(11.1.3)}{=} \frac{\partial^k}{\partial s^k} < e^{xs} > \Big|_{s=0} \overset{(11.1.2)}{=} \frac{\partial^k}{\partial s^k} e^{\frac{s^2}{2}} \Big|_{s=0}. \tag{11.1.4}$$

As a consequence, we obtain $< x^k > = 0$ for odd k, and $< x^k >= (k-1)!! = \prod_{i=1}^{k/2}(k-2i+1)$ for even k, which is the same as the number of ways to partition k elements into $k/2$ pairs.

Gaussian matrix integral. Let $M = (M_{ij})$ be an $N \times N$ Hermitian matrix and let

$$dM = \prod_i dM_{ii} \prod_{i<j} d\,\mathrm{Re}(M_{ij}) d\,\mathrm{Im}(M_{ij})$$

denote the standard Haar measure, where $\mathrm{Re}(M_{ij})$ and $\mathrm{Im}(M_{ij})$ denote the real part and the imaginary part of M_{ij}.

The Gaussian (Hermitian) matrix integral of an arbitrary function f is defined as

$$< f > \quad = \quad \frac{1}{Z_0(N)} \int e^{-N\ \mathrm{Tr}(\frac{M^2}{2})} f(M) dM, \tag{11.1.5}$$

where the integration is over $N \times N$ Hermitian matrices, and $Z_0(N)$ is the normalization factor making $< 1 > = 1$, i.e., $Z_0(N) = \int e^{-N\ \mathrm{Tr}(\frac{M^2}{2})} dM$.

As before we are particularly interested in a function of the form

$$f(M) = \sum_I a_I \prod_{(ij)\in I} M_{ij},$$

where I ranges over a finite system of multisets of elements of $N \times N$. We introduce also the source integral $< e^{\mathrm{Tr}(MS)} >$ for a given $N \times N$ Hermitian matrix S, where $\mathrm{Tr}(MS)$ denotes the trace of the matrix MS. It can easily be computed by

$$\begin{aligned} < e^{\mathrm{Tr}(MS)} > \quad &= \quad \frac{1}{Z_0(N)} \int e^{-N\ \mathrm{Tr}(\frac{M^2}{2})} e^{\mathrm{Tr}(MS)} dM \\ &= \quad \frac{1}{Z_0(N)} \int e^{-N\ \mathrm{Tr}(\frac{1}{2}(M-\frac{S}{N})^2)} e^{\frac{\mathrm{Tr}(S^2)}{2N}} dM \\ &= \quad e^{\frac{\mathrm{Tr}(S^2)}{2N}}, \end{aligned} \tag{11.1.6}$$

since the trace is linear and $\mathrm{Tr}(MS) = \mathrm{Tr}(SM)$, and thus we get

$$\begin{aligned} -N\ \mathrm{Tr}\left(\frac{M^2}{2}\right) + \mathrm{Tr}(MS) \quad &= \quad -N\ \mathrm{Tr}\left(\frac{M^2}{2} - \frac{MS+SM}{2N}\right) \\ &= \quad -N\ \mathrm{Tr}\left(\frac{1}{2}\left(M - \frac{S}{N}\right)^2\right) + \frac{\mathrm{Tr}\left(S^2\right)}{2N}. \end{aligned}$$

On the other hands, for any $1 \le i, j \le N$ we get

$$\begin{aligned} \frac{\partial}{\partial S_{ji}} e^{\mathrm{Tr}(MS)}\Big|_{S=0} \quad &= \quad \left(\frac{\partial}{\partial S_{ji}} \mathrm{Tr}(MS)\right) e^{\mathrm{Tr}(MS)}\Big|_{S=0} \\ &= \quad \left(\frac{\partial}{\partial S_{ji}} \sum_{m,n} M_{mn} S_{nm}\right) e^{\mathrm{Tr}(MS)}\Big|_{S=0} \\ &= \quad M_{ij}. \end{aligned}$$

Thus the derivatives of the source integral becomes

$$\begin{aligned} & \frac{\partial}{\partial S_{ji}}\frac{\partial}{\partial S_{lk}}\cdots < e^{\mathrm{Tr}(MS)} > \Big|_{S=0} \\ = & \frac{1}{Z_0(N)}\int e^{-N\ \mathrm{Tr}(\frac{M^2}{2})}\frac{\partial}{\partial S_{ji}}\frac{\partial}{\partial S_{lk}}\cdots e^{\mathrm{Tr}(MS)}\Big|_{S=0} dM \\ = & \frac{1}{Z_0(N)}\int e^{-N\ \mathrm{Tr}(\frac{M^2}{2})} M_{ij}M_{kl}\cdots dM \\ = & < M_{ij}M_{kl}\cdots >, \end{aligned} \tag{11.1.7}$$

where the first equality is due to the Leibniz integral rule. Using (11.1.7) and (11.1.6), we obtain

$$\begin{aligned} < M_{ij}M_{kl}\cdots > & \overset{(11.1.7)}{=} \frac{\partial}{\partial S_{ji}}\frac{\partial}{\partial S_{lk}}\cdots < e^{\mathrm{Tr}(MS)} > \Big|_{S=0} \\ & \overset{(11.1.6)}{=} \frac{\partial}{\partial S_{ji}}\frac{\partial}{\partial S_{lk}}\cdots e^{\frac{\mathrm{Tr}(S^2)}{2N}}\Big|_{S=0} \end{aligned} \tag{11.1.8}$$

and in particular

$$\begin{aligned} < M_{ij}M_{kl} > &= \frac{\partial}{\partial S_{ji}}\frac{\partial}{\partial S_{lk}} e^{\frac{\mathrm{Tr}(S^2)}{2N}}\Big|_{S=0} \\ &= \frac{\partial}{\partial S_{ji}}\left(\frac{\partial}{\partial S_{lk}}\frac{\mathrm{Tr}(S^2)}{2N}\right) e^{\frac{\mathrm{Tr}(S^2)}{2N}}\Big|_{S=0} \\ &= \frac{\partial}{\partial S_{ji}}\left(\frac{\partial}{\partial S_{lk}}\frac{\sum_{m,n} S_{mn}S_{nm}}{2N}\right) e^{\frac{\mathrm{Tr}(S^2)}{2N}}\Big|_{S=0} \\ &= \frac{\partial}{\partial S_{ji}}\frac{S_{kl}}{N} e^{\frac{\mathrm{Tr}(S^2)}{2N}}\Big|_{S=0} \\ &= \frac{\delta_{il}\delta_{jk}}{N}. \end{aligned} \tag{11.1.9}$$

Further, the derivatives in (11.1.8) and (11.1.9) must be taken in pairs (e.g., S_{ji} and S_{lk} with $l = i$ and $k = j$) to get a non-zero contribution. This yields the following result known as *Wick's theorem* or the matrix Wick Theorem.

Theorem 11.1.1 (Wick's theorem)**.** *Let $M = (M_{ij}) \in \mathbb{C}^{N\times N}$ be Hermitian and I be a finite multiset of index pairs. Then*

$$\begin{aligned} < \prod_{(ij)\in I} M_{ij} > &= \sum_{pairings, P\subset I^2} \prod_{((ij),(kl))\in P} < M_{ij}M_{kl} > \\ &= \sum_{pairings, P\subset I^2} \prod_{((ij),(kl))\in P} \frac{\delta_{il}\delta_{jk}}{N}. \end{aligned}$$

11.2 Pictorial interpretation

In this section we will show how to interpret the Gaussian matrix integral using concepts of graph theory.

Fat graphs. A *map* is a graph together with a fixed cyclic ordering of the incident edges of each vertex. It defines an embedding of the graph on an orientable 2-dimensional surface (see [101]). A map is also called a *fat graph*. We prefer this term since it corresponds to a helpful pictorial representation. In a fat graph F the vertices are made into discs (islands) and connected by fattened edges (bridges) prescribed by the cyclic orders. This defines a two-dimensional orientable surface with boundary which we also denote by F. Each component of the boundary of F will be called a *face* of F. Each face is an embedded circle (see e.g., [101]). We will denote by $G(F)$ the underlying graph of F. We denote by $e(F), v(F), p(F), c(F), g(F)$ the number of edges, vertices, faces, connected components, and the genus of F. We recall that $2g(F) = 2c(F) + e(F) - v(F) - p(F)$.

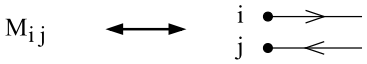
In the next sections we will count fat graphs and their relatives. To avoid confusion we assume that a fat graph has *labeled* vertices, i.e., two fat graphs are equal if they are equal as sets. We speak about *unlabeled* fat graphs if the equality is up to isomorphism.

Definition 11.2.1. *A fat graph is* pointed *if for each vertex, one fat edge incident to it is specified as its initial fat edge.*

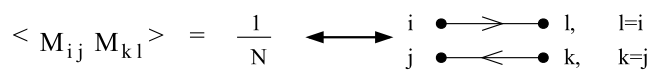
Observation 11.2.2. *Let F be a pointed fat graph. Then there is a unique orientation of the faces of F defined in each component as follows: let v be a vertex and let e be its incident fat edge. Orient the first (clockwise) shore of e from v, and the second shore of e to v.*

Proof. We need to observe that the described process consistently orients each face of F, and that is simple. □

Pictorial interpretation. Here we will interprete the non-zero contributions to $< f >$, where $f(M) = \sum_I a_I \prod_{(ij)\in I} M_{ij}$, pictorially as follows. We represent M_{ij} as a half- fat edge consisting of two end points and two parallel lines with opposite orientation such that i is associated with the outgoing line and j the incoming line:



Further (11.1.9) saying $< M_{ij}M_{kl} >= \delta_{il}\delta_{jk}/N$ can be interpreted as that two half- fat edges M_{ij} and M_{kl} construct a fat edge with oppositely oriented shores and with weight $1/N$ if and only if $i = l$ and $j = k$:



A fat edge with oppositely oriented shores will be called a *decorated fat edge*.

Example 11.1. Let us consider $\mathrm{Tr}(M^n)$, the trace of the nth power of the matrix M. By definition of the trace we get

$$\mathrm{Tr}(M^n) \;=\; \sum_{1\le i_1,i_2,\cdots,i_n\le N} M_{i_1 i_2} M_{i_2 i_3} \cdots M_{i_n i_1}.$$

Following the above pictorial interpretation we represent $\mathrm{Tr}(M^n)$ as a star fat diagram with n decorated half- fat edges arranged in a clockwise *pointed* order, such that for each half- fat edge, its first shore (clockwise along the centre) is oriented from the centre, as in Figure 11.1.

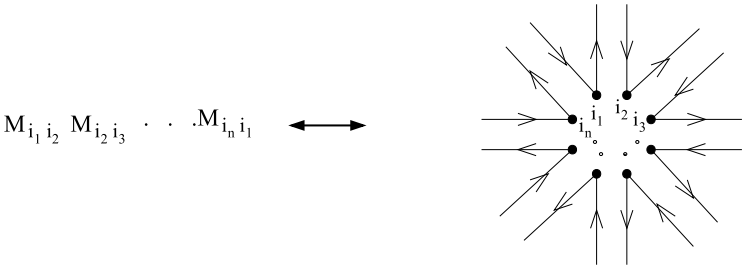


Figure 11.1: $\mathrm{Tr}(M^n)$ and its pictorial interpretation as a star fat diagram.

Moreover, using Wick's Theorem we can compute

$$\begin{aligned} < \mathrm{Tr}(M^n) > \;&=\; < \sum_{1\le i_1,i_2,\cdots,i_n\le N} M_{i_1 i_2} M_{i_2 i_3} \cdots M_{i_n i_1} > \\ &=\; \sum_{1\le i_1,i_2,\cdots,i_n\le N} \prod_{\text{pairing}} < M_{i_k i_{k+1}} M_{i_l i_{l+1}} > \\ &=\; \sum_{1\le i_1,i_2,\cdots,i_n\le N} \prod_{\text{pairing}} \frac{\delta_{i_k i_{k+1}} \delta_{i_l i_{l+1}}}{N}. \end{aligned} \tag{11.2.1}$$

Note that n should be even in order to get a non-zero contribution to (11.2.1) and thus set $n = 2m$. Further, observe that only a *fraction* of the $(2m-1)!!$ possible pairings have non-zero contribution to (11.2.1); equivalently such pairings form decorated fat edges. In other words, we can think of a pairing with non-zero contribution to (11.2.1) as a pointed fat graph with one island, whose faces are oriented as in Observation 11.2.2. It indeed defines uniquely an embedding on a surface (see Figure 11.2).

Let F be a contributing pointed fat graph. Certainly it has $n/2 = m$ edges. Since each edge contributes $1/N$ to (11.2.1), each pairing gets $1/N^m$ from all its edges. However, we should count the contributions due to the summations over $1 \le i_1, i_2, \cdots, i_n \le N$. Notice that each (oriented) face attains independently any index from 1 to N and thus the faces contribute $N^{p(F)}$ to (11.2.1). In summary each pointed fat graph F with one island and m faces contribute $N^{p(F)-m}$. Thus pointed fat graphs with genus zero contribute the leading term in N as N tends to ∞. Such a pointed fat graph with genus zero is called a *planar fat graph* (or a *plane graph* or *plane map* among graph theorists). The first two fat graphs in Figure 11.2 are planar fat graphs.

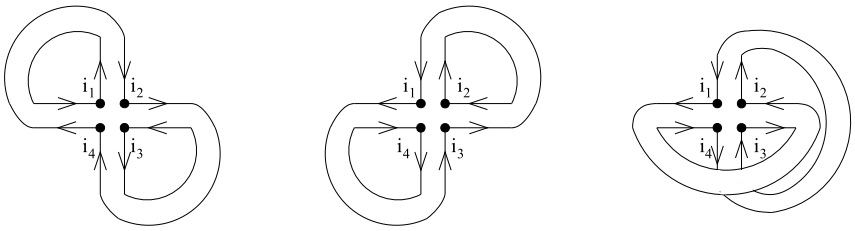


Figure 11.2: All possible fat graphs with one island and $n = 4$.

In [51] the term *a planar graph* was used to mean a planar fat graph or a plane graph. However we will distinguish these, as this is more common among graph theorists: a planar graph is a graph *embeddable* on a sphere without crossing edges, whereas a plane graph is a graph *embedded* on a sphere without crossing edges.

Now we count the planar fat graphs interpreted from $< \mathrm{Tr}(M^n) >$. Let C_m denote the set and c_m the number of the planar pointed fat graphs with one vertex and with $m = n/2$ fat edges. Note that the half- fat edge $M_{i_1 i_2}$ would be paired by one of the half- fat edges $M_{i_2 i_3}, M_{i_4 i_5}, \cdots, M_{i_{2m} i_1}$, say $M_{i_{2k} i_{2k+1}}$, forming a fat edge. We can interpret each element of C_m as being composed of an element of C_{k-1} and an element of C_{m-k}.

It implies the following recursion

$$c_m \;=\; \sum_{k=1}^{m} c_{k-1} c_{m-k}, \quad c_0 = c_1 = 1,$$

which yields a well-known *Catalan number*

$$c_m \;=\; \frac{1}{m+1}\binom{2m}{m} \quad \text{for } m \geq 1.$$

As a consequence we can compute the limit of Gaussian matrix integral of $\mathrm{Tr}(M^n)$:

$$\lim_{N\to\infty} \frac{< \mathrm{Tr}(M^n) >}{N} \;=\; \begin{cases} c_m & \text{if } n = 2m \\ 0 & \text{otherwise .} \end{cases}$$

Example 11.2. Our next example is $\mathrm{Tr}(M^3)^4\,\mathrm{Tr}(M^2)^3$. As before we rewrite it as

$$\begin{aligned} & < \mathrm{Tr}(M^3)^4\,\mathrm{Tr}(M^2)^3 > \\ = \quad & < \left(\sum_{1\leq i_1,i_2,i_3\leq N} M_{i_1 i_2} M_{i_2 i_3} M_{i_3 i_1} \right)^4 \left(\sum_{1\leq j_1,j_2\leq N} M_{j_1 j_2} M_{j_2 j_1} \right)^3 > . \end{aligned}$$

Analogously to the previous example, this equals $\sum_F N^{p(F)-e(F)}$, where the sum is over all pointed fat graphs F consisting of four fat vertices of degree 3, and three fat vertices of degree 2 (see Figure 11.3).

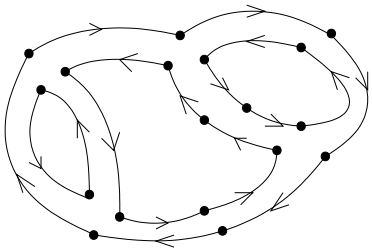


Figure 11.3: A planar fat graph interpreted from $< \mathrm{Tr}(M^3)^4\,\mathrm{Tr}(M^2)^3 >$.

In summary, if $f(M)$ is a product of the traces of powers of M then $< f >$ is equal to the number of pointed fat graphs F with the degree sequence given by the powers, and weighted by $N^{p(F)-e(F)}$.

11.3 Planar fat graphs

In this section we derive the formula for the number of planar fat graphs with a given degree sequence using Wick's theorem. For this we first recall some necessary concepts from complex analysis [126]. For every complex number z the exponential function $\exp(z)$, or e^z for short, is defined by the formula

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!} = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \frac{z^4}{4!} + \cdots .$$

Fix a region (i.e., a nonempty connected open subset of complex plane) Ω such that $\exp(\Omega) = B(1,1)$, where $B(c,r) = \{z : |z-c| < r\}$ for c complex and $r > 0$ real. The natural logarithm function $\log(z)$ may be defined as its inverse function, that is, it satisfies $w = \log(z)$ and $e^w = z$ for $z \in B(1,1)$. It is holomorphic, i.e., infinitely often complex differentiable in $B(1,1) \setminus \{z \in \mathbb{R} : z \leq 1\}$ and thus it can be described by its Taylor series

$$\begin{aligned} \log(z) &= \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}(z-1)^n \\ &= (z-1) - \frac{(z-1)^2}{2} + \frac{(z-1)^3}{3} - \frac{(z-1)^4}{4} \cdots . \end{aligned}$$

The definition of the exponential function given above can be extended for every Banach algebra. For each $i \geq 1$ let z_i be a non-zero complex variable with $|z_i| \in (0, \varepsilon_i)$ where $\varepsilon_i > 0$ will be chosen later. Let I be the group consisting of all non-negative integer vectors with only a finite number of non-zero entries $a = (n_1, n_2, \cdots)$, $n_i \in \mathbb{N}$ equipped with a linear order, which then forms a linearly ordered Abelian group. The ring $\mathbb{R}[[Z]]$ of formal power series over $\mathbb{R}$ in variable $Z = (z_1, z_2, \cdots, z_i, \cdots)$, $i \in \mathbb{N}$ with a linearly ordered Abelian group I as the index set is a Banach algebra. Thus we can extend the definition of the exponential function for the ring $\mathbb{R}[[Z]]$. In particular, for $X \in \mathbb{R}[[Z]]$ with

$X = \sum_{i\geq 1} c_i z_i$, $c_i \in \mathbb{R}$,

$$e^X = \sum_{n=0}^{\infty} \frac{X^n}{n!} = \sum_{\substack{a \in I \\ a=(n_1,\cdots,n_k)}} \prod_{i\leq k} \frac{(c_i z_i)^{n_i}}{n_i!} \tag{11.3.1}$$

$$= \lim_{\bar{a}\to\infty} \sum_{\substack{a \leq \bar{a} \\ a=(n_1,\cdots,n_k)}} \prod_{i\leq k} \frac{(c_i z_i)^{n_i}}{n_i!}, \tag{11.3.2}$$

where instead of $a = (n_1, n_2, \cdots)$ we write $a = (n_1, \cdots, n_k)$ for short if $n_i = 0$ for $i \geq k+1$.

Now we consider a function f which maps each Hermitian matrix M to

$$f(M) = e^{N\sum_{i\geq 1} z_i \mathrm{Tr}\left(\frac{M^i}{i}\right)},$$

and investigate the relation between the Gaussian matrix integral of f and the asymptotic number of planar fat graphs.

Observe first that for each $i \geq 1$, $\mathrm{Tr}(M^i) = \sum_{j=1}^N \lambda_j(M)^i$ where $\lambda_j(M)$'s are the eigenvalues of M, which are all real since M is a Hermitian matrix, and hence $\mathrm{Tr}(\frac{M^i}{i}) = \sum_{j=1}^N \frac{\lambda_j(M)^i}{i} \in \mathbb{R}$. Thus $N\sum_{i\geq 1} z_i \mathrm{Tr}\left(\frac{M^i}{i}\right) \in \mathbb{R}[[Z]]$ and we can use (11.3.2) and get

$$f(M) = \lim_{\bar{a}\to\infty} \sum_{\substack{a \leq \bar{a} \\ a=(n_1,\cdots,n_k)}} \prod_{i\leq k} \frac{(N\ z_i)^{n_i}}{n_i!} \left[\mathrm{Tr}\left(\frac{M^i}{i}\right)\right]^{n_i}$$

$$= \lim_{\bar{a}\to\infty} \sum_{\substack{a \leq \bar{a} \\ a=(n_1,\cdots,n_k)}} \prod_{i\leq k} \frac{(N\ z_i)^{n_i}}{i^{n_i} n_i!} \left[\mathrm{Tr}(M^i)\right]^{n_i}.$$

We let

$$\eta_{\bar{a}}(M) = \sum_{\substack{a \leq \bar{a} \\ a=(n_1,\cdots,n_k)}} \prod_{i\leq k} \frac{(N\ z_i)^{n_i}}{i^{n_i} n_i!} \left[\mathrm{Tr}(M^i)\right]^{n_i}$$

$$\eta(M) = \lim_{\bar{a}\to\infty} \sum_{\substack{a \leq \bar{a} \\ a=(n_1,\cdots,n_k)}} \prod_{i\leq k} \frac{(N\ z_i)^{n_i}}{i^{n_i} n_i!} \left[\mathrm{Tr}(M^i)\right]^{n_i}.$$

We first claim the following:

Claim 11.3.1. *For each z_i with $|z_i| \in (0, \varepsilon_i)$, we have*

$$< e^{N\sum_{i\geq 1} z_i \mathrm{Tr}\left(\frac{M^i}{i}\right)} > = \lim_{\bar{a}\to\infty} < \sum_{\substack{a \leq \bar{a} \\ a=(n_1,\cdots,n_k)}} \prod_{i\leq k} \frac{(N\ z_i)^{n_i}}{i^{n_i} n_i!} \left[\mathrm{Tr}(M^i)\right]^{n_i} > .$$

That is,

$$< \eta(M) > = \lim_{\bar{a}\to\infty} < \eta_{\bar{a}}(M) > .$$

To prove it we use the following lemma, which is reformulated from Theorem 4 and Lemma 2 in Section 2, Chap. II in [131]:

Lemma 11.3.2. *Let $\{\eta_n\}_{n\geq 1}$ be a sequence of random variables. If $\{\eta_n\}_{n\geq 1}$ satisfies*

(i) $\mathbb{E}(|\eta_n|)$ *is uniformly bounded, i.e.,* $\sup_n \mathbb{E}(|\eta_n|) < \infty$,

(ii) $\mathbb{E}(|\eta_n|I_A)$ *is uniformly absolutely continuous, i.e., for any* $\varepsilon > 0$ *there exists* $\delta = \delta(\varepsilon) > 0$ *such that if* $\mathbb{E}(I_A) < \delta$, *then* $\sup_n \mathbb{E}(|\eta_n|I_A) < \varepsilon$, *and*

(iii) $\lim_{n\to\infty} \eta_n = \eta$,

then

$$\mathbb{E}(\eta) = \lim_{n\to\infty} \mathbb{E}(\eta_n).$$

We set $\mathbb{E}(F) = < F(M) >$ for any integrable function F. Let $\lambda_i = \lambda_i(M)$, $1 \leq i \leq N$, be the eigenvalues of M and let $\lambda = \lambda(M) = \max_{1\leq i\leq N} |\lambda_j(M)|$. Then the following holds.

Lemma 11.3.3. *For any non-negative integer k, $\mathbb{E}(\lambda^k) < \infty$.*

Proof of Lemma 11.3.3. We will show that for fixed $1 \leq i \leq N$ the function λ_i^k is measurable (with respect to the Haar measure dM), and so $\mathbb{E}(\lambda_i^k) < \infty$. It follows then that for fixed $1 \leq i \leq N$ the absolute function $|\lambda_i|^k$ and the maximum function $\lambda^k = \max_{1\leq i\leq N} |\lambda_j|^k$ are also measurable, and so $\mathbb{E}(\lambda^k) < \infty$ (see e.g.,[126]).

Let $U = U(M)$ be the unitary matrix that diagonalize M,

$$M = U\Lambda U^*$$

where U^* is the conjugate transpose of U, i.e., $(U^*)_{ij} = \overline{U}_{ji}$ and $\Lambda = \Lambda(M)$ is a diagonal matrix defined by $\Lambda_{ij} = \lambda_i \delta_{ij}$. Since $\mathrm{Tr}(M^2) = \mathrm{Tr}(\Lambda^2)$ and the Jacobian of the transformation $M \mapsto M = U\Lambda U^*$ is

$$J = \prod_{1\leq i\leq j\leq N} (\lambda_i - \lambda_j)^2,$$

we get $dM = Jd\lambda_1 \cdots d\lambda_N$. Further we obtain

$$\begin{aligned}
Z_0(N) &= \int e^{-N \,\mathrm{Tr}(\frac{M^2}{2})} dM \\
&= \int e^{-N \,\mathrm{Tr}(\frac{1}{2}\Lambda^2)} Jd\lambda_1 \cdots d\lambda_N, \\
< e^{\mathrm{Tr}(\Lambda S)} > &= \frac{1}{Z_0(N)} \int e^{-N \,\mathrm{Tr}(\frac{M^2}{2})} e^{\mathrm{Tr}(\Lambda S)} dM \\
&= e^{\frac{\mathrm{Tr}(S^2)}{2N}} \frac{1}{Z_0(N)} \int e^{-N \,\mathrm{Tr}(\frac{1}{2}(\Lambda - \frac{S}{N})^2)} Jd\lambda_1 \cdots d\lambda_N \\
&= e^{\frac{\mathrm{Tr}(S^2)}{2N}}.
\end{aligned}$$

Thus, similarly to (11.1.7) and (11.1.8) we get

$$\begin{aligned}
< \Lambda_{ii}^k > = \frac{\partial^k}{\partial S_{ii}^k} < e^{\mathrm{Tr}(\Lambda S)} > \Big|_{S=0} &= \frac{\partial^k}{\partial S_{ii}^k} e^{\frac{\mathrm{Tr}(S^2)}{2N}} \Big|_{S=0} \\
&= \begin{cases} 0 & \text{if } k \text{ is odd} \\ \frac{(k-1)!!}{N^{k/2}} & \text{if } k \text{ is even.} \end{cases}
\end{aligned}$$

Thus $\mathbb{E}(\lambda_i^k) = <\lambda_i^k> = <\Lambda_{ii}^k> < \infty$.

□

Proof of Claim 11.3.1. We observe first that $[\mathrm{Tr}(M^i)]^{n_i} = [\sum_{j=1}^N \lambda_j(M)^i]^{n_i} \leq (N\lambda(M)^i)^{n_i}$.

Now we will show that $\mathbb{E}(|\eta_{\bar{a}}|)$ is uniformly bounded. We let $\bar{a} = (\bar{n_1}, \cdots, \bar{n_k})$ be given and we select $\varepsilon_i(\bar{a}) > 0$ so that for every z_i with $|z_i| \in (0, \varepsilon_i(\bar{a}))$ we have

$$\xi_{\bar{a}} = \sum_{\substack{a \leq \bar{a} \\ a = (n_1, \cdots, n_k)}} \prod_{i \leq k} \frac{(N^2 |z_i|)^{n_i}}{i^{n_i} n_i!} < \infty .$$

Let

$$\varepsilon_i = \inf_{\bar{a}} \varepsilon_i(\bar{a}) > 0. \tag{11.3.3}$$

From Lemma 11.3.3 we know that $\bar{\lambda} = \mathbb{E}(\lambda(M)^{\sum_{i \leq k} i n_i}) < \infty$. It follows that

$$\begin{aligned}
\mathbb{E}(|\eta_{\bar{a}}|) &= \mathbb{E}\left(\left| \sum_{\substack{a \leq \bar{a} \\ a = (n_1, \cdots, n_k)}} \prod_{i \leq k} \frac{(N z_i)^{n_i}}{i^{n_i} n_i!} \left[\mathrm{Tr}(M^i)\right]^{n_i} \right| \right) \\
&\leq \sum_{\substack{a \leq \bar{a} \\ a = (n_1, \cdots, n_k)}} \mathbb{E}\left(\prod_{i \leq k} \frac{(N |z_i|)^{n_i}}{i^{n_i} n_i!} \left[N\lambda(M)^i\right]^{n_i} \right) \\
&= \sum_{\substack{a \leq \bar{a} \\ a = (n_1, \cdots, n_k)}} \mathbb{E}\left(\prod_{i \leq k} \frac{(N^2 |z_i|)^{n_i}}{i^{n_i} n_i!} \lambda(M)^{\sum_{i \leq k} i n_i} \right) \\
&= \sum_{\substack{a \leq \bar{a} \\ a = (n_1, \cdots, n_k)}} \prod_{i \leq k} \frac{(N^2 |z_i|)^{n_i}}{i^{n_i} n_i!} \mathbb{E}\left(\lambda(M)^{\sum_{i \leq k} i n_i} \right) \\
&= \xi_{\bar{a}} \bar{\lambda} < \infty .
\end{aligned}$$

Thus $\sup_{\bar{a}} \mathbb{E}(|\eta_{\bar{a}}|) < \infty$.

Second we will show that $\mathbb{E}(|\eta_{\bar{a}}| I_A)$ is uniformly absolutely continuous. We

let $\xi = \sup_{\bar{a}} \xi_{\bar{a}} < \infty$, $\delta = \varepsilon/(2\xi\bar{\lambda})$. Then if $\mathbb{E}(I_A) < \delta$, then

$$\begin{aligned}
\mathbb{E}(|\eta_{\bar{a}}|I_A) &= \mathbb{E}\left(\left|\sum_{\substack{a \le \bar{a} \\ a=(n_1,\cdots,n_k)}} \prod_{i\le k} \frac{(N\ z_i)^{n_i}}{i^{n_i} n_i!} \left[\mathrm{Tr}(M^i)\right]^{n_i}\right| I_A\right) \\
&\le \sum_{\substack{a \le \bar{a} \\ a=(n_1,\cdots,n_k)}} \mathbb{E}\left(\prod_{i\le k} \frac{(N\ |z_i|)^{n_i}}{i^{n_i} n_i!} \left[N\lambda(M)^i\right]^{n_i} I_A\right) \\
&\le \sum_{\substack{a \le \bar{a} \\ a=(n_1,\cdots,n_k)}} \prod_{i\le k} \frac{(N^2\ \lambda|z_i|)^{n_i}}{i^{n_i} n_i!}\, \mathbb{E}\left(\lambda(M)^{\sum_{i\le k} i n_i} I_A\right) \\
&\le \xi\bar{\lambda}\delta\ < \varepsilon\,.
\end{aligned}$$

Thus $\sup_{\bar{a}} \mathbb{E}(|\eta_{\bar{a}}|I_A) < \varepsilon$ for $\mathbb{E}(I_A) < \delta$.

As a consequence Claim 11.3.1 follows immediately from Lemma 11.3.2. □

From Claim 11.3.1 and Example 11.2 we have that

$$\begin{aligned}
< f(M) > &= \lim_{\bar{a}\to\infty} < \sum_{\substack{a \le \bar{a} \\ a=(n_1,\cdots,n_k)}} \prod_{i\le k} \frac{(N\ z_i)^{n_i}}{i^{n_i} n_i!} \left[\mathrm{Tr}(M^i)\right]^{n_i} > \\
&= \lim_{\bar{a}\to\infty} \sum_{\substack{a \le \bar{a} \\ a=(n_1,\cdots,n_k)}} \prod_{i\le k} \frac{(N\ z_i)^{n_i}}{i^{n_i} n_i!} < \prod_{i\le k} \left[\mathrm{Tr}(M^i)\right]^{n_i} > \\
&= \lim_{\bar{a}\to\infty} \sum_{\substack{a \le \bar{a} \\ a=(n_1,\cdots,n_k)}} \sum_{\Gamma} N^{v(\Gamma)+p(\Gamma)-e(\Gamma)} \prod_{i\le k} \frac{z_i^{n_i}}{i^{n_i} n_i!}\,, \qquad (11.3.4)
\end{aligned}$$

where the second sum is over all pointed fat graphs Γ with n_i vertices of degree i.

Note that

$$\begin{aligned}
&\text{the number of pointed (labeled) fat graphs} \qquad (11.3.5) \\
&= \text{the number of } \textit{unlabeled} \text{ fat graphs} \times \prod_i \frac{i^{n_i} n_i!}{|\mathrm{Aut}(\Gamma)|}.
\end{aligned}$$

Hence

$$< f(M) > = \lim_{\bar{a}\to\infty} \sum_{\substack{a \le \bar{a} \\ a=(n_1,\cdots,n_k)}} \sum_{\Gamma \text{ unlabeled}} \frac{N^{v(\Gamma)+p(\Gamma)-e(\Gamma)}}{|\mathrm{Aut}(\Gamma)|} \prod_{i\le k} z_i^{n_i}. \quad (11.3.6)$$

Next we observe in (11.3.4) that $v(\Gamma) + p(\Gamma) - e(\Gamma) = -2g(\Gamma) + 2c(\Gamma)$ is additive for the operation of the disjoint union. Taking the logarithm of $< f(M) >$ we get the following formula for the connected pointed fat graphs.

Proposition 11.3.4. *For every z_i with $|z_i| \in (0, \varepsilon_i)$ where $\varepsilon_i > 0$ satisfying (11.3.3) we have*

$$\begin{aligned}
&\log < f(M) > \\
&= \lim_{\bar{a}\to\infty} \sum_{\substack{a \le \bar{a} \\ a=(n_1,\cdots,n_k)}} \sum_{\Gamma pointed\ connected} N^{v(\Gamma)+p(\Gamma)-e(\Gamma)} \prod_{i\le k} \frac{z_i^{n_i}}{i^{n_i} n_i!} \\
&= \lim_{\bar{a}\to\infty} \sum_{\substack{a \le \bar{a} \\ a=(n_1,\cdots,n_k)}} \sum_{\Gamma pointed\ connected} N^{-2g(\Gamma)+2} \prod_{i\le k} \frac{z_i^{n_i}}{i^{n_i} n_i!}.
\end{aligned}$$

Again due to (11.3.6) we have

$$\begin{aligned}
&N^{-2} \log < f(M) > \\
&= \lim_{\bar{a}\to\infty} \sum_{\substack{a \le \bar{a} \\ a=(n_1,\cdots,n_k)}} \sum_{\Gamma \text{ unlabeled connected}} \frac{N^{-2g(\Gamma)}}{|\mathrm{Aut}(\Gamma)|} \prod_{i\le k} z_i^{n_i}. \quad (11.3.7)
\end{aligned}$$

Now we are ready to state the formula for the asymptotic number of unlabeled connected planar fat graphs. For the proof we need Fatou's lemma (Theorem 2 (c) in Section 2, Chap. II in [131]):

Lemma 11.3.5. *Let $\{\xi_n\}$ be a sequence of random variables. If $\xi_n \le \eta$ for every $n \ge 1$ and $\mathbb{E}(\eta) < \infty$, then*

$$\mathbb{E}(\liminf_n \xi_n) \le \liminf_n E(\xi_n) \le \limsup_n \mathbb{E}(\xi_n) \le \mathbb{E}(\limsup_n \xi_n).$$

Theorem 11.3.6. *For every z_i with $|z_i| \in (0, \varepsilon_i)$ where $\varepsilon_i > 0$ satisfying (11.3.3) and (11.3.8) we have*

$$\begin{aligned}
&\lim_{N\to\infty} N^{-2} \log < f(M) > \\
&= \lim_{\bar{a}\to\infty} \sum_{\substack{a \le \bar{a} \\ a=(n_1,\cdots,n_k)}} \sum_{\Gamma\ unlabeled\ connected\ planar} \frac{1}{|\mathrm{Aut}(\Gamma)|} \prod_{i\le k} z_i^{n_i},
\end{aligned}$$

where Γ has n_i vertices of degree i, $i \ge 0$.

Proof. From (11.3.7) we know that

$$\begin{aligned}
&\lim_{N\to\infty} N^{-2} \log < f(M) > \\
&= \lim_{\bar{a}\to\infty} \sum_{\substack{a \le \bar{a} \\ a=(n_1,\cdots,n_k)}} \sum_{\Gamma \text{ unlabeled connected planar}} \frac{1}{|\mathrm{Aut}(\Gamma)|} \prod_{i\le k} z_i^{n_i} \\
&+ \lim_{N\to\infty} \lim_{\bar{a}\to\infty} \sum_{\substack{a \le \bar{a} \\ a=(n_1,\cdots,n_k)}} \sum_{\Gamma \text{ unlabeled connected non-planar}} \frac{N^{-2g(\Gamma)}}{|\mathrm{Aut}(\Gamma)|} \prod_{i\le k} z_i^{n_i}.
\end{aligned}$$

It is enough to show that

$$\lim_{N\to\infty} \lim_{\bar{a}\to\infty} \sum_{\substack{a \le \bar{a} \\ a=(n_1,\cdots,n_k)}} \sum_{\Gamma \text{ unlabeled connected non-planar}} \frac{N^{-2g(\Gamma)}}{|\mathrm{Aut}(\Gamma)|} \prod_{i\le k} z_i^{n_i} = 0.$$

For a given sequence $\eta = \{\eta(a)\}_{a\in I}$, where I is a linearly ordered Abelian group introduced above, we can think of the series

$$\lim_{\bar{a}\to\infty} \sum_{\substack{a \le \bar{a} \\ a = (n_1, \cdots, n_k)}} \eta(a)$$

as the expectation over I equipped with measure μ such that $\mu(a) = 1$ iff $a \in I$. That is,

$$\lim_{\bar{a}\to\infty} \sum_{\substack{a \le \bar{a} \\ a = (n_1, \cdots, n_k)}} \eta(a) = \mathbb{E}_\mu(\eta)\,.$$

We define a sequence $\eta = \{\eta(a)\}_{a\in I}$ by, for each $a = (n_1, \cdots, n_k)$,

$$\eta(a) = \sum_{\Gamma_{np}} \prod_{i\le k} |z_i|^{n_i}\,,$$

where the sum is over all unlabeled connected *non-planar* fat graphs Γ_{np} with n_i vertices of degree i. We select ε_i's so that for all z_i with $|z_i| \in (0, \varepsilon_i)$

$$\mathbb{E}_\mu(\eta) = \lim_{\bar{a}\to\infty} \sum_{\substack{a \le \bar{a} \\ a = (n_1, \cdots, n_k)}} \sum_{\Gamma_{np}} \prod_{i\le k} |z_i|^{n_i} < \infty\,, \tag{11.3.8}$$

where the second sum is over all unlabeled connected *non-planar* fat graphs Γ_{np} with n_i vertices of degree i.

We consider a sequence $\xi_N = \{\xi_N(a)\}_{a\in I}$ defined for each $a = (n_1, \cdots, n_k)$ by,

$$\xi_N(a) = \sum_{\Gamma_{np}} \frac{N^{-2g(\Gamma_{np})}}{|\mathrm{Aut}(\Gamma_{np})|} \prod_{i\le k} z_i^{n_i}\,,$$

where the sum is over all unlabeled connected non-planar fat graphs Γ_{np} with n_i vertices of degree i. Note that we have

$$\mathbb{E}_\mu(\xi_N) = \lim_{\bar{a}\to\infty} \sum_{\substack{a \le \bar{a} \\ a = (n_1, \cdots, n_k)}} \xi_N(a)\,,$$

and $|\xi_N(a)| \le \eta(a)$ for every $N \ge 1$ since $g(\Gamma_{np}) > 0$. In addition the sequence $\lim_{N\to\infty} \xi_N = \{\lim_{N\to\infty} \xi_N(a)\}_{a\in I}$ satisfies

$$\begin{aligned}
\lim_{N\to\infty} \xi_N(a) &= \lim_{N\to\infty} \sum_{\Gamma_{np}} \frac{N^{-2g(\Gamma_{np})}}{|\mathrm{Aut}(\Gamma_{np})|} \prod_{i\le k} z_i^{n_i} \\
&= \sum_{\Gamma_{np}} \prod_{i\le k} z_i^{n_i} \lim_{N\to\infty} \frac{N^{-2g(\Gamma_{np})}}{|\mathrm{Aut}(\Gamma_{np})|} \\
&= 0\,,
\end{aligned}$$

because $g(\Gamma_{np}) > 0$ for non-planar fat graphs Γ_{np}. Thus we obtain

$$\mathbb{E}_\mu(\lim_{N\to\infty} \xi_N) = \lim_{\bar{a}\to\infty} \sum_{\substack{a \le \bar{a} \\ a = (n_1, \cdots, n_k)}} \lim_{N\to\infty} \xi_N(a) = 0\,.$$

By Fatou's lemma we have that

$$\liminf_{N\to\infty} \mathbb{E}_\mu(\xi_N) = \limsup_{N\to\infty} \mathbb{E}_\mu(\xi_N) = 0\,.$$

It follows then that

$$0 = \lim_{N\to\infty} \mathbb{E}_\mu(\xi_N) \quad = \quad \lim_{N\to\infty} \lim_{\bar{a}\to\infty} \sum_{\substack{a \le \bar{a} \\ a = (n_1, \cdots, n_k)}} \sum_{\Gamma_{np}} \frac{N^{-2g(\Gamma_{np})}}{|\mathrm{Aut}(\Gamma_{np})|} \prod_{i\le k} z_i^{n_i}\,,$$

where the second sum is over all unlabeled connected non-planar fat graphs Γ_{np} with n_i vertices of degree i. □

11.4 Directed graphs

In this section we move from matrices to directed graphs. Let M be an $N \times N$ matrix and let $D = D(M) = (N, N \times N)$ be a directed graph with weights on directed edges given by M.

Pointed closed walks. Now $\mathrm{Tr}(M^n)$ may be interpreted as $\sum_p \prod_{e\in p} M_e$, where the sum is over all pointed (i.e., with a prescribed beginning) closed walks in D of length n. Similarly, $\mathrm{Tr}(M^3)^4\mathrm{Tr}(M^2)^3$ may be interpreted as

$$(\sum_{p_1} \prod_{e\in p_1} M_e)^4 (\sum_{p_2} \prod_{e\in p_2} M_e)^3,$$

where the first sum is over all pointed closed walks p_1 in D of length 3, and the second sum is over all pointed closed walks p_2 in D of length 2.

Hence if $f(M) = \mathrm{Tr}(M^3)^4\mathrm{Tr}(M^2)^3$ we get:

$$< f > = \sum_{q=q_1q_2\ldots q_7} \sum_{P} \prod_{\{e,e'\}\in P} 1/N,$$

where the second sum is over all *proper pairings* P of the directed edges of the disjoint union $q = q_1q_2\ldots q_7$ of 7 pointed closed walks, from which 4 have length 3 and remaining 3 have length 2. Two directed edges form a proper pairing if one is reversed the other. We also say that pair (q, P), P proper, *contributes* to $< f >$.

Flows and even sets of directed edges. Toward the enumeration of planar graphs with a given degree sequence, the starting idea is to replace pointed closed walks by *flows*. A non-negative integer function on the directed edges of D is *flow* if for each vertex v of D, the sum of $f(e)$ over the incoming edges is the same as over the outgoing edges. It turns out that a restrictive subset of the set of all flows consisting of $0, 1$ flows is already interesting. These flows naturally lead to even subsets of edges defined below.

Definition 11.4.1. *A subset $A \subset E(D)$ of directed edges is even if A can be written as a union of edge-disjoint cycles of length bigger than two.*

We further denote by $\mathcal{A}(r_3,\dots,r_k)$ the set of all even sets of edges which can be decomposed into r_i cycles of length i $(i=3,\dots,k)$ and by $\mathcal{A}'(r_3,\dots,r_k)$ the set of all even sets of edges *with a fixed decomposition* into r_i cycles of length i $(i=3,\dots,k)$. We let $\mathcal{A}(r)=\cup_{r=r_3+\dots r_k}\mathcal{A}(r_3,\dots,r_k)$, $\mathcal{A}=\cup_{r>0}\mathcal{A}(r)$, and $\mathcal{A}'(r)=\cup_{r=r_3+\dots r_k}\mathcal{A}'(r_3,\dots,r_k)$, $\mathcal{A}'=\cup_{r>0}\mathcal{A}'(r)$.

A *proper pairing* of an even set A is a partition of A into pairs $(ij),(ji)$ of oppositely directed edges.

Cycle double cover conjectures. It is not true that a set of directed edges, which induces the same indegree and outdegree at each vertex, is a union of disjoint directed cycles of length bigger than 2. This is closely related to the *cycle double cover conjectures.*

Definition 11.4.2. *Let G be an undirected graph. A collection of its cycles is called a* cycle double cover *(CDC) if each edge belongs to exactly two of the cycles. Moreover it is called a* directed cycle double cover *(DCDC) if it is possible to orient the cycles so that they go oppositely on each edge.*

Some of the most puzzling conjectures of discrete mathematics are centred around this notion. A graph is *bridgeless* if it cannot be disconnected by deletion of a single edge. Clearly a graph with a bridge does not have a CDC. On the other hand, there are

- *Cycle double cover conjecture*: Is it true that each bridgeless graph has a CDC?
- *Directed cycle double cover conjecture*: Is it true that each bridgeless graph has a DCDC?

The following observation is straightforward.

Observation 11.4.3. *Let $q\in\mathcal{A}'(r_3,\dots,r_k)$ come with a fixed decomposition into r_i cycles $C_{i,1},\dots,C_{i,r_i}$ of length i $(i=3,\dots,k)$. Let P be a proper pairing of q. Then the cycles form a DCDC of the simple graph with vertices $\{1,\dots,N\}$ and the edges given by the pairings.*

Definition 11.4.4. *Let M be a matrix. We let*

$$g_r(M) \;=\; \sum_{c\in\mathcal{A}(r)}\prod_{e\in c}M_e,$$

$$g'_r(M) \;=\; \sum_{c\in\mathcal{A}'(r)}\prod_{e\in c}M_e,$$

and we define $g,g',g_{r_3,\dots,r_k}(M),g'_{r_3,\dots,r_k}(M)$ analogously.

We call $\prod_{e\in c}M_e$ the weight of the cycle c.

Next definition and proposition are crucial.

Definition 11.4.5. *Let G be a finite simple undirected graph with at most N vertices and with a DCDC. Then let $c(G)$ be the set of all pairs (q,P) so that there is a colouring d of the vertices of G by colours $\{1,\dots,N\}$, where each vertex gets a different color, $q=\{(d(x),d(y)),(d(y),d(x));\{x,y\}\in E(G)\}$, and P consists of all the pairs $([(d(x),d(y)),(d(y),d(x))];\{x,y\}\in E(G))$.*

We remark that each such q is an even subset of directed edges of $D = (N, N \times N)$ since G has a DCDC, and $|c(G)| = N(N-1)\dots(N-|V(G)|+1)$.

Proposition 11.4.6. *A term (q, P) contributes to $< g_{r_3,\dots,r_k}(M) >$ if and only if there is a simple graph G with a DCDC consisting of r_i cycles of length i $(i = 3, \dots, k)$ such that $(q, P) \in c(G)$.*

Proof. If $(q, P) \in c(G)$, then any DCDC provides a partition of q into its cycles and hence (q, P) contributes to $< g_{r_3,\dots,r_k}(M) >$. On the other hand if (q, P) contributes to $< g_{r_3,\dots,r_k}(M) >$, then letting G be the graph with the vertices from $\{1, \dots, N\}$ and the edges given by P we get that G is simple since q consists of edge-disjoint directed cycles, it has a DCDC consisting of r_i cycles of length i $(i = 3, \dots, k)$, and $(q, P) \in c(G)$. □

Proposition 11.4.7. *If $c(G) \cap c(G') \neq \emptyset$, then G is isomorphic to G'. Moreover, if G is isomorphic to G', then $c(G) = c(G')$.*

Proof. If $(q, P) \in c(G) \cap c(G')$, then the construction of q induces a function between the sets of vertices of G and G', and P gives the edges of both G, G'. Hence they are isomorphic. The second part is true since the definition of $c(G)$ does not depend on 'names' of the vertices. □

As a consequence we have

Theorem 11.4.8.

$$< g_{r_3,\dots,r_k}(M) > = \sum_{[G]} \frac{N(N-1)\dots(N-|V(G)|+1)}{N^{e(G)}},$$

where the sum is over all isomorphism classes of simple graphs with at most N vertices that have a DCDC consisting of r_i cycles of length i $(i = 3, \dots, k)$. Moreover

$$< g'_{r_3,\dots,r_k}(M) > = \sum_{[G]'} \frac{N(N-1)\dots(N-|V(G)|+1)}{N^{e(G)}},$$

where the sum is over all isomorphism classes of simple graphs with at most N vertices and with a specified *DCDC consisting of r_i cycles of length i $(i = 3, \dots, k)$.*

Analogous statements hold for g_r, g'_r, g, g'.

11.5 Calculations

The integral $< g'(M) >$ counts all the directed cycle double covers of graphs on at most N vertices and hence its calculation is an attractive task which need not be hopeless. We show next a curious formula for $g'(M)$ which identifies it with an Ihara-Selberg-type function (see Theorem 11.5.3). Let us recall that

$$g'(M) = \sum_{c \in \mathcal{A}'} \prod_{e \in c} M_e, \tag{11.5.1}$$

is the generating function (with variables M_e's) of the collections of edge-disjoint directed cycles of length at least three, in the directed graph $D = D(M) = (N, N \times N)$.

Construction of digraph D'. We first construct a directed graph D' with the weights on the *transitions between the edges*. First we split each vertex of D, i.e., we replace each vertex v by new edge $e(v)$ and we let all the edges of D entering v enter the initial vertex of $e(v)$, and all the edges of D leaving v leave the terminal vertex of $e(v)$. If edge g enters v in D then we define the weight of the transition $w(g, e(v)) = M_g$. We let all the remaining transition be equal to one (see Figure 11.4, the first two parts).

Finally, for each pair g_1, g_2 of oppositely directed edges of D, say $g_1 = (uv), g_2 = (vu)$ we introduce new vertex v_g and we let both g_1, g_2 pass through it; equivalently, we subdivide both g_1, g_2 by one vertex and identify this pair of vertices into unique vertex called v_g (and thus we have new edges $(uv_g), (v_g v)$ from $g_1 = (uv)$, and new edges $(vv_g), (v_g u)$ from $g_2 = (vu)$) (see Figure 11.4, the last two parts).

We let the weights of the transitions at vertex v_g *between* g_1 and g_2 (i.e., between (uv_g) and $(v_g u)$ and between (vv_g) and $(v_g v)$) be equal to zero, the transitions *along* g_1 and g_2 (i.e., between (uv_g) and $(v_g v)$ and between (vv_g) and $(v_g u)$) be equal to one, and the transitions between $(v_g v)$ and $e(v)$ be equal to M_{g_1} and between $(v_g u)$ and $e(u)$ be equal to M_{g_2}. See an example in Figure 11.5.

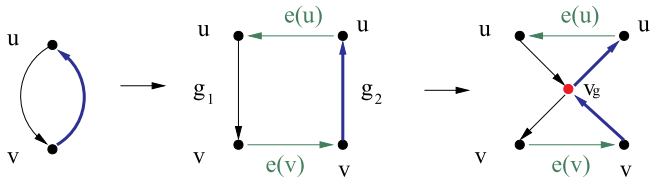


Figure 11.4: Construction of $e(v)$ and v_g

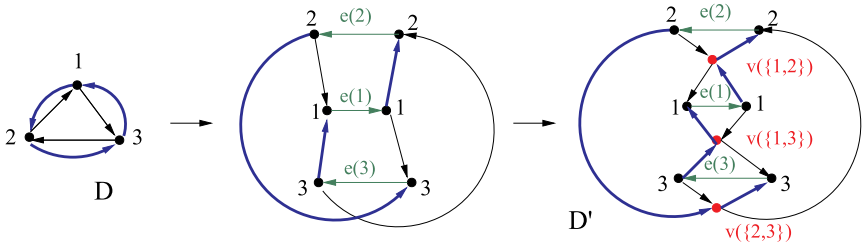


Figure 11.5: An example of the construction of digraph D'

In what follows, the directed closed walk is considered *not pointed*. We let the weight of the directed closed walk be the product of the weights of its transitions.

Observation 11.5.1. *There is a weight preserving bijection between the set of the directed cycles of D of length at least three and the set of the closed directed*

walks of D' of a non-zero weight which go through each directed edge and through each vertex v_g at most once.

Proof. This follows directly from the construction of D'. □

Definition 11.5.2. *We define the rotation number for each closed walk w of D' with a non-zero weight by induction as follows: first order the directed edges of D', say as $a_1, \ldots, a_m$, so that the edges $e(v), v \in V(D)$ form the terminal segment. Then*

(1) If w is a directed cycle, then we let $r(w) = -1$.

(2) Let w go at least twice through a directed edge. Let a be the first such edge in the fixed ordering. Hence w is a concatenation of two shorter closed walks w_1, w_2, both containing a. If $a \neq e(v)$ for some v then we let $r(w) = r(w_1)r(w_2)$. If $a = e(v)$, then we let $r(w) = 0$.

(3) If none of 1.,2. applies, w must go through a vertex v_g (introduced in the definition of D') at least twice. Then we again let $r(w) = 0$.

Theorem 11.5.3. *Let $g'(M)$ be defined as (11.5.1). Then*

$$g'(M) = \prod_p (1 - r(p)w(p)),$$

where the product is over all aperiodic closed directed walks p in D' and $w(p)$ denotes the weight of p.

To prove Theorem 11.5.3 we will need a curious lemma on coin arrangements stated below. It has been introduced by Sherman [130] in the study of 2-dimensional Ising problem.

Lemma 11.5.4 (A lemma on coin arrangements.)**.** *Suppose we have a fixed collection of N objects of which m_1 are of one kind, m_2 are of second kind, $\cdots$, and m_n are of n-th kind. Let b_k be the number of exhaustive unordered arrangements of these symbols into k disjoint, nonempty, circularly ordered sets such that no two circular orders are the same and none are periodic. For example let us have 10 coins of which 3 are pennies, 4 are nickles and 3 are quarters. Then $\{(p,n),(n,p),(p,n,n,q,q,q)\}$ is not a correct arrangement since (p,n) and (n,p) represent the same circular order. If $N > 1$ then $\sum_{i=1}^{N}(-1)^{i+1}b_i = 0$.*

Proof of Lemma 11.5.4. The lemma follows immediately if we expand the LHS of the following *Witt Identity* and collect terms where the sums of the exponents of the z_i's are the same.

Witt Identity (see [75]): Let $z_1, ..., z_k$ be commuting variables. Then

$$\prod_{m_1,\ldots,m_k \geq 0} (1 - z_1^{m_1} ... z_k^{m_k})^{M(m_1,\ldots,m_k)} = 1 - z_1 - z_2 - ... - z_k,$$

where $M(m_1, ..., m_k)$ is the number of different nonperiodic sequences of z_i's taken with respect to circular order. □

Proof of Theorem 11.5.3. We first show that the coefficients corresponding to the products of variables where at least one $M_e, e \neq e(v)$, appears with the exponent greater than one, are all equal to zero.

Let us denote $W(p) = -r(p)w(p)$. Let A_1 be the set of all non-periodic closed walks p such that a_1 appears in p. Each $p \in A_1$ has a unique factorization into words $(W_1, ..., W_k)$ each of which starts with a_1 and has no other appearance of a_1.

Let S be a monomial summand in the expansion of $\prod_{p \in A_1}(1+W(p))$. Hence S is a product of finitely many $W(p), p \in A_1$.

Each $p \in A_1$ has a unique factorization into words defined above. Each word may appear several times in the factorization of p and also in the factorization of different non-periodic closed walks. Let $B(D')$ be the set-system of all the words (with repetition) appearing in the factorizations of the aperiodic closed walks of D'.

It directly follows from Lemma 11.5.4, the lemma on coin arrangements, that the sum of all monomial summands S in the expansion of $\prod_{p \in A_1}(1 + W(p))$, which have the same 'coins' $B(D')$ of more than one element is zero. Hence the monomial summands S which survive in the expansion of $\prod_{p \in A_1}(1 + W(p)$ all have $B(D')$ consisting of exactly one word. Hence they cannot have a_1 with exponent bigger than one. Now we can repeat the same consideration for the other edges different from $e(v), v \in V$.

Hence the only terms of the expansion of the infinite product that survive have all $M_e, e \neq e(v)$, with the exponent at most one.

We know from Observation 11.5.1 that the collections of the edge-disjoint directed cycles of length at least three in D' correspond to the collections of the directed closed aperiodic walks of D' where each edge $e \neq e(v)$ of D' appears at most once; by above, these exactly *have chance* to survive.

Each term of $g'(M)$ may be expressed several times as a product of aperiodic closed walks of D', *but* only one such expression survives in the infinite product since if a closed walk goes through an edge $e(v)$ or through a vertex v_g more than once, its rotation is defined to be zero. Hence $g'(M)$ is counted correctly in the infinite product. □

Remark. Let us write $r(p) = q^{\mathrm{rot}(p)}$, where $q = -1$. Without the zero values of $r(p)$, function $\mathrm{rot}(p)$ is additive when we 'smoothen' p into directed cycles. The integer lattice generated by the directed cycles has a basis which may be constructed e.g., from the ear-decomposition [68]; the function $\mathrm{rot}(p)$ may be split into contributions of the edge-transitions for the basis, and since it is a basis, it may be split also for all the directed cycles. Hence **i**f the additivity property holds, $\mathrm{rot}(p)$ may be split into the contributions $\mathrm{rot}(t)$ of the edge-transitions t for the aperiodic closed walks. Hence

$$\prod_p (1 - r(p)w(p)) = \prod_p (1 - \prod_{t \in p} (-1)^{\mathrm{rot}(t)} w(t)).$$

This formula transforms the infinite product into the *Ihara-Selberg function*. It was studied by Bass in [10] who proved that it is equal to a determinant. A combinatorial proof was given by Foata and Zeilberger in [65].

Due to the zero values of $r(p)$ it is not clear how to split the rotation into individual edge-transitions. A determinant-type formula, perhaps non-commutative, may however exist. Moreover the Ihara-Selberg function and its

inverse appear frequently in theoretical physics and so the matrix integral of $g'(M)$ may have, via the formula of Theorem 11.5.3, an interesting physics interpretation.

11.6 Planar graphs with given degree sequence

Back to fat graphs. Loopless fat graph F is called *cyclic* if each face of F is a cycle of $G(F)$. For cyclic F we define its *dual* F^* as the fat graph whose islands are the discs bounded by the faces of F, and whose bridges may be identified with the bridges of F. Note that F^* is again loopless and thus we have:

Observation 11.6.1. *A fat graph F is cyclic if and only if it is a dual of a cyclic fat graph; in particular $F^{**} = F$.*

Definition 11.6.2. *A fat graph F is called* relevant *if it is cyclic and F^* is a* simple *fat graph.*

By definition a loopless fat graph is relevant if each face is a cycle of $G(F)$, and each pair of faces of F shares at most one bridge. ($G(F^*)$ does not have multiple edges). Moreover, if W^* is simple then W does not have vertices of degree at most two.

A *compressed* fat graph is a pair (F, P) where F is a fat graph and P is a partition of the set of its vertices. We denote by $G_P(F)$ the (abstract) graph which is obtained by the identification of the vertices of each class of P in $G(F)$.

Next definition is more technical.

Definition 11.6.3. *A pair (W, Q) where W is a relevant fat graph and Q is a partition of the set of its faces is called* relevant *if W is relevant and $G_Q(W^*)$ is a simple graph.*

When Q partitions of faces into themselves, we denote it by $\emptyset$.

Proposition 11.6.4. *There is a natural bijection between the set of the relevant pairs (W, Q) such that W has exactly r_i vertices of degree i ($i \geq 3$) and the set of the simple graphs G with a specified DCDC consisting of r_i cycles of length i. The bijection sends (W, Q) to $G_Q(W^*)$.*

Proof. We realize each cycle from the DCDC as a disc and glue the discs together along the pairs of corresponding oppositely oriented edges. We get a surface where some vertices are identified. When we split the identified vertices, we get an honest compact 2-dimensional orientable surface with a graph G' embedded on it. We change the embedding of G' into the fat graph F'; it is cyclic since its faces are exactly the cycles of the DCDC we started with.

Graph G' is obtained from G by splitting off some vertices. This defines a partition Q of the set of vertices of G'.

This produces a relevant pair (F'^*, Q).

Moreover, it is not difficult to observe that this construction may be reversed and thus we get a bijection. □

This implies the following.

Corollary 11.6.5. *Let* $< g_{r_3,\ldots,r_k}(M) >$ *be as in Definition 11.4.4.*

$$< g_{r_3,\ldots,r_k}(M) > = \sum_{[(W,Q)]} \frac{N(N-1)\ldots(N-\alpha+1)}{N^{e(G_Q(W^*))}},$$

where (W,Q) *is a relevant pair such that* W *has exactly* r_i *vertices of degree* i *(*$i \geq 3$*), and partition* Q *of its faces into* $\alpha \leq N$ *parts;* $[\cdot]$ *denotes the isomorphism equivalence class of* $G_Q(W^*)$*.*

Extracting planar graphs. We are grateful to Bojan Mohar for the following characterization of relevant planar fat graphs.

Proposition 11.6.6. *A planar fat graph is relevant iff each of its connected components is 2-vertex connected and 3-edge-connected.*

Proof. A planar graph is 2-connected iff its dual is 2-connected iff each face is a cycle; here parallel edges in the dual are allowed, but no loops. The parallel edges are eliminated by the 3-edge-connectivity. □

A theorem analogous to Theorem 11.3.6 holds.

Theorem 11.6.7. *For every* z_i *with* $|z_i| \in (0, \varepsilon_i)$ *we have*

$$\lim_{N\to\infty} N^{-2} \log \sum_{r_3,\ldots} < g_{r_3,\ldots}(M) > \prod_i \frac{(Nz_i)^{r_i}}{r_i!} \quad = \quad \sum_{r_3,\ldots} \sum_{[\Gamma]^*} \prod_i \frac{z_i^{r_i}}{r_i!},$$

where Γ *is a 2-connected and 3-edge-connected planar fat graph with* r_i *vertices of degree* i*,* $i \geq 0$*, and* $[]^*$ *is the isomorphism equivalence class of* Γ^**.*

Note that the coefficient of $\prod_i z_i^{r_i}/r_i!$ in the right hand side of the power series in Theorem 11.6.7 is the same as the number of unlabelled planar graphs with r_i faces of length i, whose dual is 2-connected and 3-edge-connected.

Proof. From Corollary 11.6.5 we get

$$\begin{aligned}
&\lim_{N\to\infty} N^{-2} \log \sum_{r_3,\ldots} < g_{r_3,\ldots}(M) > \prod_i \frac{(Nz_i)^{r_i}}{r_i!} \\
= \;& \lim_{N\to\infty} N^{-2} \log \sum_{r_3,\ldots} \sum_{[(W,Q)]} N^{v(W)-e(W)} N(N-1)\ldots(N-\alpha+1) \prod_i \frac{z_i^{r_i}}{r_i!} \\
= \;& \lim_{N\to\infty} N^{-2} \log \sum_{r_3,\ldots} \sum_{[(W,Q)]} N^{v(W)-e(W)+p(W)} \\
&\qquad \times \prod_{1\leq i\leq \alpha-1} \left(1-\frac{i}{N}\right) N^{\alpha-p(W)} \prod_i \frac{z_i^{r_i}}{r_i!} \qquad (11.6.1)
\end{aligned}$$

where (W,Q) is a relevant pair such that W has exactly r_i vertices of degree i ($i \geq 3$), and partition Q of its faces into $\alpha \leq N$ parts, and $[\cdot]$ denotes the isomorphism equivalence class, as in Corollary 11.6.5.

As in Proposition 11.3.4, it is natural to express (11.6.1) in terms of *connected* W. As before $v(W) - e(W) + p(W) = -2g(W) + 2c(W)$ is additive for the

operation of the disjoint union, but the rest is only submultiplicative. Hence we proceed in two steps, bounding (11.6.1) from above and from below.

First, we upper bound the LHS of (11.6.1): using a partition $\bar{Q}$ of faces of W into $\alpha \leq N$ parts, each of whose partition classes lies in the same connected component of W, we obtain

$$\begin{aligned}
&\sum_{[(W,Q)]} N^{v(W)-e(W)+p(W)} \prod_{1\leq i\leq \alpha-1}\left(1-\frac{i}{N}\right) N^{\alpha-p(W)} \prod_i \frac{z_i^{r_i}}{r_i!} \\
&\leq \sum_{[(W,\bar{Q})]} N^{v(W)-e(W)+p(W)} \prod_{1\leq i\leq \alpha-1}\left(1-\frac{i}{N}\right) N^{\alpha-p(W)}(1+O(N^{-1})) \prod_i \frac{z_i^{r_i}}{r_i!}
\end{aligned}$$

for large N. Thus we can upper bound Equation (11.6.1) by

$$\begin{aligned}
\lim_{N\to\infty} N^{-2} \log \sum_{r_3,\dots} \sum_{[(W,\bar{Q})]} & N^{v(W)-e(W)+p(W)} \\
&\times \prod_{1\leq i\leq \alpha-1}\left(1-\frac{i}{N}\right) N^{\alpha-p(W)}(1+O(N^{-1})) \prod_i \frac{z_i^{r_i}}{r_i!}, \qquad (11.6.2)
\end{aligned}$$

which is equal to

$$\begin{aligned}
&\lim_{N\to\infty} N^{-2} \log (1+O(N^{-1})) + \lim_{N\to\infty} N^{-2} \log \sum_{r_3,\dots} \sum_{[(W,\bar{Q})]} N^{v(W)-e(W)+p(W)} \\
&\qquad \times \prod_{1\leq i\leq \alpha-1}\left(1-\frac{i}{N}\right) N^{\alpha-p(W)} \prod_i \frac{z_i^{r_i}}{r_i!} \\
&= \lim_{N\to\infty} N^{-2} \log \sum_{r_3,\dots} \sum_{[(W,\bar{Q})]} N^{v(W)-e(W)+p(W)} \\
&\qquad \times \prod_{1\leq i\leq \alpha-1}\left(1-\frac{i}{N}\right) N^{\alpha-p(W)} \prod_i \frac{z_i^{r_i}}{r_i!}. \qquad (11.6.3)
\end{aligned}$$

Further using the relation between the logarithm of the generating function with respect to $[(W,\bar{Q})]$ and the generating function w.r.t. $[(W_C,\bar{Q})] = [(W_C,Q)]$, where W_C denotes *connected* W, Equation (11.6.3) becomes

$$\begin{aligned}
&\lim_{N\to\infty} N^{-2} \sum_{r_3,\dots} \sum_{[(W_C,Q)]} N^{v(W_C)-e(W_C)+p(W_C)} \prod_{1\leq i\leq \alpha-1}\left(1-\frac{i}{N}\right) N^{\alpha-p(W_C)} \prod_i \frac{z_i^{r_i}}{r_i!} \\
&= \lim_{N\to\infty} \sum_{r_3,\dots} \sum_{[(W_C,Q)]} N^{-g(W_C)} \prod_{1\leq i\leq \alpha-1}\left(1-\frac{i}{N}\right) N^{\alpha-p(W_C)} \prod_i \frac{z_i^{r_i}}{r_i!} \\
&= \lim_{N\to\infty} \sum_{r_3,\dots} \sum_{[(W_C,Q)],\ W_C \text{ non-planar}} N^{-g(W_C)} \prod_{1\leq i\leq \alpha-1}\left(1-\frac{i}{N}\right) N^{\alpha-p(W_C)} \prod_i \frac{z_i^{r_i}}{r_i!} \\
&\quad + \lim_{N\to\infty} \sum_{r_3,\dots} \sum_{[(W_C,Q)],\ W_C \text{ planar}} \prod_{1\leq i\leq \alpha-1}\left(1-\frac{i}{N}\right) N^{\alpha-p(W_C)} \prod_i \frac{z_i^{r_i}}{r_i!}.
\end{aligned}$$

which, due to Fatou's lemma as in Theorem 11.3.6, becomes

$$\begin{aligned}
&\sum_{r_3,\dots}\sum_{[(W_C,Q)],\ W_C \text{ non-planar}} \lim_{N\to\infty} N^{-g(W_C)} \prod_{1\le i\le\alpha-1}\left(1-\frac{i}{N}\right) N^{\alpha-p(W_C)} \prod_i \frac{z_i^{r_i}}{r_i!} \\
&\quad + \sum_{r_3,\dots}\sum_{[(W_C,Q)],\ W_C \text{ planar}} \lim_{N\to\infty} \prod_{1\le i\le\alpha-1}\left(1-\frac{i}{N}\right) N^{\alpha-p(W_C)} \prod_i \frac{z_i^{r_i}}{r_i!} \\
&= \sum_{r_3,\dots}\sum_{[(W_C,Q)],\ W_C \text{ planar}} \lim_{N\to\infty} \prod_{1\le i\le\alpha-1}\left(1-\frac{i}{N}\right) N^{\alpha-p(W_C)} \prod_i \frac{z_i^{r_i}}{r_i!} \\
&= \sum_{r_3,\dots}\sum_{[(W_C,\emptyset)],\ W_C \text{ planar}} \prod_i \frac{z_i^{r_i}}{r_i!},
\end{aligned}$$

where the last equality follows from the fact that for a relevant pair (W_C, Q) with $Q \neq \emptyset$, $\lim_{N\to\infty}\prod_{1\le i\le\alpha-1}\left(1-\frac{i}{N}\right)N^{\alpha-p(W_C)} = 0$, since $\alpha < p(W)$. On the other hand,

$$\lim_{N\to\infty} \prod_{1\le i\le\alpha-1}\left(1-\frac{i}{N}\right) N^{\alpha-p(W_C)} = 1$$

for a relevant pair (W_C, Q) with $Q = \emptyset$, since $\alpha = p(W)$. This together with Proposition 11.6.6 implies that

$$\begin{aligned}
&\lim_{N\to\infty} N^{-2} \log \sum_{r_3,\dots} < g_{r_3,\dots}(M) > \prod_i \frac{(Nz_i)^{r_i}}{r_i!} \\
\le\ & \sum_{r_3,\dots}\sum_{[(W_C,\ \emptyset)],\ W_C \text{ relevant planar}} \prod_i \frac{z_i^{r_i}}{r_i!} \ = \ \sum_{r_3,\dots}\sum_{[\Gamma]^*}\prod_i \frac{z_i^{r_i}}{r_i!}.
\end{aligned}$$

where Γ is 2-connected and 3-edge connected planar with r_i vertices of degree i for $i \ge 3$ and $[]^*$ is the isomorphism equivalence class of Γ^*.

Finally, we lower bound the LHS of (11.6.1):

$$\begin{aligned}
&\lim_{N\to\infty} N^{-2} \log \sum_{r_3,\dots} < g_{r_3,\dots}(M) > \prod_i \frac{(Nz_i)^{r_i}}{r_i!} \\
\overset{(11.6.1)}{\ge}\ & \lim_{N\to\infty} N^{-2} \sum_{r_3,\dots}\sum_{[(W_C,\emptyset)]} N^{v(W_C)-e(W_C)+p(W_C)} \prod_{1\le i\le\alpha-1}\left(1-\frac{i}{N}\right)\prod_i \frac{z_i^{r_i}}{r_i!} \\
=\ & \lim_{N\to\infty} \sum_{r_3,\dots}\sum_{[(W_C,\ \emptyset)]} N^{-g(W_C)} \prod_{1\le i\le\alpha-1}\left(1-\frac{i}{N}\right)\prod_i \frac{z_i^{r_i}}{r_i!} \\
=\ & \sum_{r_3,\dots}\sum_{[(W_C,\ \emptyset)],W_C,\text{ relavant planar}} \prod_i \frac{z_i^{r_i}}{r_i!},
\end{aligned}$$

where the last equality follows from Fatou's lemma, similarly as for the upper bound.

Since the two bounds are equal, the theorem follows.

□

Part IV

Random Graph Processes

Chapter 12

Mininum Degree Random Multigraph Process

In this chapter we study how the minimum degree random multigraph process evolves as the number of edges increases. The minimum degree random multigraph process $(G_{\min}(n,M))_{M\geq 0}$ is a Markov chain, whose state space is the set of graphs with vertex set $[n] := \{1, 2, \cdots, n\}$. It starts with an empty graph on n vertices, and for $M > 0$, $G_{\min}(n, M+1)$ is obtained from $G_{\min}(n, M)$ by first choosing a vertex of minimum degree in $G_{\min}(n, M)$ uniformly at random, and then connecting it by a new edge to another vertex chosen uniformly at random among all the remaining vertices in $G_{\min}(n, M)$.

We show that a typical $G_{\min}(n, M)$ becomes connected quite quickly. In particular, when the minimum degree of $G_{\min}(n, M)$ reaches three it is a.a.s. connected. To this end, we investigate how the number of vertices of a given degree and the number of isolated edges change during the process. We prove that $G_{\min}(n, M)$ consists of one giant component with $(1 - o(1))n$ vertices and possibly a number of small isolated cycles when the minimum degree is two. Then, we show that the probability that $G_{\min}(n, tn)$ is connected for $t \neq h_2 \doteq 1.2197$ tends to a certain constant $\rho(t)$. The function $\rho(t)$ is continuous for all $t \neq h_2$.

We study also the phase transition of the minimum degree process. We prove that for a constant $h_{\rm cr} \doteq 0.8607$, a.a.s. $G_{\min}(n, M)$ consists of small components on $O(\log n)$ vertices when $M < h_{\rm cr}n$, and the largest component is of order roughly $n^{2/3}$ when $M = h_{\rm cr}n$, whereas the graph consists of one giant component on $\Theta(n)$ vertices and small components on $O(\log n)$ vertices when $M > h_{\rm cr}n$.

The rest of the chapter is organized as follows. In Sections 12.1 and 12.2 we study the number of vertices of small degrees and the number of isolated edges. In Sections 12.3 and 12.4 we investigate the structure of components when the number of edges is larger than the number of vertices, and study how the probability of the graph being connected changes. In Section 12.6 we model the minimum degree random multigraph process by a multi-type branching process. In Section 12.7 we study the phase transition.

12.1 Vertices of small degree

Let $X_k = X_k(n, M)$ denote the number of vertices of degree k in $G_{\min}(n, M)$, and let $H_k = H_k(n)$ denote the hitting time for the property that $G_{\min}(n, M)$ has minimum degree k, i.e., $H_k = H_k(n)$ is a random variable defined by

$$H_k = \min\left\{M : \delta(G_{\min}(n, M)) \geq k\right\}.$$

Thus, $H_0 \equiv 0$ and $H_k \geq kn/2$. We will investigate the limit behaviour of $X_i(n, M)$, for $i = 0, 1, 2$, and $H_j(n)$, for $j = 1, 2, 3$, when $n \to \infty$.

Let us define constants

$$\begin{aligned} h_1 &= \ln 2 \doteq 0.6931\,, \\ h_2 &= \ln 2 + \ln(1 + \ln 2) \doteq 1.2197\,, \\ h_3 &= \ln((\ln 2)^2 + 2(1 + \ln 2)(1 + \ln(1 + \ln 2))) \doteq 1.7316\,, \end{aligned} \tag{12.1.1}$$

and functions

$$\alpha_0(t) = \begin{cases} 2e^{-t} - 1 & \text{for } 0 \leq t \leq h_1 \\ 0 & \text{for } t \geq h_1\,, \end{cases} \tag{12.1.2}$$

$$\alpha_1(t) = \begin{cases} 2te^{-t} & \text{for } 0 \leq t \leq h_1 \\ e^{h_2 - t} - 1 & \text{for } h_1 \leq t \leq h_2 \\ 0 & \text{for } t \geq h_2\,, \end{cases} \tag{12.1.3}$$

$$\alpha_2(t) = \begin{cases} t^2 e^{-t} & \text{for } 0 \leq t \leq h_1 \\ te^{h_2 - t} - h_1(h_1 + 2)e^{-t} & \text{for } h_1 \leq t \leq h_2 \\ e^{h_3 - t} - 1 & \text{for } h_2 \leq t \leq h_3 \\ 0 & \text{for } t \geq h_3\,. \end{cases} \tag{12.1.4}$$

We will show that, typically, the value of random variable H_k/n is close to h_k and the random variable $X_k(n, M)/n$ can be uniformly approximated by $\alpha_k(M/n)$.

Theorem 12.1.1. *A.a.s. the process $(G_{\min}(n, M))_{M \geq 0}$ is such that for every $M \geq 0$ and $k = 0, 1, 2$,*

$$|X_k(n, M) - n\alpha_k(M/n)| = o(n)\,,$$

and for $k = 1, 2, 3$,

$$|H_k - nh_k| = o(n)\,.$$

For $M \leq (1 - \varepsilon)h_1 n$, $\varepsilon > 0$, this theorem was proved by Wormald [148], with much better estimates for the error terms. He also described how his result can be extended to cover all values of M. Thus, here we only outline his argument.

Proof. We check first that the random variables $X_k = X_k(n, M)$ satisfy the assumptions of Lemma 3.4.5.

Since $X_k(n, M)$ denotes the number of vertices of degree k in the random graph $G_{\min}(n, M)$, we have $|X_k(n, M)| \leq n$. Furthermore, adding an edge can only change the degree of its ends, so

$$|X_k(n, M + 1) - X_k(n, M)| \leq 2\,.$$

Hence, the condition (i) of Lemma 3.4.5 holds.

Now we compute the expected change of the number of vertices of degree k, $k = 0, 1, 2$, when a new edge is added to $G_{\min}(n, M)$.

Suppose first that $M < H_1$. Then, when we add a new edge, we lose an isolated vertex as the first end of the edge and lose another isolated vertex as the second end of the edge with probability $\frac{X_0(n,M)-1}{n-1}$. Hence

$$\mathbb{E}[X_0(n, M+1) - X_0(n, M)|G_{\min}(n, M)] = -1 - \frac{X_0(n, M)}{n} + o(1).$$

Meanwhile, we gain a vertex of degree one as the first end of the new edge and one another if an isolated vertex is chosen as the second end, which happens with probability $\frac{X_0(n,M)-1}{n-1}$.

On the other hand we can destroy a vertex of degree one if it is chosen as the second end of the edge, which can occur with probability $\frac{X_1(n,M)}{n-1}$. Thus

$$\mathbb{E}[X_1(n, M+1) - X_1(n, M)|G_{\min}(n, M)] \\ = 1 + \frac{X_0(n, M)}{n} - \frac{X_1(n, M)}{n} + o(1).$$

Finally, the probability that a vertex of degree one is chosen as the second end of the new edge is $\frac{X_1(n,M)}{n-1}$ and the probability that a vertex of degree two is chosen as the second end of this edge is $\frac{X_2(n,M)}{n-1}$. Consequently,

$$\mathbb{E}[X_2(n, M+1) - X_2(n, M)|G_{\min}(n, M)] = \frac{X_1(n, M)}{n} - \frac{X_2(n, M)}{n} + o(1).$$

If $H_1 \le M < H_2$, then clearly $X_0(n, M) = 0$. When a new edge is added, a vertex of degree one should be chosen as the first end of the edge. As the second end of the edge, a vertex of degree one is chosen with probability $\frac{X_1(n,M)-1}{n-1}$ and a vertex of degree two with probability $\frac{X_2(n,M)}{n-1}$. Thus, we have

$$\mathbb{E}[X_1(n, M+1) - X_1(n, M)|G_{\min}(n, M)] = -1 - \frac{X_1(n, M)}{n} + o(1),$$

and

$$\mathbb{E}[X_2(n, M+1) - X_2(n, M)|G_{\min}(n, M)] = \\ 1 + \frac{X_1(n, M)}{n} - \frac{X_2(n, M)}{n} + o(1)\,.$$

In a similar way, for $H_2 \le M < H_3$, we have $X_0(n, M) = 0$, $X_1(n, M) = 0$, and

$$\mathbb{E}[X_2(n, M+1) - X_2(n, M)|G_{\min}(n, M)] = -1 - \frac{X_2(n, M)}{n} + o(1),$$

while for $M \ge H_3$, $X_k(n, M) = 0$ for $k = 0, 1, 2$.

Thus, let

$$\begin{aligned} f_{0,0}(t,\alpha_0,\alpha_1,\alpha_2) &= -1-\alpha_0,\\ f_{1,0}(t,\alpha_0,\alpha_1,\alpha_2) &= 1+\alpha_0-\alpha_1,\\ f_{2,0}(t,\alpha_0,\alpha_1,\alpha_2) &= \alpha_1-\alpha_2,\\ f_{1,1}(t,\alpha_1,\alpha_2) &= -1-\alpha_1,\\ f_{2,1}(t,\alpha_1,\alpha_2) &= 1+\alpha_1-\alpha_2,\\ f_{2,2}(t,\alpha_2) &= -1-\alpha_2. \end{aligned}$$

Note that all these functions are linear and so they satisfy the Lipschitz condition on, say,

$$D=\{(t,\alpha_0,\alpha_1,\alpha_2): 0\le t\le 3,\ 0\le\alpha_i\le 1,\ \text{ for } i=0,1,2\}\,.$$

Thus, the only obstacle to apply Lemma 3.4.5 is that, say, the value of $\mathbb{E}[X_2(n,M+1)-X_2(n,M)|G_{\min}(n,M)]$ is given by either $f_{2,0}$, $f_{2,1}$, or $f_{2,2}$ depending on the value of M; furthermore, the intervals of M in which we are supposed to use one of these three functions are determined by the values of random variables H_i.

However, there is a simple way to overcome this problem. Let $m(n)=(1-\varepsilon)h_1 n$, where $\varepsilon>0$ is a small constant. Then, Lemma 3.4.5 implies that for $M\le m(n)$ random variables $X_k(n,M)/n$, $k=0,1,2$, are well approximated by the solutions of the differential equations

$$\frac{d\alpha_k}{dt}=f_{k,0}(t,\alpha_0,\alpha_1,\alpha_2),\qquad k=0,1,2,$$

with initial conditions $\alpha_0(0)=1$, $\alpha_1(0)=0$, and $\alpha_2(0)=0$. It is easy to check that the solutions of these equations are given by (12.1.2)–(12.1.4), where $t\in(0,(1-\varepsilon)h_1)$. Furthermore, one can easily verify that $\alpha_0(t)>0$ for $t\in(0,(1-\varepsilon)h_1)$, so a.a.s. $H_1\ge(1-o(1))h_1 n$.

Note also that each edge added to $G_{\min}(n,M)$ for $M<H_1$, decreases the number $X_0(n,M)$ of isolated vertices by at least one and at most two. Hence, for every $M<H_1$

$$X_0(n,M)/2\le H_1-M\le X_0(n,M)\,,$$

and so a.a.s. for any $\varepsilon>0$

$$\begin{aligned} \frac{\alpha_0((1-\varepsilon)h_1)}{2}-o(1)\le\frac{X_0(n,(1-\varepsilon)h_1 n)}{2n}&\le\frac{H_1}{n}-(1-\varepsilon)h_1\\ &\le\frac{X_0(n,(1-\varepsilon)h_1 n)}{n}\le\alpha_0((1-\varepsilon)h_1)+o(1)\,. \end{aligned}$$

Thus, since $\lim_{t\to h_1^-}\alpha_0(t)=0$, a.a.s. $H_1=(1+o(1))h_1 n$. Note also that, since adding an edge can only affect degrees of two vertices of $G_{\min}(n,M)$ and $H_1=(1+o(1))h_1 n$, a.a.s. we have $X_1(n,H_1)=(1+o(1))\alpha_1(h_1)n$ and $X_2(n,H_1)=(1+o(1))\alpha_2(h_1)n$.

Now let us consider the behaviour of $X_k(n,M)$, $k=1,2$, for $(1+\varepsilon)h_1 n\le M\le(1-\varepsilon)h_2 n$. Note that we can consider this part of the process as a 'new' process $G'_{\min}(n,M)$ which starts with a graph $G_{\min}(n,H_1)$ which, as we know

from the part of theorem we have just proved, a.a.s. contains roughly $\alpha_k(h_1)n$ vertices of degree k, for $k = 1, 2$. Then, using Lemma 3.4.5, one can approximate uniformly the random variables $X_k(n, M)/n$ with the solutions of the equations

$$\frac{d\alpha_k}{dt} = f_{k,1}(t, \alpha_1, \alpha_2), \qquad k = 1, 2,$$

in the interval $t \in (h_1, (1-\varepsilon)h_2)$, with boundary conditions

$$\alpha_1(h_1) = \lim_{t \to h_1^-} \alpha_1(t) = \ln 2$$

and

$$\alpha_2(h_1) = \lim_{t \to h_1^-} \alpha_2(t) = (\ln 2)^2/2\,.$$

The solutions of the above system for $t \in (h_1, (1-\varepsilon)h_2)$ are given by (12.1.3), (12.1.4). Then, again, we can approximate the value of H_2 by $h_2 n$, where h_2 is the minimum $t > 0$ for which $\alpha_1(t) = 0$, and find an approximation of $X_2(n, M)/n$ in the interval (h_2, h_3) solving the differential equation

$$\frac{d\alpha_2}{dt} = f_{2,2}(t, \alpha_2)$$

with boundary condition $\alpha_2(h_2) = \lim_{t \to h_2^-} \alpha_2(t)$. Finally, solving the equation $\alpha_2(t) = 0$ one can find h_3 so that a.a.s. $H_3 = (1 + o(1))h_3 n$. □

12.2 Number of isolated edges

Let $Y(n, M)$ count the number of isolated edges in $G_{\min}(n, M)$. Then, one can use Lemma 3.4.5 to prove the following result.

Theorem 12.2.1. *A.a.s. for every $M \geq 0$*

$$\big|Y(n, M) - n\beta(M/n)\big| = o(n)\,,$$

where

$$\beta(t) = \begin{cases} 2e^{-t} - \frac{3}{2}e^{-2t} - \frac{1}{2} & \text{for } 0 \leq t \leq h_1 \\ \frac{1}{8h_1^2}(e^{h_2 - t} - 1)^2 & \text{for } h_1 \leq t \leq h_2 \\ 0 & \text{for } t \geq h_2\,. \end{cases} \tag{12.2.1}$$

Proof. The argument is analogous to the one we used in the proof of Theorem 12.1.1, so we only sketch it here. For $0 \leq M < H_1$, when an edge is added, an isolated vertex is chosen as the first end of the new edge. If another isolated vertex is chosen as the second end, then we gain an isolated edge. This case happens with probability $\frac{X_0(n,M)-1}{n-1}$. However, if a vertex of an isolated edge is chosen as the second end, which happens with probability $\frac{2Y(n,M)}{n-1}$, the number of isolated edges will decrease by one. Hence

$$\mathbb{E}[Y(n, M+1) - Y(n, M) | G_{\min}(n, M)] = \frac{X_0(n, M)}{n} - \frac{2Y(n, M)}{n} + o(1)\,.$$

For $H_1 \leq M < H_2$, when an edge is added, a vertex of degree one is chosen as the first end of the new edge. If this vertex belongs to an isolated edge,

which happens with probability $\frac{2Y(n,M)}{X_1(n,M)}$, we lose a path on two vertices. If a vertex of another path on two vertices is chosen as the other end of the edge, which can occur with probability $\frac{2Y(n,M)-2}{n-1}$, one more isolated edge is destroyed. Consequently,

$$\mathbb{E}[Y(n,M+1)-Y(n,M)|G_{\min}(n,M)] = -\frac{2Y(n,M)}{X_1(n,M)} - \frac{2Y(n,M)}{n} + o(1)\,.$$

Finally, for $M \geq H_2$, clearly $Y(n,M) \equiv 0$.

Using Theorem 12.1.1 to approximate $X_k(n,tn)/n$ by $\alpha_k(t)$ for $k = 0, 1$, we arrive at

$$\frac{d\beta}{dt} = \begin{cases} 2e^{-t} - 1 - 2\beta\,, & \text{for } t \in (0,h_1) \\ -\frac{2\beta}{2(1+\ln 2)e^{-t}-1} - 2\beta\,, & \text{for } t \in (h_1,h_2), \end{cases}$$

with the boundary condition $\beta(0) = 0$. One can verify that the solution to the above system is given by (12.2.1) and so, arguing as in the proof of Theorem 12.1.1, one can use Lemma 3.4.5 to infer that $Y(n,tn)$ can be well approximated by $n\beta(t)$ in the intervals $t \in (0,h_1)$ and $t \in (h_1,h_2)$. □

12.3 Component structure

In this section we will show that when the number of edges in $G_{\min}(n,M)$ is larger than the number of vertices, a.a.s. there is only one largest component with more than a half of the vertices, and all other possible components are small, i.e., of at most logarithmic order, and are unicyclic.

Theorem 12.3.1. *Let $\delta > 0$. Then with probability $1-O(1/n)$ each component of $G_{\min}(n,(1+\delta)n)$ smaller than $n/2$ has at most $(2/\delta)\log n$ vertices and contains at most one cycle. Moreover, for every function $\omega = \omega(n) \to \infty$, a.a.s. the number of all vertices contained in unicyclic components of $G_{\min}(n,(1+\delta)n)$ is smaller than ω.*

Proof. We first note that the probability that, for some k, $(2/\delta)\log n \leq k \leq n/2$, a graph $G_{\min}(n,(1+\delta)n)$ contains a component with k vertices (and hence with at least $k-1$ edges) is bounded from above by

$$\sum_{k=\lceil (2/\delta)\log n\rceil}^{n/2} \sum_{i=k-1}^{(1+\delta)n} \binom{n}{k} \Big(\frac{k-1}{n-1}\Big)^i \Big(\frac{n-k-1}{n-1}\Big)^{(1+\delta)n-i}, \tag{12.3.1}$$

where the first factor counts all possible choices of the vertex set S of a component of k vertices, the second one bounds the probability that every edge with one end in S has the other end in S too, and the last factor is the probability that every vertex outside S has chosen the other end outside S. Note also, that

$$\binom{n}{k} \leq \frac{n^n}{k^k(n-k)^{n-k}}\,,$$

and so, for $1 \leq k \leq n/2$, we have

$$\binom{n}{k}\Big(\frac{k-1}{n-1}\Big)^k\Big(\frac{n-k-1}{n-1}\Big)^{(1+\delta)n-k} \leq e^{-\delta k}\,.$$

Hence, for large enough n, one can bound (12.3.1) from above by

$$\sum_{k=\lceil (2/\delta)\log n\rceil}^{n/2}\ \sum_{s=-1}^{(1+\delta)n-k}\Big(\frac{k-1}{n-k-1}\Big)^s e^{-\delta k} \le (2+\delta)n \sum_{k=\lceil (2/\delta)\log n\rceil}^{n/2} e^{-\delta k} \le \frac{(2+\delta)}{1-e^{-\delta}} n e^{-2\log n} = O\Big(\frac{1}{n}\Big)\,. \tag{12.3.2}$$

Note that any component with at least two cycles has more edges than vertices. Hence the probability that a component of $G_{\min}(n,(1+\delta)n)$ smaller than $\lceil (2/\delta)\log n\rceil$ contains at least two cycles is, for n large enough, bounded by

$$\begin{aligned}&\sum_{k=2}^{(2/\delta)\log n}\ \sum_{s=1}^{(1+\delta)n-k}\binom{n}{k}\Big(\frac{k-1}{n-1}\Big)^{k+s}\Big(\frac{n-k-1}{n-1}\Big)^{(1+\delta)n-k-s}\\ &\le 2\sum_{k=2}^{(2/\delta)\log n}\Big(\frac{k-1}{n-k-1}\Big)e^{-\delta k}\\ &\le \frac{4}{n}\sum_{k=2}^{\infty} k e^{-\delta k} \le \frac{4}{n}\frac{e^{-\delta}}{(1-e^{-\delta})^2}\,.\end{aligned} \tag{12.3.3}$$

Hence, from (12.3.2) and (12.3.3) it follows that the probability that a component of $G_{\min}(n,(1+\delta)n)$ smaller than $n/2$ contains at least two cycles is $O(1/n)$.

Finally, let $\omega=\omega(n)\to\infty$, and let U_k be the number of unicyclic components of k vertices in $G_{\min}(n,(1+\delta)n)$. Note that a unicyclic component has the same number of vertices and edges and thus

$$\begin{aligned}\sum_{k=2}^{n} k\mathbb{E}U_k &\le \sum_{k=2}^{n} k\binom{n}{k}\Big(\frac{k-1}{n-1}\Big)^{k}\Big(\frac{n-k-1}{n-1}\Big)^{(1+\delta)n-k}\\ &\le \sum_{k=2}^{\infty} k e^{-\delta k} \le \frac{e^{-\delta}}{(1-e^{-\delta})^2}\,.\end{aligned}$$

Thus, the probability that at least ω vertices of $G(n,(1+\delta)n)$ belong to unicyclic components is, by Markov's inequality, bounded above by $O(1/\omega)=o(1)$. □

Theorems 12.2.1 and 12.3.1 imply that the giant component must have appeared when the number of edges is between $h_1 n$ and n with $h_1 \doteq 0.6931$. We will show in Section 12.7 that it happens, in fact, when the number of edges becomes $h_{\rm cr} n$ with $h_{\rm cr} \doteq 0.8607$.

12.4 Connectedness

In this section we study how the probability that $G_{\min}(n,M)$ is connected changes as M grows. The main result of this section determines this probability quite precisely for most of the stages of the process $G_{\min}(n,M)$.

Theorem 12.4.1. *Let constants h_2, h_3 be defined as in (12.1.1) and let $\rho_n(t)$ denote the probability that $G_{\min}(n,tn)$ is connected. Then, for every constant $t \neq h_2$, the limit*

$$\rho(t) = \lim_{n\to\infty} \rho_n(t)$$

exists and $\rho(t) = 0$ for $t < h_2$ while $\rho(t) = 1$ for $t \geq h_3$. If $t \in (h_2, h_3)$, then $0 < \rho(t) < 1$, where

$$\rho^+ = \lim_{t\to h_2+} \rho(t) > 0 \quad \text{and} \quad \lim_{t\to h_3-} \rho(t) = 1\,.$$

Proof of Theorem 12.4.1. For each $t < h_2$, Theorem 12.2.1 implies $G_{\min}(n,tn)$ a.a.s. contains many isolated edges and so it is a.a.s. disconnected; hence $\rho(t) = 0$ for $t < h_2$.

If $t > h_2$ then, by Theorems 12.1.1 and 12.3.1, a.a.s. $G_{\min}(n,tn)$ consists of one large component and, perhaps, some short isolated cycles.

In particular, if $t > h_3$, then Theorem 12.1.1 implies a.a.s. $\delta(G_{\min}(n,tn)) \geq 3$ and thus $G_{\min}(n,tn)$ contains no isolated cycles. Consequently, for $t > h_3$ a.a.s. $G_{\min}(n,tn)$ is connected and $\rho(t) = 1$.

It is enough to consider the case $t \in (h_2, h_3]$. Let us fix $t \in (h_2, h_3)$. Note that from Theorem 12.1.1 it follows that for some function $\omega = \omega(n) \to \infty$ with probability $1 - O(\omega^{-2})$ for the process $\{G_{\min}(n,M)\}$ the following holds:

(i) $|H_2 - h_2 n| \leq n/\omega^3$;

(ii) for $M = H_2$, we have $|X_2(n,M) - \alpha_2(h_2)n| \leq n/\omega^3$;

(iii) $|X_2(n,tn) - \alpha_2(t)n| \leq n/\omega^3$.

In our further argument we shall often condition on the event B that (i)–(iii) hold for $(G_{\min}(n,M))_{M\geq 0}$. Note that, since $\Pr(B) = 1 - O(\omega^{-2})$, for any event A we have

$$\Pr(A|B) = \frac{\Pr(A) - O(\omega^{-2})}{1 - O(\omega^{-2})} = O(\omega^{-2}) + (1 - O(\omega^{-2}))\Pr(A). \tag{12.4.1}$$

Let $Z_k = Z_k(n,tn)$, $k = 2,3,\ldots$, denote the number of isolated cycles of length k in $G_{\min}(n,tn)$. We first estimate the expectation of Z_k in the conditional probability space, when we condition on B. In $G_{\min}(n,tn)$ there exist $\binom{n}{k}$ candidates for the set of vertices of an isolated cycle of length k. Let us fix one such subset S. Note that if $G_{\min}(n,tn)$ contains an isolated cycle with vertex set S, then all of its edges appear in $G_{\min}(n,M)$ already at the moment H_2, when the minimum degree of a graph reaches two. If at this moment a cycle is isolated, then each time we chose the first end of an edge outside S we had to pick as the second end of an edge a vertex outside S as well. By (12.4.1) the probability of that event is given by

$$\begin{aligned} &O(\omega^{-2}) + (1 + O(\omega^{-2}))\Big(\frac{n-k-1}{n-1}\Big)^{(1+O(\omega^{-3}))h_2 n - k} \\ &\qquad = O(\omega^{-2}) + (1 + O(k\omega^{-3} + k^2/n))e^{-h_2 k}. \end{aligned} \tag{12.4.2}$$

If $G_{\min}(n,M)$ contains an isolated cycle on vertex set S, it means that until this moment each time we have picked up one end of an edge in S the second

end has been chosen also in S, in such a way that it created with edges which had already been selected a forest which consisted of paths, and, eventually, a cycle of length k. Thus, the probability that in $G(n, H_2)$ the subset S spans a cycle is equal to

$$\Big(\frac{k-1}{n-1}\Big)^k \Pr(G_{\min}(k,k) \text{ is a cycle}). \tag{12.4.3}$$

Hence the probability that, conditioned on B, there exists an isolated cycle on S in $G(n, H_2)$ is given by a product of (12.4.2) and (12.4.3).

The probability that a cycle on the set S which is isolated at the moment $M = H_2$ remains isolated also in $G_{\min}(n, tn)$ is the probability that each vertex of the cycle has degree two also in $G_{\min}(n, tn)$. It is easy to see that if by W_2 and $W_2(t)$ we denote the sets of vertices of degree two in $G_{\min}(n, M)$ at the moments $M = H_2$ and $M = tn$ respectively, then each subset of W_2 of $|W_2(t)|$ elements is equally likely to become $W_2(t)$ later in the process. Hence the probability that $W_2(t)$ contains a given subset of $S \subseteq W_2$ of k elements is equal to

$$\frac{\binom{|W_2|-k}{|W_2(t)|-k}}{\binom{|W_2|}{|W_2(t)|}} = \Big(\frac{|W_2(t)|}{|W_2|}\Big)^k \Big(1 + O\Big(\frac{k^2}{|W_2|}\Big)\Big).$$

Consequently, the probability that, conditioned on B, an isolated cycle on S present in $G_{\min}(n, M)$ for $M = H_2$ remains isolated also in $G_{\min}(n, tn)$ is given by

$$O(\omega^{-2}) + (1 + O(\omega^{-2} + k^2/n))\Big(\frac{\alpha_2(t)}{\alpha_2(h_2)}\Big)^k.$$

Thus,

$$\begin{aligned}
\mathbb{E}Z_k &= O(\omega^{-2}) + (1 + O(\omega^{-2} + k\omega^{-3} + k^2/n))\binom{n}{k}\Big(\frac{k-1}{n-1}\Big)^k \\
&\qquad \times e^{-h_2 k}\Big(\frac{\alpha_2(t)}{\alpha_2(h_2)}\Big)^k \Pr(G_{\min}(k,k) \text{ is a cycle}) \\
&= O(\omega^{-2}) + (1 + O(\omega^{-2} + k\omega^{-3} + k^2/n))\frac{(k-1)^k}{k!} \\
&\qquad \times \Big(\frac{e^{h_3 - t} - 1}{e^{h_3} - e^{h_2}}\Big)^k \Pr(G_{\min}(k,k) \text{ is a cycle}).
\end{aligned} \tag{12.4.4}$$

Now let $Z(n, tn) = \sum_{k=2}^{\omega} Z_k(n, tn)$. Since a.a.s. $G_{\min}(n, tn)$ contains no cycles longer than ω, a.a.s. $Z(n, tn)$ is equal to the number of all isolated cycles of $G_{\min}(n, tn)$. From (12.4.4) we infer that

$$\mathbb{E}Z(n, tn) = (1 + o(1))\lambda(t), \tag{12.4.5}$$

where

$$\lambda(t) = \sum_{k=2}^{\infty} \frac{(k-1)^k}{k!}\Big(\frac{e^{h_3 - t} - 1}{e^{h_3} - e^{h_2}}\Big)^k \Pr(G_{\min}(k,k) \text{ is a cycle }). \tag{12.4.6}$$

Furthermore, for the ith factorial moment of $Z(n,tn)$, $i=2,3,\dots$, we get

$$\begin{aligned}\mathbb{E}_i Z(t,tn) &= \mathbb{E}Z(t,tn)(Z(n,tn)-1)\cdots(Z(n,tn)-i+1)\\ &= \mathbb{E}\sum_{k_1=2}^{\omega} Z_{k_1}(n,tn)\mathbb{E}\sum_{k_2=2}^{\omega} Z_{k_2}(n-k_1,tn-k_1)\\ &\qquad \cdots \mathbb{E}\sum_{k_i=2}^{\omega} Z_{k_i}(n-\sum_{s=1}^{i-1}k_s, tn-\sum_{s=1}^{i-1}k_s) \qquad (12.4.7)\\ &= (1+o(1))\left[\mathbb{E}\sum_{k=2}^{\omega} Z_k(n,tn)\right]^i = (1+o(1))\lambda^i(t)\,.\end{aligned}$$

Hence, $Z(t,nt)$ has asymptotically Poisson distribution $Po(\lambda(t))$, i.e., for every $s=0,1,\dots$,

$$\Pr(Z(n,tn)=s) = (1+o(1))\frac{\lambda^s(t)}{s!}e^{-\lambda(t)}\,.$$

In particular, Theorem 12.3.1 implies that

$$\begin{aligned}\Pr(G_{\min}(n,tn) \text{ is connected}) &= (1+o(1))\Pr(Z(n,tn)=0)\\ &= (1+o(1))e^{-\lambda(t)}\,,\end{aligned} \qquad (12.4.8)$$

where $\lambda(t)$ is given by (12.4.6).

Observe that if $t\to h_2$ then $\lambda(t)\to\lambda(h_2)>0$. Hence

$$\lim_{t\to h_2^+}\rho(t) = e^{-\lambda(h_2)} > 0\,.$$

Note also that $\lambda(t)\to 0$ for $t\to h_3$, so $\lim_{t\to h_3^-}\rho(t)=1$. Finally, since for every n the probability that $G_{\min}(n,nt)$ is connected is a non-decreasing function of t, we have $\rho(h_3)=1$. □

Observe that the limit behaviour of the minimum degree multigraph process $(G_{\min}(n,M))_M$ is very different from the classical random graph process $(G(n,M))_M$ mentioned in Section 3.1, in which a.a.s. $G(n,M)$ becomes connected for $M=\frac{n}{2}(\log n+\omega(n))$ with $\omega(n)\to\infty$. However, it is worthwhile to compare Theorem 12.4.1 with analogous results for two other random graph process models in which the minimum degree grows quickly with the number of edges.

The first one is the uniform graph process $(U(n,M))_M$, in which the Mth edge of $U(n,M)$ has one end at vertex $M-\lfloor M/n\rfloor$ while its other end is chosen uniformly at random from all $n-1$ possibilities (the vertex set of $U(n,M)$ is $\{0,1,\dots,n-1\}$). Jaworski and Łuczak [83] proved that for every $t\ge 0$ the probability that $U(n,tn)$ is connected tends to a limit $f(t)$ as $n\to\infty$, where $f(t)$ is an explicit continuous function such that $f(t)=0$ for $t\le 1$, $f(t)=1$ for $t\ge 2$ and for $1<t<2$ we have $0<f(t)<1$.

The second random graph process is the random d-process $(G^d(n,M))_{M=0}^{\lfloor dn/2\rfloor}$. In this process $G^d(n,M)$ is obtained from $G^d(n,M-1)$ by adding to it an edge chosen uniformly at random among all pairs of vertices e such that the graph $G^d(n,M-1)\cup e$ has the maximum degree at most d. It was shown by Ruciński and Wormald [124] that if $0\le t<d/2$ then a.a.s. $G^d(n,tn)$ is disconnected.

On the other hand, they proved that the final stage of the process, the graph $G(n, \lfloor nd/2 \rfloor)$, is a.a.s. connected, provided $d \geq 3$. Hence, the limit probability that $G^d(n, tn)$ is connected has a 'degenerate' discontinuity at $t = d/2$, where it jumps from 0 to 1. We also remark that for $t \in (h_2, h_3)$ the structure of $G_{\min}(n, tn)$ is somewhat similar to that of the random graph considered by Karoński and Pittel in [90].

12.5 Two phases

In this section we study $G_{\min}(n, M)$ through two phases, as it will turn out to be useful when we study the phase transition by approximating the graph process by a branching process in Section 12.7. The first phase is when the minimum degree is zero, and the other is when the minimum degree is one. We will represent the phases by colouring the edges in the following way: When an edge is chosen to be added, we colour it *red* if the minimum degree of the graph (before the addition of the edge) is zero, and *blue* if the minimum degree is one. Other edges are uncoloured. In this chapter we will only consider the stages of the process where all the edges are *a.a.s.* either red or blue, namely when $t < h_2$. We let the *red phase* be the part of the process where the graph still contains isolated vertices. The *blue phase* is the phase where the minimum degree is one. In the red phase $G_{\min}(n, M)$ is a red forest; in the blue phase $G_{\min}(n, M)$ is a union of a red and a blue forest.

Red trees. We will first determine how many red trees there are of different order in $G_{\min}(n, tn)$, by using the differential equation method, due to Wormald (see Lemma 3.4.5).

Let $R_k(n, M)$ be the number of components of order k in $G_{\min}(n, M)$, when we are still in the red phase. Note that all components are trees in this phase. We will say that a component or tree is *trivial* if it contains only a single vertex, and thus no edges. First we will show that there are no components of larger than logarithmic order.

Lemma 12.5.1. *The largest red tree in* $G_{\min}(n, H_1)$ *has* a.a.s. $O(\log n)$ *vertices.*

Proof. We have to prove that there is a positive constant c, such that there *a.a.s.* is no red tree of order $c \log n$ or greater.

When an edge (v, w) is added to the graph, we can think of it the way that we first choose a vertex v of minimum degree, and then let v choose the vertex w randomly. Then v is the *choosing* vertex, while w is the *chosen* vertex. Consider the graph $G_{\min}(n, H_1)$, which is the state of the process at the precise moment when the minimum degree becomes 1, or in other words, at the end of the red phase. Set $k = \lceil c \log n \rceil - 1$, and suppose that there is a component of order at least $k + 1$. Let E be the set of edges in $G_{\min}(n, H_1)$. Then there is a set of edges, $E' = \{e_1, \ldots, e_k\} \subset E$, with the following property:

For $i = 1, \ldots, k$, let $e_i = (v_i, w_i)$, where v_i is the choosing vertex of the edge. Then for every $i = 2, \ldots, k$, $w_i \in \{v_1, v_2, \ldots, v_{i-1}, w_1\}$.

Let $E'' \subset E$ be any subset of E with $|E''| = k$. The probability that E'' satisfies the above property is at most $(k-1)!/(n-1)^{k-2}$. Since $|E| = nh_1 + o(n)$ *a.a.s.*, there are about $\binom{nh_1}{k}$ ways to choose a set of k edges from

E. Hence the probability that there is a set E' as described above, is bounded from above by

$$\binom{0.7n}{k}\frac{k!}{(n-1)^{k-2}} \le \frac{(0.7n)^k}{k!}\frac{k!}{(n-1)^{k-2}} = n^2 0.7^{k-2},$$

which tends to 0 for sufficiently large c. □

Lemma 12.5.2. *If $t < h_1$, then the number of components with exactly k vertices in $G_{\min}(n, tn)$ is* a.a.s.

$$\frac{1}{k}(1-e^{-t})^{k-1}((k+1)e^{-t}-1)n + o(n). \tag{12.5.1}$$

Proof. Let $R_k(n, M)$ be the number of components with k vertices in $G_{\min}(n, tn)$ for $k \ge 0$. In particular, $R_0(n, M) = 0$ and $R_1(n, M)$ is the number of isolated vertices. By Lemma 12.5.1, all components have order $O(\log n)$ *a.a.s.*, so we only have to consider $R_k(n, M)$ with $k = O(\log n)$. Then we can use Lemma 3.4.5 to determine the asymptotic values for $R_k(n, M)$.

We now find an expression for the expected amount of change in $R_k(n, M)$ through the addition of a single edge to the graph, and then use Lemma 3.4.5 to find functions $\rho_k(t)$ such that *a.a.s.*, $R_k(n, tn) = \rho_k(t)n + o(n)$ for $t < h_1$. It is clear that

$$|R_k(n, M)| \le n, \quad \text{and} \quad |R_k(n, M+1) - R_k(n, M)| \le 2.$$

When an edge is added to the graph, the first end of the edge is in a component of order 1, and we therefore always lose one such component. If the other end is in a component of order k, we lose one component of order k, and if it is in a component of order $k-1$, we gain one component of order k. The probabilities of these two events are $\frac{kR_k(n,M)}{n-1}$ and $\frac{(k-1)R_{k-1}(n,M)}{n-1}$, respectively. Hence for $k \ge 1$,

$$\begin{aligned}&\mathbb{E}\left[R_k(n, M+1) - R_k(n, M) \mid G_{\min}(n, M)\right]\\ &= -\delta_{1k} - \frac{kR_k(n, M)}{n} + \frac{(k-1)R_{k-1}(n, M)}{n} + o(1),\end{aligned} \tag{12.5.2}$$

where δ_{ij} is the Kronecker delta.

By Lemma 3.4.5 the functions $\rho_k(t)$ satisfy the differential equation

$$\frac{d}{dt}\rho_k(t) = -\delta_{1k} - k\rho_k(t) + (k-1)\rho_{k-1}(t), \tag{12.5.3}$$

with $\rho_0(t) = 0$ for all t. Since this equation is linear, it satisfies a Lipschitz condition in a suitable domain D. Since all components consist of one vertex when $M = 0$, we have the boundary conditions $\rho_1(0) = 1$ and $\rho_k(0) = 0$ for $k \ge 2$. Solving the differential equation (12.5.3) and using the boundary condition, we get

$$\begin{aligned}\rho_1(t) &= 2e^{-t} - 1\,,\\ \rho_k(t) &= \frac{1}{k}\left(1-e^{-t}\right)^{k-1}\left((k+1)\,e^{-t} - 1\right),\end{aligned}$$

which satisfies the condition 3 in Lemma 3.4.5 in a suitable domain D. □

Corollary 12.5.3. *When the red phase is finished, the number of red trees with exactly k vertices is* a.a.s.

$$\frac{k-1}{k2^k}n + o(n). \tag{12.5.4}$$

Proof. When $t > h_1$, there are *a.a.s.* no more isolated vertices in the graph. We therefore get the number of red trees with k vertices by setting $t = h_1 = \log 2$ in (12.5.1). □

For a vertex v we let $C_{red}(v)$ be the red tree containing v. From Corollary 12.5.3 we get

$$\Pr\left[|C_{red}(v)| = k\right] = \frac{k-1}{2^k} + o(1). \tag{12.5.5}$$

We say that a tree is a (k,p)-tree if it consists of k vertices, exactly p of which are leaves, and we let $e_{k,p}$ be the probability that a red tree on k vertices contains exactly p leaves. Note that leaves correspond to light vertices, and non-leaves to heavy vertices. When $t = h_1$, there are *a.a.s.* $n\log 2 + o(n)$ vertices of degree 1. Thus, when $t > h_1$ there are *a.a.s.* $n\log 2 + o(n)$ vertices incident to precisely one red edge. From this and (12.5.5) it follows that

$$\begin{aligned}\Pr\left[C_{red}(v) \text{ is a } (k,p)\text{-tree}\middle|\, \deg_r(v) = 1\right]& \\ = \frac{p(k-1)}{(\log 2)k2^k}e_{k,p} + o(1)\,,& \qquad (12.5.6)\\ \Pr\left[C_{red}(v) \text{ is a } (k,p)\text{-tree}\middle|\, \deg_r(v) > 1\right]& \\ = \frac{(k-p)(k-1)}{(1-\log 2)k2^k}e_{k,p} + o(1),& \qquad (12.5.7)\end{aligned}$$

where $\deg_r(v)$ denotes the number of red edges incident to v and is called the *red degree* of v.

To study the distribution of $e_{k,p}$ we define, for $p \geq 2$,

$$E_p(z) = \sum_{k\geq p} e_{k,p}z^k$$

to be the probability generating function for $e_{k,p}$. Let $E_p'(z) := \frac{dE_p(z)}{dz}$.

Lemma 12.5.4.

$$E_p(z) = -\frac{z}{p(p-1)} + \sum_{i=2}^{p}\frac{(-iz)^{p-i}e^{iz}}{(p-i)!i^2}(i-1)(iz+p-i). \tag{12.5.8}$$

Proof. It is easy to see that $e_{k,1} = 0$ for all $k \geq 2$, $e_{k,k} = 0$ for all $k > 2$ and $e_{2,2} = 1$. Suppose that $k > 2$ and $2 \leq p < k$. A (k,p)-tree can either be constructed from a $(k-1,p)$-tree by attaching v_k to a leaf, or from a $(k-1,p-1)$-tree by attaching v_k to a non-leaf. Hence $e_{k,p}$ satisfies the recursion

$$e_{k,p} = \frac{k-p}{k-1}e_{k-1,p-1} + \frac{p}{k-1}e_{k-1,p}.$$

Then (12.5.8) follows by induction on p. □

Blue trees. Now we will assume that we are somewhere in the blue phase, that is, $h_1 < t < h_2$. We say that a vertex is *light* if it is incident to precisely one red edge, and *heavy* otherwise. Every non-trivial blue tree begins as an edge, and then possibly continues to grow one vertex at a time. When a non-trivial blue tree is first created, at most one of the two vertices in the tree can be heavy. Every subsequent vertex added to the tree must be light. Hence a blue tree cannot contain more than one heavy vertex. We say that a blue tree in $G_{\min}(n, M)$ is *simple* if every vertex in the tree is light, and *non-simple* otherwise. From the above explanation, a non-simple tree must contain precisely one heavy vertex.

If a vertex is not incident to any blue edges, we consider it a blue tree of order 1. If this vertex is light, we consider it a simple tree, and if it is heavy, we consider it a non-simple tree. Hence every vertex is part of both a red and a blue tree.

We will now determine how many simple and non-simple blue trees there are in $G_{\min}(n, tn)$. In order to simplify the formulas, we define

$$u = u(t) := 2e^{-t}.$$

Lemma 12.5.5. *The number of simple blue trees with exactly k vertices in $G_{\min}(n, tn)$ is* a.a.s.

$$\frac{1}{k}(1-u)^{k-1}(u + ku\log 2 - 1)n + o(n). \tag{12.5.9}$$

The number of non-simple blue trees with exactly k vertices in $G_{\min}(n, tn)$ is a.a.s.

$$(1 - \log 2)u(1-u)^{k-1}n + o(n). \tag{12.5.10}$$

Proof. Let $S_k(n, M)$ be the number of simple blue trees with exactly k vertices, and $T_k(n, M)$ be the number of non-simple blue trees with exactly k vertices. Lemma 12.5.1 can be adapted to blue trees as well as red, so we can assume that $k = O(\log n)$.

Consider a blue tree of order $k \geq 2$. The probability that this tree grows with one vertex when an edge is added to $G_{\min}(n, M)$, is the same as the probability that a red tree of order k grows with one vertex in the red phase. Hence (12.5.2) holds for blue trees as well, with $R_k(n, M)$ exchanged with $S_k(n, M)$ and $T_k(n, M)$, respectively. Hence $\sigma_k(t)$ and $\tau_k(t)$ both satisfy the recursion (12.5.3). The behaviour of the blue trees deviates from the red trees when $k = 1$. Every edge added causes a simple blue tree of order one to disappear. The expected amount of change of $S_1(n, M)$ and $T_1(n, M)$ are given by the equations

$$\begin{aligned}
\mathbb{E}\left[S_1(n, M+1) - S_1(n, M) | G_{\min}(n, M)\right] &= -1 - \frac{S_1(n, M)}{n} + o(1)\,, \\
\mathbb{E}\left[T_1(n, M+1) - T_1(n, M) | G_{\min}(n, M)\right] &= -\frac{T_1(n, M)}{n} + o(1).
\end{aligned}$$

Hence the differential equations

$$\frac{d}{dt}\sigma_1(t) = -1 - \sigma_1(t) \quad \text{and} \quad \frac{d}{dt}\tau_1(t) = -\tau_1(t), \tag{12.5.11}$$

are satisfied. As mentioned earlier, when $t = h_1 = \log 2$, there are *a.a.s.* $n \log 2 + o(n)$ vertices of degree 1 in $G_{\min}(n, tn)$. Hence we have the boundary conditions $\sigma_1(\log 2) = \log 2$ and $\tau_1(\log 2) = 1 - \log 2$. Furthermore $\sigma_k(\log 2) = \tau_k(\log 2) = 0$ for $k \geq 2$. Solving the differential equation (12.5.11), and using the boundary conditions just given, we get

$$\sigma_1(t) = 2(1 + \log 2)e^{-t} - 1 \quad \text{and} \quad \tau_1(t) = 2(1 - \log 2)e^{-t}.$$

It follows by induction that

$$\begin{aligned} \sigma_k(t) &= \frac{1}{k}(1-u)^{k-1}(u + ku\log 2 - 1) \\ \tau_k(t) &= (1 - \log 2)u(1-u)^{k-1}. \end{aligned}$$

By Lemma 3.4.5, this implies the theorem. □

For each vertex v, we let $C_{blue}(v)$ be the blue tree containing v. Recall that the number of vertices of red degree one when $t > h_1$ is *a.a.s.* $n \log 2 + o(n)$. Then Lemma 12.5.5 implies the following:

$$\Pr\left[|C_{blue}(v)| = k | \deg_r(v) > 1\right] = u(1-u)^{k-1} + o(1), \tag{12.5.12}$$

$$\begin{aligned} &\Pr\left[C_{blue}(v) \text{ is simple and } |C(v)| = k | \deg_r(v) = 1\right] \\ &\qquad = \frac{1}{\log 2}(1-u)^{k-1}(u + ku\log 2 - 1) + o(1), \end{aligned} \tag{12.5.13}$$

$$\begin{aligned} &\Pr\left[C_{blue}(v) \text{ is non-simple and } |C(v)| = k | \deg_r(v) = 1\right] \\ &\qquad = \left(\frac{1}{\log 2} - 1\right)(k-1)u(1-u)^{k-1} + o(1). \end{aligned} \tag{12.5.14}$$

12.6 Branching process.

In this section we model $G_{\min}(n, M)$ by a multi-type branching process (see Section 3.4). In a usual branching process, in particular the one for a standard random graph process $G(n, M)$, starting with a vertex v we build up the component containing v starting first with the immediate neighbourhood of v, and then continuing as new vertices are added to the component. Note however that in $G_{\min}(n, M)$, given two incident edges, the probability that one of them is in $G_{\min}(n, M)$ is not independent of the other. We will overcome this problem by looking at $G_{\min}(n, tn)$ through two phases and also by distinguishing vertices between *light* and *heavy* vertices.

Suppose that we are in the blue phase, and consider a vertex v. This vertex is incident to at least one red edge, and possibly one or more blue edges. If it is not incident to any blue edges, we consider it a blue tree of order 1. Thus it is part of a red tree, T_r, of order at least 2, and a blue tree, T_b, of order at least 1. As in Section 12.5 we call v a *light* vertex if it is incident to precisely one red edge, and a *heavy* vertex if it is incident to more than one red edge.

The crucial observation is that if we are given the information about whether v is light or heavy conditioned on the structure of the red forest, then the order of T_r and the order of T_b are two random variables which are independent of each other. The reason for this is that, when we add a new blue edge, the way

we choose the two vertices which will be joined does not depend on the order of the component which they are part of, but only on the degrees of the vertices.

In the branching process we build up the component containing v by alternatingly adding red and blue trees. As t grows, the branching process will produce larger components, and the value h_{cr} given in Theorem 12.7.1 corresponds to the critical point: When $t \leq h_{\mathrm{cr}}$, the branching process dies out after a finite number of steps with probability 1, while when $t > h_{\mathrm{cr}}$ it continues forever with probability strictly greater than 0. This corresponds to the existence of a giant component in $G_{\min}(n, tn)$.

The branching process we will use has four different types. A vertex is either light or heavy, and it is either an *r-vertex* or a *b-vertex*. Furthermore, we will distinguish between *saturated* and *unsaturated* vertices. (These types are defined below.)

Let $\mathcal{A}$ be the event that the number of red trees in $G_{\min}(n, tn)$ is given by (12.5.4), that the number of simple blue trees is given by (12.5.9), and that the number of non-simple blue trees is given by (12.5.10). According to Corollary 12.5.3 and Lemma 12.5.5, $\mathcal{A}$ holds with probability tending to 1. Thus, if we want to show that some event holds *a.a.s.*, it is sufficient to show that it holds *a.a.s.*, when conditioned on $\mathcal{A}$. From now on we therefore assume that the event $\mathcal{A}$ holds.

In the first step of the branching process a red tree is created. The order of the tree is given by the probability distribution (12.5.5). All the vertices created in this step are *unsaturated r-vertices*. If the tree has order k, we let p of the vertices be light and $k - p$ heavy, with probability $e_{k,p}$.

In subsequent steps, the branching process evolves as follows: We choose an unsaturated vertex v at random. If v is an r-vertex, then we create a blue tree incident to v. If v is heavy, then the order of the tree is chosen randomly with probabilities given by (12.5.12). If v is light, then the order of the tree, and whether it is simple or non-simple, is determined according to the probabilities in (12.5.13) and (12.5.14). All the vertices created are *unsaturated b-vertices*. Note that it is possible that the tree created has order 1. In this case no new vertices are created. If v is a heavy vertex, then all the newly generated vertices in the blue tree are light. If v is light, and the blue tree is simple, all the new vertices are also light, and if the blue tree is non-simple, exactly one of the new vertices is heavy, while the others are light.

If v is a b-vertex, then we create a red tree incident to v. The probability that the red tree has order k, and contains exactly p light vertices, is given by the probability distribution (12.5.6) (resp. (12.5.7)) if v is a light (resp. heavy) vertex. All the newly created vertices are unsaturated r-vertices. After the new red or blue tree has been created, we end the step by marking v as *saturated*.

Thus a vertex is an r-vertex if it was generated through the creation of a red tree, and a b-vertex otherwise.

This branching process approximates the way we might proceed if we want to find all the vertices in the component in $G_{\min}(n, M)$ which contains a given vertex v. First we find the red tree containing v. Then at every vertex w of this red tree, we find the blue tree which contains w. Note that the order of the blue tree does not depend on the order of the red tree of which w is a part, only on the information about whether w is incident to one or more than one red edge. Then we continue exploring alternatingly red and blue trees, until it stops. Some care must be taken, because the graph process will generally

contain cycles, which the branching process does not. We will consider this problem more closely in Section 12.7.

We will now calculate the generating functions for the number of vertices created in one step of the branching process. By one step it is meant that from a given vertex we create a (red or blue) tree and then count the number of vertices which have been created. In subscripts we will often use the letters r, R, b and B. In general r and b refer to light r- and b-vertices, whereas R and B refer to heavy r- and b-vertices.

Using (12.5.12),(12.5.13) and (12.5.14) we get

$$\begin{aligned}
&f_r(x,y)\\
&=\sum_{i,j}\Pr\left[\text{a light } r\text{-vertex generates } i \text{ light and } j \text{ heavy } b\text{-vertices}\right]x^iy^j\\
&=\sum_k\Bigg(\Pr\left[C_{blue}(v) \text{ is simple and } |C(v)|=k+1\right|\deg_r(v)=1]\,x^k\\
&+\Pr\left[C_{blue}(v) \text{ is not simple and } |C(v)|=k+2\right|\deg_r(v)=1]\,x^ky\Bigg)\\
&=\frac{(y\log 2-y+x)u^2+((1-\log 2)y+1+\log 2-2x)u-1+x}{(-xu+x-1)^2\log 2},
\end{aligned}$$

$$\begin{aligned}
&f_R(x,y)\\
&=\sum_{i,j}\Pr\left[\text{a heavy } r\text{-vertex generates } i \text{ light and } j \text{ heavy } b\text{-vertices}\right]x^iy^j\\
&=\sum_k\Pr\left[\,|C_{blue}(v)|=k+1\right|\deg_r(v)>1]\\
&=\frac{u}{1-(1-u)x}.
\end{aligned}$$

In order to shorten the formulas and make them more readable, we set

$$\alpha=(x-2)(x-y),\quad \beta=(y-2)(y-x),\quad \text{and}\quad \gamma=\frac{x}{y}e^{\frac{y-x}{2}}.$$

We will occasionally write $z=\frac{y}{2}$. Using (12.5.6), (12.5.7) and (12.5.8), we get

$$\begin{aligned}
&f_b(x,y)\\
&=\sum_{i,j}\Pr\left[\text{a light } b\text{-vertex generates } i \text{ light and } j \text{ heavy } r\text{-vertices}\right]x^iy^j\\
&=\sum_{k\geq 2}\sum_{p=2}^{k}\Pr\left[C_{red}(v) \text{ is a } (k,p)\text{-tree}\right|\deg_r(v)=1]\,x^{p-1}y^{k-p}\\
&=\frac{1}{x\log 2}\left(\frac{x}{y-x}+\frac{-4\gamma+4\gamma^2+\alpha\gamma^2}{4(1-\gamma)^2}+\log\frac{y-x}{y(1-\gamma)}\right),
\end{aligned}$$

$$
\begin{aligned}
&f_B(x,y)\\
&= \sum_{i,j} \Pr\left[\text{a heavy } b\text{-vertex generates } i \text{ light and } j \text{ heavy } r\text{-vertices}\right] x^i y^j\\
&= \sum_{k\geq 2}\sum_{p=2}^{k} \Pr\left[C_{red}(v) \text{ is a } (k,p)\text{-tree} \middle| \deg_r(v) > 1\right] x^p y^{k-p-1}\\
&= \frac{1}{y(1-\log 2)}\left(\log\frac{y}{y-x} + \frac{x}{x-y} + \frac{\beta\gamma^2 - 4\gamma^2 + 4\gamma}{4(1-\gamma)^2} + \log(1-\gamma)\right).
\end{aligned}
$$

We let $T = T(t)$ be the transition matrix associated with the branching process. It can be written as

$$
T = \begin{bmatrix} 0 & 0 & \phi_{1,3}(t) & \phi_{1,4}(t) \\ 0 & 0 & \phi_{2,3}(t) & \phi_{2,4}(t) \\ \psi_{3,1} & \psi_{3,2} & 0 & 0 \\ \psi_{4,1} & \psi_{4,2} & 0 & 0 \end{bmatrix}.
$$

Here the first row and the first column correspond to light r-vertices; the second row and the second column to heavy r-vertices; the third row and the third column to light b-vertices; and the fourth row and the fourth column to heavy b-vertices. The entries denote the expected number of vertices we get of the required type when we start with a light vertex.

The zeros in the matrix are there, because we will always go from an r-vertex to a b-vertex, and vice versa. The ϕ's describe the expected number of vertices generated by the creation of a blue tree at an r-vertex, so they depend on t. The ψ's, on the other hand, describe how many vertices are generated by the creation of a red tree at a b-vertex, and do not depend on t, since we are after the red phase in the graph process, and no more red edges are added.

The ϕ's can be calculated by differentiating the corresponding generating functions and evaluating them at $(x,y) = (1,1)$. It is slightly more difficult to calculate the ψ's, since $x - y$ occurs in the denominator of the expressions of $f_b(x,y)$ and $f_B(x,y)$. This is, however, a removable pole, and we can calculate the ψ's by taking limits at $(1,1)$.

Thus we have expressions for all the entries in T. Since T is not positive regular, we cannot apply Theorem 3.4.6 to it. Instead we consider the 2×2-submatrix $A = A(t)$ of T^2 consisting of the first and second rows and columns. This matrix is the transition matrix when we consider the branching process in two steps at a time: From an r-vertex we first generate a blue tree, and then we generate a red tree for each of the new vertices we get. We have that

$$
A = \begin{bmatrix} \phi_{1,3}(t)\psi_{3,1} + \phi_{1,4}(t)\psi_{4,1} & \phi_{1,3}(t)\psi_{3,2} + \phi_{1,4}(t)\psi_{4,2} \\ \phi_{2,3}(t)\psi_{3,1} & \phi_{2,3}(t)\psi_{3,2} \end{bmatrix}. \tag{12.6.1}
$$

Let $\lambda_1(t)$ be the largest eigenvalue of A. Then $\lambda_1(t)$ is an increasing and continuous function of t, and we define h_{cr} to be the value for which $\lambda_1(h_{\text{cr}}) = 1$. According to Theorem 3.4.6, the branching process dies out with probability 1 when $t \leq h_{\text{cr}}$ and it continues forever with positive probability when $t > h_{\text{cr}}$.

The largest eigenvalue of A is

$$
u_+ = \frac{3\zeta - \zeta^2 + \sqrt{27\zeta^2 - 16\zeta}}{8\zeta - \zeta^2} = \frac{3\log 2 - 1 + \log 2\sqrt{27 - 16\log 2}}{8\log 2 - 1},
$$

where $\zeta = \frac{1}{\log 2}$. Since $u = 2e^{-t}$, we get

$$h_{\mathrm{cr}} = \log \frac{2}{u_+} = \log \left(\frac{16 \log 2 - 2}{3 \log 2 - 1 + \log 2 \cdot \sqrt{27 - 16 \log 2}} \right) \doteq 0.8607.$$

Thus we have located the critical moment of the branching process, when the largest eigenvalue of the transition matrix becomes 1.

Let $P(t)$ be the extinction probability of the branching process. When $t \leq h_{\mathrm{cr}}$, we have that $P(t) = 1$. Suppose now that $t > h_{\mathrm{cr}}$. We let $g_r(x, y)$ and $g_R(x, y)$ be the generating functions for the number of light and heavy r-vertices generated in two steps (first generating a blue tree and then red trees afterwards) starting with one light r-vertex and one heavy r-vertex, respectively. We have

$$\begin{aligned} g_r(x, y) &= f_r(f_b(x, y), f_B(x, y)), \\ g_R(x, y) &= f_R(f_b(x, y), f_B(x, y)). \end{aligned}$$

Let $q_r(x, y)$ be the generating function counting the following: Take a vertex v, which is a light r-vertex, and start by generating a blue tree, and then continue the branching process until it dies out. Then we count the number of r-vertices throughout the process, including the vertex v. The coefficient of $x^i y^j$ is then the probability that precisely i light r-vertices and j heavy r-vertices were created throughout the process. We define $q_R(x, y)$ similarly (when we start with a heavy vertex).

Suppose we start with a light vertex v. The generating function for the total number of vertices in the process, including v, conditioned on the event that s light and S heavy vertices are created in the first step, is $x q_r(x, y)^s q_R(x, y)^S$. Let

$$\begin{aligned} &p_{s,S} \\ &= \Pr\left[\text{a light } r\text{-vertex generates } s \text{ light and } S \text{ heavy } r\text{-vertices in one step}\right]. \end{aligned}$$

Hence

$$q_r(x, y) = \sum_{s,S} p_{s,S} x q_r(x, y)^s q_R(x, y)^S = x g_r(q_r(x, y), q_R(x, y)).$$

Similarly we have

$$q_R(x, y) = y g_R(q_r(x, y), q_R(x, y)).$$

If y_r and y_R are the extinction probabilities, when starting with a light and heavy r-vertex, respectively, we get

$$\begin{aligned} y_r &= \sum_{s,S} \Pr\left[\text{a total of } s \text{ light and } S \text{ heavy } r\text{-vertices are generated}\right] \\ &= q_r(1, 1), \\ y_R &= q_R(1, 1). \end{aligned}$$

Thus y_r and y_R satisfy the equations

$$y_r = g_r(y_r, y_R), \quad y_R = g_R(y_r, y_R). \tag{12.6.2}$$

When $t > h_{\mathrm{cr}}$, there is, according to Theorem 3.4.6, a unique solution (y_r, y_R) of (12.6.2) satisfying $0 \leq y_r, y_R < 1$.

The very first step in the branching process consists of creating a red tree. We call this tree $C_1(v)$. The probability of the process dying out is

$$\begin{aligned} P(t) &= \sum_{k,p} \Pr\left[C_1(v) \text{ is a } (k,p)\text{-tree}\right] y_r^p y_R^{k-p} \\ &= \sum_{k,p} \frac{k-1}{2^k} e_{k,p} y_r^p y_R^{k-p} \\ &= \frac{y_r y_R (y_r - y_R)^2 e^{(y_R - y_r)/2}}{4\left(y_r e^{(y_R-y_r)/2} - y_R\right)}. \end{aligned}$$

12.7 Phase transition

In this section we will prove the following phase transition, using the eigenvectors of the transition matrix of the branching process defined in the previous section.

Theorem 12.7.1. *Let*

$$h_{cr} = \log\left(\frac{16\log 2 - 2}{3\log 2 - 1 + \log 2 \cdot \sqrt{27 - 16\log 2}}\right) \doteq 0.8607.$$

(1) If $t < h_{cr}$, then a.a.s. *every component in $G_{\min}(n, tn)$ has $O(\log n)$ vertices.*

(2) If $t = h_{cr}$, and $\omega(n) \to \infty$, then $G_{\min}(n, tn)$ a.a.s. *contains no component of order greater than $n^{2/3}\omega(n)$, and at least one component of order greater than $n^{2/3}/\omega(n)$.*

(3) If $t > h_{cr}$, then a.a.s. *the largest component in $G_{\min}(n, tn)$ has $\Theta(n)$ vertices and every other component, if any, has $O(\log n)$ vertices.*

Proof. We first prove part (1). Assume that $t < h_{\mathrm{cr}}$. Let v be a vertex in $G_{\min}(n, tn)$, and let $C(v)$ be the component containing v. We will bound the probability that $C(v)$ has more than $O(\log n)$ vertices, by using the branching process of the previous section.

Since every vertex in $G_{\min}(n, tn)$ is incident to at least one red edge, every b-vertex in the branching process gives rise to at least one r-vertex. Thus it is sufficient to count the number of r-vertices, since the total number of vertices is at most twice the number of r-vertices.

Let A be as (12.6.1), and let $\lambda_1 = \lambda_1(t)$ be its largest eigenvalue. Since $t < h_{\mathrm{cr}}$, we have that $\lambda_1 = 1 - \varepsilon$ for some $\varepsilon > 0$. For a fixed k, we let $[v_1 v_2]$ be the left eigenvector of A, corresponding to the eigenvalue λ_1, such that $v_1 + v_2 = k$.

Imagine that we start with u_1 light r-vertices and u_2 heavy r-vertices. From each of these vertices we first generate a blue tree, and from each of the new vertices we then generate a red tree. Let V_1 be the number of newly generated light r-vertices, and V_2 be the same for heavy r-vertices. We define

$$P(x_1, x_2) = \Pr\left[V_1 \geq v_1 \text{ or } V_2 \geq v_2 | u_1 = x_1, u_2 = x_2\right].$$

The probability that $C(v)$ has at least $2k$ vertices is bounded from above by the probability that the branching process generates at least k r-vertices. This

probability is again bounded from above by

$$\max_{x_1 \le v_1, x_2 \le v_2} P(x_1, x_2) = P(v_1, v_2).$$

Set $u_1 = v_1$ and $u_2 = v_2$. Since $[v_1 v_2]$ is an eigenvector, $\mathbb{E}[V_1] = (1-\varepsilon)v_1$ and $\mathbb{E}[V_2] = (1-\varepsilon)v_2$.

Let $W_{r\to r}$ denote the random variable counting the number of light r-vertices generated from one light r-vertex. Similarly define $W_{r\to R}$, $W_{R\to r}$ and $W_{R\to R}$. Then $V_1 = \sum_{i=1}^{v_1} W_{r\to r,i} + \sum_{i=1}^{v_2} W_{R\to r,i}$ and $V_2 = \sum_{i=1}^{v_1} W_{r\to R,i} + \sum_{i=1}^{v_2} W_{R\to R,i}$.

Thus both V_1 and V_2 are the sum of two random variables, each of which is a sum of a number of non-negative, integral, mutually independent, and identically distributed random variables. Hence we can use Lemma 3.4.4 to show that if we set $k = c\log n$, we can always choose c so large that

$$\Pr[V_i \ge v_i] = o(n^{-1}),$$

for $i = 1, 2$, and so $P(v_1, v_2) = o(n^{-1})$.

It follows that the expected number of vertices in components of size greater than $2k$ is $n \cdot o(n^{-1}) = o(1)$, which finishes the proof of part (1).

We prove part (2) the case $t = h_{\mathrm{cr}}$, i.e., when $\lambda_1(t) = 1$. This is called the *critical phase*. It turns out that the largest component in this case has order roughly $n^{2/3}$.

It will be convenient to consider the branching process in Section 12.6 as a single-type, rather than a multi-type branching process. We still distinguish between light and heavy vertices and between r-vertices and b-vertices, but we only count one type, say the light r-vertices. A step in the branching process consists of taking a light r-vertex and generating a blue tree incident to it, and then generating a red tree for each of the newly generated b-vertices. However, instead of stopping here as we did in the previous section, if there are now heavy r-vertices, we continue generating trees from these vertices, until we only have light r-vertices.

We let p_i be the probability that one light r-vertex generates precisely i new light r-vertices in this process. We let $p(z) = \sum_{i\ge 0} p_i z^i$ be the corresponding generating function. Since we are at the critical moment of the branching process, i.e., when $\lambda_1(t) = 1$, according to Theorem 3.4.6, the process dies out with probability 1. Hence the number of vertices created throughout the process is finite with probability 1, and so $p(1) = 1$.

We let $h(z)$ be the generating function for the number of light r-vertices that are created when starting with one *heavy* r-vertex. Then

$$p(z) \quad = \quad g_r(z, h(z)), \quad h(z) = g_R(z, h(z)).$$

Now we let q_i be the probability that the branching process starting from one light r-vertex dies out after having produced precisely i light r-vertices, including the vertex we started with. Then

$$q(z) = zp(q(z)).$$

We will use Lemma 2.4.1 with $f(z) = q(z)$ and $\phi(u) = p(u)$. As explained above, $p(1) = 1$. The expected number of light r-vertices produced by a light

r-vertex is given by $p'(1)$. The value $p'(1)$ is in fact an increasing function of t, which is continuous. If $p'(1) > 1$, the process continues forever with positive probability. Since this is not the case when $t = h_{\mathrm{cr}}$, we must have $p'(1) \leq 1$. But for any $t > h_{\mathrm{cr}}$, the process does continue forever with positive probability, so in this case $p'(1) > 1$. Because of continuity, we must therefore have $p'(1) = 1$ when $t = h_{\mathrm{cr}}$.

Hence, by Lemma 2.4.1, the dominant singularity of $q(z)$ is $\rho = 1$. To apply Lemma 2.4.1, we only have to show that $p''(1) \neq 0$. This holds, since

$$p''(1) = \sum_{i \geq 2} i(i-1)p_i > 0.$$

Thus from Formula (2.4.4) we obtain

$$q_i \sim ci^{-\frac{3}{2}}. \tag{12.7.1}$$

Given a vertex v, we can build up the component $C(v)$, as explained earlier, by starting with a red tree, and then adding blue and red trees alternatingly. The vertices in $C(v)$ are then labelled light and heavy, and r-vertices and b-vertices. Let A_k be the event that for some vertex v in $G_{\min}(n, h_{\mathrm{cr}}n)$, $C(v)$ contains a component with at least k light r-vertices, and let X_k be the number of light r-vertices contained in such components. Clearly $\Pr[A_k] = \Pr[X_k \geq k]$. Let Q_k be the probability that a branching process starting with one light r-vertex lasts until at least k light r-vertices have been created. Then

$$Q_k = \sum_{i \geq k} q_k \sim \sum_{i \geq k} ci^{-\frac{3}{2}} \sim c \int_k^\infty x^{-\frac{3}{2}} dx = 2ck^{-\frac{1}{2}}.$$

The branching process is only an approximation to the process of exposing the components. In the branching process the probability that we choose a red or blue tree of order k remains fixed throughout, but in the graph, every time we choose a tree of some order, the number of such trees in the graph is reduced by one, so the probability that we choose such a tree again later on is changed slightly. Every time a tree of order k is chosen, the number of vertices in such trees is reduced by k. Therefore the number of vertices in large trees is reduced at a greater rate than the number of vertices in smaller trees, so the expected growth of the component exposure process is bounded by the expected growth of the branching process.

It follows that the probability that $C(v)$ is in a component with at least k light r-vertices is bounded above by Q_k. Hence $\mathbb{E}[X_k] \leq nQ_k$, and by Markov's inequality,

$$\Pr[A_k] = \Pr[X_k \geq k] \leq \frac{\mathbb{E}[X_k]}{k} \sim 2cnk^{-\frac{3}{2}}.$$

Let $\omega(n) \to \infty$. If $k = n^{2/3}\omega(n)$, then

$$\Pr[A_k] \leq \frac{2c}{\omega(n)} \to 0.$$

Hence, *a.a.s.* $G_{\min}(n, h_{\mathrm{cr}}n)$ is such that no matter which vertex v we start with, if we build up $C(v)$, we will encounter fewer than $n^{2/3}\omega(n)$ light r-vertices, for any function $\omega(n) \to \infty$.

By duplicating what we have done so far, but considering *heavy* r-vertices, rather than light r-vertices, we can show that the same is true for them. It follows that *a.a.s.*, for every vertex v, $C(v)$ will contain less than $n^{2/3}\omega(n)$ r-vertices. In Section 12.7 it was explained that the total number of vertices in $C(v)$ is at most twice the number of r-vertices. Hence we conclude that *a.a.s.*, for every vertex v, $C(v)$ contains at most $n^{2/3}\omega(n)$ vertices.

Now we consider the lower bound on the order of the largest component. We will prove that for any function $\omega(n) \to \infty$, there is at least one component $C(v)$ which contains $n^{2/3}/\omega(n)$ light r-vertices.

Let $\rho'(n,k)$ be the probability that a given vertex v in $G_{\min}(n, h_{\mathrm{cr}}n)$ is in a component of order at least k. As mentioned above the probability $\rho'(n,k)$ is bounded above by Q_k. To find a lower bound for $\rho'(n,k)$ we define a modified branching process, as in the previous section: Whenever a new vertex is generated, we discard it with probability $n^{-1/3}\omega'(n)$, for some function $\omega'(n) \to \infty$.

When we build the component $C(v)$, a newly "generated" vertex w may coincide with an already generated, but unsaturated vertex w'. In this case we cannot generate new trees from both w and from w' - we solve this by disregarding both the vertex w and the vertex w'. The probability that a vertex has to be disregarded in this manner is bounded above by $n^{2/3}\omega'(n)/n = n^{-1/3}\omega'(n)$, since we have already proved that there are no components of order larger than $n^{2/3}\omega'(n)$.

Furthermore, as mentioned above, the probability that we choose a red or blue tree of some order l, changes slightly throughout the process. Let for instance r_l be the probability that a given vertex is in a red tree of order l, at the beginning of the exposure. Thus the graph contains $r_l n$ vertices in such trees. After we have exposed $cn^{2/3}$ vertices, the expected number of vertices exposed which are in red trees of order l is $r'_l n^{2/3}$, for some constant r'_l. If we now choose a vertex at random from the non-exposed vertices, the probability that this vertex is in a red tree of order k is therefore

$$\frac{r_l n - r'_l n^{2/3}}{n - cn^{2/3}} = r_l + O(n^{-1/3}).$$

Similar calculations hold if we instead consider the probability that a red (resp. blue) tree of order l is generated from a light or heavy b-vertex (resp. r-vertex). Thus, the probability that the branching process in one step chooses a tree of "wrong" order is smaller than $n^{-1/3}\omega'(n)$.

By discarding a newly generated vertex with probability $n^{-1/3}\omega'(n)$, the order of the components generated by the modified branching process therefore gives us a lower bound on the order of the components in the graph process.

Now let $\omega(n) \to \infty$ and $k = n^{2/3}\omega(n)$, and choose $\omega'(n)$ such that $\omega'(n)^2 = o(\omega(n))$. If we let $p_-(z)$ be the generating function for the number of light r-vertices which are generated in one step from one light r-vertex, we have

$$\begin{aligned} p_-(z) &= \sum_{i\geq 0} p_i((1-n^{-1/3}\omega'(n))z + n^{-1/3}\omega(n))^i \\ &= p((1-n^{-1/3}\omega'(n))z + n^{-1/3}\omega'(n)). \end{aligned}$$

Defining $q_-(z)$ analogously to $q(z)$, we get $q_-(z) = zp((1-n^{-1/3}\omega'(n))q_-(z) + n^{-1/3}\omega'(n))$. Let ρ_- be the dominant singularity of $q_-(z)$, as in Lemma 2.4.1. By studying the functions $p_-(z)$ and $q_-(z)$ close to $z = 1$, one can show that

$\log \rho_- = c' \left(n^{-1/3}\omega'(n)\right)^2$ for some constant c'. Hence, $\rho_-^{-k} = e^{-c'kn^{-2/3}\omega'(n)^2} = e^{-o(1)}$, so the dominant singularity of $q_-(z)$ approaches 1. The constant in (12.7.1) depends only on the derivatives of $p(z)$, so when $n \to \infty$, we get $[z^i]q_-(z) \sim [z^i]q(z)$. It follows that $\rho'(n,k)$ is asymptotically bounded below by Q_k, and hence that $\rho'(n,k) \sim Q_k$ and $\mathbb{E}[X_k] \sim nQ_k$.

According to the above argument there is *a.a.s.* no component with more than $k' = n^{2/3}\omega'(n)$ vertices. We want to prove that there is at least one component with at least k vertices. We let X_k be as above, such that

$$\mathbb{E}[X_k] \sim nQ_k = 2cnk^{-\frac{1}{2}} = 2cn^{2/3}\sqrt{\omega(n)}.$$

Let v be a vertex in a component with at least k vertices. *A.a.s* $C(v)$ contains at most k' vertices. Hence the expected number of pairs of vertices in components of order at least k, which include v, is bounded above by $k' + \mathbb{E}[X_k]$. Thus

$$\begin{aligned}
\mathbb{E}[X_k(X_k-1)] &\leq \mathbb{E}[X_k]\,(k' + \mathbb{E}[X_k]) \\
&\sim 2cn^{2/3}\sqrt{\omega(n)}(n^{2/3}\omega'(n) + 2cn^{2/3}\sqrt{\omega(n)}) \\
&= \mathbb{E}[X_k]^2\,(1+o(1)),
\end{aligned}$$

so again by Chebyshev's inequality, we get that *a.a.s.* $X_k \sim \mathbb{E}[X_k]$. In particular there is *a.a.s.* a component in $G_{\min}(n, h_{\mathrm{cr}}n)$ of order at least k.

Now we prove part (3). Assume that $t > h_{\mathrm{cr}}$. In the branching process we distinguish between *saturated* and *unsaturated* vertices as in Section 12.6. Saturated vertices are those from which we already have added a red or a blue tree, while unsaturated vertices are those which have been generated, but from which we have not yet generated a new tree.

One problem appears when we try to use the branching process to model the graph process: In the branching process, newly generated vertices will always be distinct from earlier generated vertices. In the graph process it may happen that a newly generated vertex is the same as one of the unsaturated vertices. If the number of vertices we already have in the component is k, then the probability that a given newly generated vertex is one of the unsaturated vertices, is bounded above by $\frac{k}{n}$. As long as $k \ll n$, for any $\varepsilon' > 0$ this value is smaller than ε' for large n. We therefore introduce a modified branching process: Whenever a new vertex is generated, we discard this vertex with probability $\varepsilon' > 0$. Since $t > h_{\mathrm{cr}}$, and the eigenvalue function is continuous, we can always find an ε' small enough, so that the largest eigenvalue remains strictly greater than one. We will from now on assume that we are using this modified branching process. $\lambda_1 = 1 + \varepsilon$ will be the largest eigenvalue, and $\varepsilon > 0$.

We will now prove that there exist constants $c > 0$ and $\delta > 0$, such that for any η with $0 < \eta < \frac{1}{2}$, the following is true. Let $k_- = c \log n$ and $k_+ = n^{1-\eta}$. The probability that there is a vertex v such that for some k with $k_- < k < k_+$, the branching process starting with v has less than δk unsaturated vertices after k steps, given that the branching process has not died out before k_- steps, is $o(1)$.

To prove this, we fix k, with $k_- < k < k_+$, and assume that the branching process has not died out after k_- steps. The expected number of vertices generated from k vertices is $k(1+\varepsilon)$. Let δ be a constant with $0 < \delta < \varepsilon$. Lemma 3.4.4 implies that the c which k_- depends on can be chosen so large

that the probability that the branching process after k steps (with $k > k_-$) has fewer than $k(1+\delta)$ vertices is $o\left(n^{-2}\right)$. Thus the number of unsaturated vertices is at least $k(1+\delta) - k = \delta k$. The probability that for some vertex v and some k with $k_- < k < k_+$ the number of unsaturated vertices is less than δk is then bounded from above by

$$n \sum_{k=k_-}^{k_+} o\left(n^{-2}\right) = o(1).$$

In other words, there is *a.a.s.* no component with between k_- and k_+ vertices.

Now we want to show that there is no more than one component with more than k_+ vertices. Let v and w be two vertices belonging to components with at least k_+ vertices. Now we run the branching processes starting with v and w. According to what we showed above, when we have reached k_+ vertices in each of the branching processes, each of them has δk_+ unsaturated vertices. By Lemma 3.4.4, the δk_+ unsaturated vertices in the component containing v generate at least $\delta' k_+$ vertices for some $\delta' > 0$. The probability that none of these vertices is one of the unsaturated vertices in w's component is bounded from above by

$$\left(\frac{n - \delta k_+}{n}\right)^{\delta' k_+} = \left(1 - \delta n^{-\eta}\right)^{\delta' n^{1-\eta}} \leq e^{-\delta\delta' n^{1-2\eta}} = o\left(n^{-2}\right).$$

So *a.a.s.*, if v and w are in components with more than k_+ vertices, then v and w are in the same component.

We call a component *small* if it has less than k_- vertices and *large* if it has more than k_+ vertices. We let $\rho(n,t)$ be the probability that a vertex v in $G_{\min}(n, tn)$ is in a small component. In Section 12.6 we proved that when $t > h_{\mathrm{cr}}$ the probability that the branching process dies out is $P(t)$ with $0 < P(t) < 1$. Thus $\rho(n,t)$ is bounded from below by $P(t)+o(1)$. We let $P_{\varepsilon'}(t)$ be the extinction probability of the modified branching process defined earlier in the proof, where a newly created vertex is discarded with probability ε'. As long as $\varepsilon' > 0$, $P_{\varepsilon'}(t)$ is an upper bound for the probability that $C(v)$ is small. As $\varepsilon' \to 0$, $P_{\varepsilon'}(t)$ converges to $P(t)$, so $\rho(n,t) \to P(t)$ as $n \to \infty$.

Letting Y be the number of vertices in small components, we get that $\mathbb{E}[Y] = (P(t) + o(1))\, n$. Furthermore,

$$\mathbb{E}[Y(Y-1)] \leq n\rho(n,t)(k_- + n\rho(n - O(k_-), t) = (1 + o(1))\mathbb{E}[Y]^2,$$

so by Chebyshev's inequality, *a.a.s.* $Y = (P(t) + o(1))\, n$. Hence the number of vertices in the largest component is $(1 - P(t) + o(1))\, n$. □

Chapter 13

Min-min Random Graph Process

In this chapter we investigate the evolution of the *min-min random graph process* $(G_M)_{0 \le M \le N}$. The process starts with an empty graph G_0 with vertex set $V := [n]$, where n is even. Furthermore, in each step G_{M+1} is obtained from G_M by first choosing a pair $\{v, w\}$ of distinct non-adjacent vertices of G_M of minimum degree uniformly at random among all such pairs and adding a new edge $\{v, w\}$ to G_M.

Note that it may happen that at a certain step of the min-min random graph process, G_M say, we cannot anymore select a pair $\{v, w\}$ of distinct non-adjacent vertices in G_M where *both* v, w are of minimum degree in G_M, which happens with probability tending to zero as $n \to \infty$. In this case we restart the process from the empty graph. Let $\mathcal{B}$ be an event that the min-min random graph process continues at least until $M = rn/2$ for a fixed $r \ge 3$. This happens with probability tending to one as $n \to \infty$. Hence it is sufficient to show that a certain event holds *a.a.s.* conditioned on $\mathcal{B}$, if we want to show that *a.a.s.* it holds. Thus we assume that the event $\mathcal{B}$ holds below.

Note that the maximum and the minimum degree of G_M differ by at most one for all M. More precisely, if we let $r(M) = \lfloor 2M/n \rfloor$ and $\nu(M) = (r+1)n - 2M$, then G_M has exactly $\nu(M)$ vertices of degree $r(M)$ and $n - \nu(M)$ vertices of degree $r(M) + 1$. Hence, if $M = rn/2$, then G_M is r-regular, although, e.g., in the case $r = 2$ G_M is *not* a uniformly distributed r-regular graph.

Let $M = n + tn$ for $-1 \le t \le 1/2$. At the beginning of the min-min random graph process the evolution is quite simple: If $-1 \le t < -1/2$, then G_M will have less than $n/2$ edges. The minimum degree equals 0 and the maximum degree of G_M equals one. Thus G_M simply consists of isolated vertices and isolated edges. When $t = -1/2$, the minimum degree of G_M reaches one, and $G_M = G_{n/2}$ is a perfect matching of the vertex set V. In fact, G_M is a uniformly distributed random perfect matching of V, because the distribution of G_M is invariant under permutations of the vertex set.

In the range $-1/2 \le t < 0$, G_{M+1} is obtained from G_M by connecting two randomly chosen vertices of degree one by a new edge, and G_{M+1} consists of isolated paths and isolated even cycles.

When $t = 0$, the minimum degree of G_M increases to two, and $G_M = G_n$

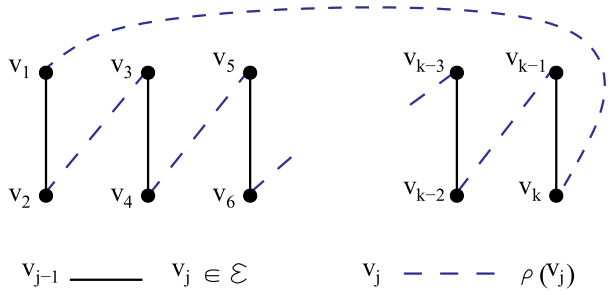


Figure 13.1: An $\mathcal{E}$-cycle.

consists of isolated even cycles, that is, it is a random bipartite 2-regular graph. However, G_n is *not* uniformly distributed. We will study the distribution of G_n in Sections 13.1.

In the range $t > 0$, the cycles of G_n glue together to form a large component of order $\Omega(n)$ in G_M, and indeed *a.a.s.* the random graph G_M consists of a giant component on $(1 - o(1))n$ vertices and a number of small isolated even cycles as we will see in Section 13.3.

When $t = 1/2$, the minimum degree of G_M increases to three, and $G_M = G_{3n/2}$ is a random 3-regular graph. However, we do not know if $G_{3n/2}$ is uniformly distributed or is contiguous to the uniformly distributed 3-regular graph $G_n^{(3)}$. Moreover, all the isolated cycles appeared before have joined into the giant component and thus $G_{3n/2}$ is *a.a.s.*connected.

In the rest of the chapter we study further typical properties of G_M when $0 \le t \le 1/2$. In Section 13.1 we study the configuration model to determine the distribution of G_n. In Sections 13.2 and 13.3 we investigate the distribution of isolated cycles and the structure of components of G_M. In Sections 13.4 and 13.5 we study the probability that G_M is connected and determine the asymptotic distribution of the order of the largest component of G_M. To this end, we combine "classical" probabilistic methods with the singularity analysis and the analysis of characteristic functions (i.e., Fourier transforms).

13.1 Configuration model

To study the distribution of G_n, we make a little detour via the well-known concept of configuration.

The configuration model was invented by Bender and Canfield [15] and by Bollobás [34] to study random regular graphs. We will use a result on the number of cycles in the configuration model to investigate the min-min random graph process.

Let $\nu \ge 2$ be an even integer, and let $W = \{1, \dots, \nu\}$. Let $\mathcal{E} = \{e_1, \dots, e_{\nu/2}\}$ be a set of pairwise disjoint subsets $e_j \subset V$ of cardinality 2. For an even integer $k \ge 2$ we call a sequence $C = (v_1, \dots, v_k)$ of pairwise distinct elements of W an *$\mathcal{E}$-cycle of length k* when $\{v_{j-1}, v_j\} \in \mathcal{E}$ for even j with $2 \le j \le k$ (see Figure 13.1).

Moreover, let $\rho : V \to V$ be a *perfect matching*; that is, $\rho\circ\rho = \mathrm{id}$ and $\rho(v) \neq v$ for all $v \in W$. Then we say that the $\mathcal{E}$-cycle C *occurs in* ρ if $\rho(v_j) = v_{j+1}$ for $2 \leq j < k$ even, and $\rho(v_k) = v_1$.

The number of $\mathcal{E}$-cycle satisfies the following property. A proof of it can be found in [33, Sec. 2.4].

Proposition 13.1.1. *Let K be an arbitrarily large even number that remains fixed as $n \to \infty$. Furthermore, let Y_k be the number of $\mathcal{E}$-cycles of length k occurring in a perfect matching ρ, where ρ is chosen uniformly at random among all $(\nu-1)!!$ possible perfect matchings. Let $\lambda_k = k^{-1}$. Then $(Y_k)_{2\leq k\leq K,\text{ even}}$ are asymptotically independent Poisson variables with means $(\lambda_k)_{2\leq k\leq K,\text{ even}}$ as $\nu \to \infty$.*

We call a perfect matching ρ an *$\mathcal{E}$-configuration* if $\{v, \rho(v)\} \notin \mathcal{E}$ for all $v \in V$.

Corollary 13.1.2. *The number of $\mathcal{E}$-configurations is $(1+o(1))\exp(-1/2)(\nu-1)!!$ as $\nu \to \infty$.*

Proof. A random perfect matching ρ is an $\mathcal{E}$-configuration iff $Y_2 = 0$, and by Proposition 13.1.1 we have $\Pr[Y_2 = 0] \sim \exp(-\lambda_2) = \exp(-1/2)$. □

Let $\mathcal{E}$ denote the set of all sequences $(e_1, \ldots, e_{n/2})$ of pairwise disjoint subsets of V of cardinality 2. Moreover, fix any perfect matching G_* of V, and let $\mathcal{E}(G_*)$ signify the set of all $(e_1, \ldots, e_{n/2}) \in \mathcal{E}$ such that $e_i \notin E(G_*)$ for all i. Then each G_*-configuration ρ corresponds to $(n/2)!$ elements of $\mathcal{E}(G_*)$, which specify the order in that the edges $\{v, \rho(v)\}$ are added. We shall first work with the tuples $(e_1, \ldots, e_{n/2}) \in \mathcal{E}(G_*)$ rather than with the configurations $\rho \in \mathcal{K}(G_*)$.

Lemma 13.1.3. *We have $|\mathcal{E}(G_*)| \geq \frac{1}{3}|\mathcal{E}|$.*

Proof. Let $(e_1, \ldots, e_{n/2}) \in \mathcal{E}$ be chosen uniformly at random. Then the expected number of indices i such that $e_i \in E(G_*)$ equals $\frac{n}{2} \cdot \frac{(n-3)!!}{(n-1)!!} \leq \frac{2}{3}$. Therefore, the assertion follows from Markov's inequality. □

For an index $1 \leq m \leq \frac{n}{2}$ and a tuple $(e_1, \ldots, e_{n/2}) \in \mathcal{E}$ we let $Z_m = Z_m(E)$ be the number of isolated edges of $G_* + \{e_1, \ldots, e_m\}$.

Lemma 13.1.4. *Let $E = (e_1, \ldots, e_{n/2}) \in \mathcal{E}$ be chosen uniformly at random.*

(i) If $1 \leq m \leq \frac{n}{2} - n^{3/100}$, then

$$\Pr\left[|Z_m - \frac{n}{2}(1-2m/n)^2| \geq n^{99/100}(1-2m/n)\right] \leq \exp(-n^{1/100}).$$

(ii) With probability $\geq 1 - n^{-9/10}$ we have $Z_m = 0$ for all $m \geq \frac{n}{2} - n^{3/100}$.

Proof. The set S of all vertices $v \in V$ such that $v \in e_i$ for some $1 \leq i \leq m$ is a uniformly distributed subset of V of cardinality $2m$. In addition, consider a random subset W of V obtained by including every vertex $v \in V$ with probability $p = 2m/n$ independently of all others. Then $|W|$ is binomially distributed with mean $2m$, so that

$$\Pr[|W| = 2m] = \Omega(n^{-1/2}). \tag{13.1.1}$$

Furthermore, given that $|W| = 2m$, W is uniformly distributed. Thus, letting Z'_m denote the number of edges $e \in E(G_*)$ such that $e \subset V \setminus W$, the conditional distribution of Z'_m given that $|W| = 2m$ coincides with the distribution of Z_m.

Suppose that $1 \leq m \leq \frac{n}{2} - n^{3/100}$. As Z'_m is binomially distributed with mean $\frac{n}{2}(1-p)^2$, the Chernoff bound (3.4.6) entails that

$$\Pr\left[|Z'_m - \frac{n}{2}(1-2m/n)^2| \geq n^{99/100}(1-2m/n)\right] \leq \exp\left[-n^{1/50}\right]. \quad (13.1.2)$$

Combining (13.1.1) and (13.1.2), we conclude that

$$\begin{aligned}\Pr&\left[|Z'_m - \frac{n}{2}\left(1-\frac{2m}{n}\right)^2| \geq n^{\frac{99}{100}}\left(1-\frac{2m}{n}\right) \;\middle|\; |W| = 2m\right]\\ &\leq \quad O(\sqrt{n})\exp(-n^{1/50}) \leq \exp(-n^{1/100}),\end{aligned}$$

thereby proving the first assertion.

Now, let $m \geq \frac{n}{2} - n^{3/100}$. Since S is uniformly distributed, for any edge $e \in E(G_*)$ we have

$$\Pr[e \cap S = \emptyset] = \binom{n-2}{2m}\binom{n}{2m}^{-1} \leq (1-2m/n)^2.$$

In effect, the expected number of edges $e \in E(G_*)$ such that $e \cap S = \emptyset$ is $\leq \frac{n}{2}(1-2m/n)^2 \leq n^{-9/10}$. Hence, the second assertion follows from Markov's inequality. □

Let $\mathcal{E}_0(G_*)$ denote the set of all $E \in \mathcal{E}(G_*)$ such that

$$|Z_m(E) - \frac{n}{2}(1-2m/n)^2| < n^{99/100}(1-2m/n)$$

for all $1 \leq m \leq \frac{n}{2} - n^{3/100}$, and $Z_m(E) = 0$ for all $m \geq \frac{n}{2} - n^{3/100}$. As a direct consequence of Lemmas 13.1.3 and 13.1.4 we have the following result.

Corollary 13.1.5. *If $E \in \mathcal{E}(G_*)$ is chosen uniformly at random, then*

$$\Pr[E \in \mathcal{E}_0(G_*)] \geq 1 - n^{-4/5}.$$

For $E = (e_1, \ldots, e_{n/2}) \in \mathcal{E}(G_*)$ we set

$$p(E, G_*) = \Pr\left[G_{\frac{n}{2}+j+1} = G_{\frac{n}{2}+j} + e_{j+1} \text{ for } 0 \leq j < \frac{n}{2} | G_{n/2} = G_*\right].$$

Lemma 13.1.6. *(i) We have $p(E, G_*) \sim |\mathcal{E}(G_*)|^{-1}$ uniformly for all $E \in \mathcal{E}_0(G_*)$.*

(ii) For any element $E \in \mathcal{E}(G_*)$ we have $p(E, G_*) \leq n^{3/5}|\mathcal{E}(G_*)|^{-1}$.

Proof. For $E = (e_1, \ldots, e_{n/2}) \in \mathcal{E}(G_*)$ we let $\mathcal{F}_m = \mathcal{F}_m(E)$ denote the event that $E(G_{\frac{n}{2}+m}) \setminus E(G_{\frac{n}{2}}) = \{e_1, \ldots, e_m\}$. In addition, $\mathcal{F}_0$ signifies the event that $G_{n/2} = G_*$, and $p_m = p_m(E) = \Pr[\mathcal{F}_m | \mathcal{F}_i \text{ for } 0 \leq i < m]$. Furthermore, we let $T_m = \binom{n-2j}{2} - Z_m$ be the number of possible edges that can be added to G_m in order to obtain G_{m+1}; thus, the probability that any specific pair $\{v, w\}$ of distinct non-adjacent vertices of minimum degree of G_m is chosen is T_m^{-1}.

To prove the first assertion, assume that $E \in \mathcal{E}_0(G_*)$. Then given that $\mathcal{F}_i$ occurs for $0 \le i \le m$ we have

$$T_m = \begin{cases} \binom{n-2m}{2} - \frac{n}{2}(1-\frac{2m}{n})^2 + \Delta_m, & \text{if } m \le \frac{n}{2} - n^{3/100}, \\ \binom{n-2m}{2}, & \text{otherwise,} \end{cases}$$

where $|\Delta_m| \le n^{99/100}(1-\frac{2m}{n})$. Therefore, letting

$$q_{m+1} = \left[\binom{n-2m}{2} - \frac{n}{2}(1-\frac{2m}{n})^2\right]^{-1}$$

for $0 \le m \le \frac{n}{2} - n^{3/100}$ and $q_{m+1} = \binom{n-2m}{2}^{-1}$ for $\frac{n}{2} - n^{3/100} < m \le \frac{n}{2}$, we have

$$\frac{q_{m+1}}{p_{m+1}} = \begin{cases} 1 + \frac{\Delta_m}{\binom{n-2m}{2} - \frac{n}{2}(1-\frac{2m}{n})^2}, & \text{if } m \le \frac{n}{2} - n^{3/100}, \\ 1, & \text{otherwise.} \end{cases}$$

Consequently,

$$\begin{aligned} \frac{\prod_{0\le m<\frac{n}{2}} q_{m+1}}{p(E,G_*)} &= \prod_{0\le m<\frac{n}{2}} \frac{q_{m+1}}{p_{m+1}} \\ &\le \prod_{0\le m\le \frac{n}{2}-n^{\frac{3}{100}}} \left(1 + \frac{n^{\frac{99}{100}}(1-2m/n)}{\binom{n-2m}{2} - \frac{n}{2}(1-2m/n)^2}\right) \\ &\le \exp\left[O(n^{-1/100}) \sum_{0\le m\le \frac{n}{2}-n^{\frac{3}{100}}} \frac{1}{n-2m}\right] \\ &= 1 + o(1), \end{aligned} \tag{13.1.3}$$

and similarly

$$\frac{\prod_{0\le m<\frac{n}{2}} q_{m+1}}{p(E,G_*)} = \prod_{0\le m<\frac{n}{2}} \frac{q_{m+1}}{p_{m+1}} \ge 1 - o(1). \tag{13.1.4}$$

Since the quantity $\prod_{0\le m<\frac{n}{2}} q_{m+1}$ does not depend on E, the first assertion follows from (13.1.3) and (13.1.4).

In order to establish the second assertion, consider any $E \in \mathcal{E}(G_*)$. As we have

$$\binom{n-2m}{2} - \frac{1}{2}(n-2m) \le T_m \le \binom{n-2m}{2}$$

for all $0 \le m \le \frac{n}{2}$, we get

$$1 - \frac{1}{n-2m-1} \le p_{m+1}^{-1}\binom{n-2m}{2}^{-1} \le 1 \quad (0 \le m < \frac{n}{2} - 1),$$

and $p_{n/2} = 1$. Therefore,

$$\begin{aligned} 1 &\geq \left[p(E, G_*) \prod_{0 \leq m < \frac{n}{2}-1} \binom{n-2m}{2} \right]^{-1} = \prod_{0 \leq m < \frac{n}{2}-1} \frac{1}{p_{m+1}\binom{n-2m}{2}} \\ &\geq \prod_{0 \leq m < \frac{n}{2}-1} \left(1 - \frac{1}{n-2m-1}\right) \\ &\geq \exp\left[- \sum_{3 \leq j \leq n,\ \text{odd}} (j^{-1} + j^{-2}) \right] \geq \exp\left[-\frac{1}{2} \ln n - O(1) \right] \\ &\geq n^{-3/5}. \end{aligned} \tag{13.1.5}$$

Since $\prod_{0 \leq m < \frac{n}{2}-1} \binom{n-2m}{2}^{-1}$ does not depend on E, (13.1.5) implies the second assertion. □

Distribution of G_n. To describe the distribution of G_n, note first that the edge set $E(G_{n/2})$ of $G_{n/2}$ is a partition of V into $n/2$ pairwise disjoint subsets of cardinality 2, so that we can use the notion of $\mathcal{E}(G_{n/2})$-configurations; for brevity just write "$G_{n/2}$-configuration". Moreover, we let $\mathcal{K}(G_{\frac{n}{2}})$ be the set of all $G_{\frac{n}{2}}$-configurations equipped with the uniform distribution. Then to each $\rho \in \mathcal{K}(G_{\frac{n}{2}})$ we can associate a 2-regular graph $G_{\frac{n}{2}} + \rho = G_{\frac{n}{2}} + \{\{v, \rho(v)\} : v \in V\}$.

Proposition 13.1.7. *For any perfect matching G_* of V there is a set $\mathcal{M}(G_*) \subset \mathcal{K}(G_*)$ that enjoys the following properties.*

(i) If $\rho \in \mathcal{K}(G_*)$ is chosen uniformly at random, then $\Pr[\rho \in \mathcal{M}(G_*)] = 1 - o(1)$.

(ii) Given that $G_{n/2} = G_*$, with probability $1 - o(1)$ there is a configuration $\rho \in \mathcal{M}(G_*)$ such that $G_n = G_{n/2} + \rho$.

(iii) We have $\Pr\left[G_n = G_{n/2} + \rho | G_{n/2} = G_*\right] \sim |\mathcal{K}(G_*)|^{-1}$ uniformly for all $\rho \in \mathcal{M}(G_*)$.

Proposition 13.1.7 sets up a connection between the distribution of G_n and the graph $G_{n/2} + \rho$, where $\rho \in \mathcal{K}(G_{n/2})$ is chosen uniformly at random. More precisely, we define a probability space $(\Omega_{\mathcal{M}}, \Pr_{\mathcal{M}})$ as follows: let $\Omega_{\mathcal{M}}$ be the set of all pairs $(\rho, \mathcal{G})$, where $\mathcal{G} = (G_M)_{0 \leq M \leq \binom{n}{2}}$ is a min-min process, and $\rho \in \mathcal{K}(G_{n/2})$. Moreover, let

$$\Pr_{\mathcal{M}}(\rho, \mathcal{G}) = \frac{\Pr[\mathcal{G} | G_n = G_{n/2} + \rho]}{|\mathcal{K}(G_{n/2})|}.$$

In addition, if Z is a random variable that assigns to a min-min process $\mathcal{G}$ a real $Z(\mathcal{G})$, then we can extend Z to $\Omega_{\mathcal{M}}$ by letting $Z(\rho, \mathcal{G}) = Z(\mathcal{G})$. We let $\mathbb{E}_{\mathcal{M}}(Z)$ signify the expectation of Z with respect to $\Pr_{\mathcal{M}}$.

By Proposition 13.1.7 for any set $\mathcal{A}$ of min-min processes we have

$$|\Pr_{\mathcal{M}}[(\rho, \mathcal{G}) \text{ satisfies } \mathcal{G} \in \mathcal{A}] - \Pr[\mathcal{A}]| = o(1).$$

Furthermore, we can describe the distribution of G_n with respect to the measure $\Pr_{\mathcal{M}}$ as follows: first, choose a perfect matching $G_{n/2}$ uniformly at random; then, choose $\rho \in \mathcal{K}(G_{n/2})$ uniformly at random and set $G_n = G_{n/2} + \rho$.

Proof of Proposition 13.1.7. For $\rho \in \mathcal{K}(G_*)$ let $\mathcal{E}(\rho)$ be the set of all sequences $(e_1, \ldots, e_{n/2}) \in \mathcal{E}(G_*)$ such that $\{e_i : 1 \le i \le n/2\} = \{\{v, \rho(v)\} : v \in V\}$. Furthermore, we let

$$\mathcal{M}(G_*) = \left\{\rho \in \mathcal{K}(G_*) : |\mathcal{E}(\rho) \setminus \mathcal{E}_0(G_*)| \le n^{-7/10}|\mathcal{E}(\rho)|\right\}.$$

Then Corollary 13.1.5 implies that

$$|\mathcal{M}(G_*)| \ge (1 - n^{-1/10})|\mathcal{K}(G_*)|,$$

whence the first assertion follows. Moreover, Lemma 13.1.6 entails that

$$\Pr\left[G_n = G_{n/2} + \rho | G_{n/2} = G_*\right] \sim \Pr\left[G_n = G_{n/2} + \rho' | G_{n/2} = G_*\right]$$

uniformly for all $\rho, \rho' \in \mathcal{M}(G_*)$, so that the third assertion follows. Finally, invoking Corollary 13.1.5 and Lemma 13.1.6 once more, we observe that

$$\begin{aligned}\sum_{\rho \in \mathcal{K}(G_*) \setminus \mathcal{M}(G_*)} & \Pr\left[G_n = G_{n/2} + \rho | G_{n/2} = G_*\right] \\ &\le \frac{(n/2)!|\mathcal{K}(G_*) \setminus \mathcal{M}(G_*)|}{(1 - o(1))|\mathcal{E}(G_*)|} + \sum_{E \in \mathcal{E}(G_*) \setminus \mathcal{E}_0(G_*)} p(E, G_*) \\ &\le (1 + o(1))n^{-1/10} + n^{-1/5} = o(1),\end{aligned}$$

thereby proving the second assertion. □

13.2 Isolated cycles

In this section we will study the isolated cycles in G_M with $M = n + tn$ when $0 \le t < 1/2$.

As G_n is bipartite and 2-regular, it consists of isolated even cycles. Using Proposition 13.1.7, we will first prove the following statement, which yields the joint distribution of the number of cycles of bounded length in G_n.

Proposition 13.2.1. *Let C_k be the number of cycles of length k in G_n, for even k. Then for every even constant k_0 $(C_k)_{k=4,6,\ldots,k_0}$ are asymptotically independent Poisson variables with means $(1/k)_{k=4,6,\ldots,k_0}$. Moreover,* a.a.s. *G_n consists of at most $(\frac{1}{2} + o(1)) \ln n$ connected components.*

Proof. Let G_* be any perfect matching of V. Moreover, let C'_k be the number of cycles of length k in $G_* + \rho$, where $\rho \in \mathcal{K}(G_*)$ is chosen uniformly at random. Then Proposition 13.1.1 entails that $(C'_k)_{4 \le k \le k_0,\, \text{even}}$ are asymptotically independent Poisson variables with means $(k^{-1})_{4 \le k \le k_0}$. Finally, due to Proposition 13.1.7 this implies that $(C_k)_{4 \le k \le k_0}$ are asymptotically independent Poisson variables with means $(k^{-1})_{4 \le k \le k_0}$ as well.

In order to bound the number of components of $G_* + \rho$ (where $\rho \in \mathcal{K}(G_*)$ is uniformly distributed) we adapt an argument for counting cycles in a random permutation [62, p. 258]. For each vertex v we let v' denote the neighbour of v in $G_{\frac{n}{2}}$. Then we construct a uniformly distributed random $G_{\frac{n}{2}}$-configuration ρ as follows. Initially, we let the first vertex $v_1 = 1 \in V$ choose its image $v_2 = \rho(v_1)$ uniformly at random from $V \setminus \{v_1, v'_1\}$; we also set $\rho(v_2) = v_1$. Clearly, there

are $n-2$ ways to choose v_2. Now, v_2' chooses a random image $v_3 = \rho(v_2')$, where $v_3 \in V \setminus \{v_1, v_2, v_2'\}$; thus there are $n-3$ ways to choose v_3. Set $\rho(v_3) = v_2'$. If $v_3 \neq v_1'$, then we choose a further vertex $v_4 \in V \setminus \{v_1, v_2, v_2', v_3, v_3'\}$, etc. We proceed in this way until we eventually choose $v_k = v_1'$, thereby closing the cycle $(v_1, v_2, v_2', v_3, v_3', \ldots, v_{k-1}, v_{k-1}', v_k = v_1')$. Since in the $2 \leq k$-th step there are $n-2k+1$ vertices to choose from, the probability that $v_k = v_1'$ equals $p_k = (n-2k+1)^{-1}$. If $v_k = v_1'$, we let $v_{k+1} \in V = \{1, \ldots, n\}$ be the smallest vertex that has not yet been assigned an image $\rho(v_{k+1})$ and repeat the same procedure to determine the second cycle, etc.

Now, to count the cycles occurring in $G_* + \rho$, we define a random variable Z_j as follows: we let $Z_j = 1$ if the j-th step of the construction of ρ closes a cycle, and 0 otherwise. Then $Z = \sum_{j=1}^{n/2} Z_j$ equals the total number of cycles and thus the number of components of $G_{\frac{n}{2}} + \rho$.

However, we cannot characterize the distribution of Z directly, because the random variables Z_j are not independent; for $Z_j = 1$ implies $Z_{j+1} = 0$. Therefore, we consider the following family of slightly modified random variables: we set $\mathcal{Z}_j = Z_j$ if $Z_{j-1} \neq 1$ and $j > 1$. Moreover, in the case $Z_{j-1} = 1$ or $j = 1$, we let $\mathcal{Z}_j = 1$ with probability $p_j = (n-2j+1)^{-1}$, and $\mathcal{Z}_j = 0$ with probability $1-p_j$ independently of the construction of ρ and the values assigned to all other $\mathcal{Z}_i$'s. Then $\Pr[\mathcal{Z}_j = 1] = p_j$ for all $j = 1, \ldots, n/2$, and $\mathcal{Z}_1, \ldots, \mathcal{Z}_{n/2}$ are mutually independent. Furthermore, $\mathcal{Z} = \sum_{j=1}^{n/2} \mathcal{Z}_j \geq Z$ bounds the number of components of $G_{\frac{n}{2}} + \rho$ from above. Thus, we just need to estimate $\Pr\left[\mathcal{Z} \leq (\frac{1}{2} + o(1)) \ln n\right]$. Since the $\mathcal{Z}_j$'s are Bernoulli with mean p_j, we have

$$\begin{aligned} \mathbb{E}(\mathcal{Z}) &= \sum_{j=1}^{n/2} p_j = \sum_{j=1}^{n/2} \frac{1}{2j-1} \sim \frac{1}{2} \ln n, \\ \mathrm{Var}(\mathcal{Z}) &= \sum_{j=1}^{n/2} p_j(1-p_j) \leq \mathbb{E}(\mathcal{Z}). \end{aligned}$$

Therefore, Chebyshev's inequality implies that $\mathcal{Z} \leq (\frac{1}{2} + o(1)) \ln n$ *a.a.s.*, i.e., $G_* + \rho$ has at most $(\frac{1}{2} + o(1)) \ln n$ components *a.a.s.* Finally, Proposition 13.1.7 implies that the same is true for G_n. □

In the range $t > 0$ the cycles of G_n glue together to form a large component of order $\Omega(n)$. To study this process in detail, the next proposition is crucial.

Proposition 13.2.2. *Let G_* be a bipartite 2-regular graph. Suppose that $n^{-1} \leq t < 1/2 - \delta$ for an arbitrarily small $\delta > 0$ that does not depend on n. Fix a set S of s vertices, where $s = o(\sqrt{n/t})$. Let $F = E(G_M) \setminus E(G_n)$. Then*

$$\Pr[e \cap S = \emptyset \text{ for all } e \in F | G_n = G_*] \sim (1-2t)^s.$$

Proof. Let G_* be a bipartite 2-regular graph, and let $S \subset V$ be a set of vertices of order $s = o(\sqrt{n/t})$. Consider a min-min process $\mathcal{G} = (G_0, \ldots, G_{3n/2})$ such that $G_n = G_*$, where G_{i+1} is obtained from G_i by adding an edge e_i. Let $0 \leq k = tn \leq \frac{1}{2}(1-\delta)n$ for an arbitrarily small but fixed $\delta > 0$. Further, let X_j be the number of edges joining two vertices of degree 2 in G_{n+j} $(0 \leq j \leq k)$. Moreover, let X_j' denote the number of edges joining two vertices v, w of degree

2 in G_{n+j} such that $v, w \in V \setminus S$. Finally, let $\mathcal{E}_j$ be the event that the edge e_j added in the $(n+j)$-th step satisfies $e_j \cap S = \emptyset$.

Our goal is to show that

$$\Pr[\mathcal{E}_j \text{ for all } 0 \leq j < k] \sim (1-2k)^s.$$

To this end, we consider the quantity

$$p_j = \frac{\binom{n-2j-s}{2} - X'_j}{\binom{n-2j}{2} - X_j} = \frac{(n-2k-s)_2}{(n-2k)_2} \cdot \frac{1 - X'_k \binom{n-2k-s}{2}^{-1}}{1 - X_k \binom{n-2k}{2}^{-1}}. \tag{13.2.1}$$

Then p_j equals the probability that $\mathcal{E}_j$ occurs given the values of X_j and X'_j and given that all vertices in S have degree 2. For the denominator in (13.2.1) equals the total number of possible edges e_j, because there are $n-2j$ vertices of degree 2 in G_{n+j}. Moreover, since there are $n-2j-s$ vertices of degree 2 outside of S, the numerator in (13.2.1) equals the number of possible edges e_j such that $e_{n+j} \cap S = \emptyset$.

To compute $\Pr[\mathcal{E}_j \text{ for all } 0 \leq j < k]$, we estimate p_j uniformly for all possible values of X_j and X'_j. By our assumption that $n - 2k \geq \delta n$ and $s = o(n)$, we have $X'_k/\binom{n-2j-s}{2}, X_k/\binom{n-2j}{2} = O(1/n)$. Therefore,

$$\begin{aligned}
1 - \frac{X_j}{\binom{n-2j}{2}} &= \exp\left[-\frac{X_j}{\binom{n-2j}{2}} + O\left(\frac{X_j}{\binom{n-2j}{2}}\right)^2\right] \\
&= \exp\left[-\frac{X_j}{\binom{n-2j}{2}} + o(1/n)\right], \qquad (13.2.2) \\
1 - \frac{X'_j}{\binom{n-2j-s}{2}} &= \exp\left[-\frac{X'_j}{\binom{n-2j-s}{2}} + O\left(\frac{X'_j}{\binom{n-2j-s}{2}}\right)^2\right] \\
&= \exp\left[-\frac{X'_j}{\binom{n-2j-s}{2}} + o(1/n)\right]. \qquad (13.2.3)
\end{aligned}$$

Since $s = o(n)$ and the maximum degree of G_{n+j} is ≤ 3, S is incident with at most $3s = o(n)$ edges. Consequently, $X_j - X'_j = o(n)$. Hence, plugging (13.2.2) and (13.2.3) into (13.2.1), we obtain

$$\begin{aligned}
p_j &= \frac{(n-2j-s)_2}{(n-2j)_2} \cdot \exp\left[\frac{X_j}{\binom{n-2j}{2}} - \frac{X'_j}{\binom{n-2j-s}{2}} + o(1/n)\right] \qquad (13.2.4) \\
&= \frac{(n-2j-s)_2}{(n-2j)_2} \cdot \exp\left[X_j\left[\binom{n-2j}{2}^{-1} - \binom{n-2j-s}{2}^{-1}\right] + o(1/n)\right].
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\binom{n-2j}{2}\binom{n-2j-s}{2}^{-1} &= \left(1 + \frac{s}{n-2j-s}\right) \cdot \left(1 + \frac{s}{n-2k-s-1}\right) \\
&= 1 + O(s/n). \qquad (13.2.5)
\end{aligned}$$

Combining (13.2.4) and (13.2.5), we obtain

$$\begin{aligned} p_j &= \frac{(n-2j-s)_2}{(n-2j)_2} \cdot \exp\left[X_j \cdot \frac{\binom{n-2j-s}{2} - \binom{n-2j}{2}}{\binom{n-2j-s}{2} \cdot \binom{n-2j}{2}} + o(1/n) \right] \\ &= \frac{(n-2j-s)_2}{(n-2j)_2} \cdot \exp\left[\frac{X_j \cdot O(s/n)}{\binom{n-2j}{2}} + o(1/n) \right] \\ &= \frac{(n-2j-s)_2}{(n-2j)_2} \cdot \exp(o(1/n)). \end{aligned} \tag{13.2.6}$$

Finally, since the right hand side of (13.2.6) does not depend on X_j and X'_j, we obtain that

$$\Pr\left[\mathcal{E}_j | \mathcal{E}_0 \wedge \cdots \wedge \mathcal{E}_{j-1}\right] = \frac{(n-2j-s)_2}{(n-2j)_2} \cdot \exp(o(1/n)). \tag{13.2.7}$$

As a consequence, recalling that $k = tn$, we get

$$\begin{aligned} &\Pr\left[\mathcal{E}_j \text{ for all } 0 \le j < k\right] = \prod_{0 \le j < k} \Pr\left[\mathcal{E}_j | \mathcal{E}_0 \wedge \cdots \wedge \mathcal{E}_{j-1}\right] \\ = &\prod_{0 \le j < k} \left(\frac{(n-s-2j)(n-s-2j-1)}{(n-2j)(n-2j-1)} \cdot \exp(o(1/n)) \right) \\ \sim &\frac{(n-s)_{2tn}}{(n)_{2tn}} = \binom{n-s}{2tn} \binom{n}{2tn}^{-1}. \end{aligned} \tag{13.2.8}$$

Further, estimating the right hand side of (13.2.8) via Stirling's formula (3.4.1) and letting $y = s/n$, we get

$$\begin{aligned} &\Pr\left[\mathcal{E}_j \text{ for all } 0 \le j < k\right] \\ &\sim \left(\frac{n-s}{2tn}\right)^{2tn} \left(\frac{n-s}{n-s-2tn}\right)^{n-s-2tn} \left(\frac{2tn}{n}\right)^{2tn} \left(\frac{n-2tn}{n}\right)^{n-2tn} \\ &= (1-2t)^s \cdot \left(1 - \frac{s}{n}\right)^{n-s} \cdot \left(1 + \frac{s}{(1-2t)n-s}\right)^{(1-2t)n-s} \\ &= (1-2t)^s \exp\left[n\left((1-y)\ln(1-y) + (1-2t-y)\ln(1+y/(1-2t-y))\right)\right]. \end{aligned} \tag{13.2.9}$$

Now, Taylor expanding, we obtain

$$\begin{aligned} \kappa(y) &= (1-y)\ln(1-y) + (1-2t-y)\ln\left[1 + \frac{y}{1-2t-y}\right] \\ &= \sum_{k=2}^{\infty} \frac{1}{k(k-1)} \left(\frac{(1-2t)^{k-1} - 1}{(1-2t)^{k-1}} \right) y^k. \end{aligned} \tag{13.2.10}$$

Moreover, Taylor series of $(1-2t)^{k-1}$ together with the Lagrange remainder gives

$$(1-2t)^{k-1} = 1 - 2(k-1)t + t^2\eta_k, \tag{13.2.11}$$

where $|\eta_k| \leq 2(k-1)(k-2)$. Since $1-2t \geq \delta$ is bounded away from 0, $y = o(1)$, and $ty^2 = o(1/n)$, plugging (13.2.11) into (13.2.10), we conclude that

$$\begin{aligned} |\kappa(y)| &\leq \sum_{k\geq 2} \frac{y^{k-2}}{\delta^{k-1}} \left(\frac{2ty^2}{k} + \frac{2(k-1)(k-2)}{k(k-1)} y^2 t^2 \right) \\ &\leq 10\delta^{-1} t y^2 \sum_{k\geq 0} \left(\frac{y}{\delta}\right)^k = o(1/n). \end{aligned}$$

Therefore (13.2.9) yields

$$\Pr\left[\mathcal{E}_j \text{ for all } 0 \leq j < k\right] \sim (1-2t)^s \exp(n\kappa(y)) \sim (1-2t)^s,$$

as desired. □

Combining Propositions 13.2.1 and 13.2.2, we can estimate the number $Y_k(M)$ of isolated cycles of length k in G_M with $M = n + tn$ for $t > 0$ (k even): each such isolated cycle results from an isolated cycle of G_n that remained untouched during steps $n+1, n+2, \ldots, M$.

Corollary 13.2.3. *Let $0 \leq t < 1/2$, and let $Y_k(M)$ denote the number of isolated cycles of length k in G_M with $M = n + tn$ for even k. For an arbitrarily large but constant k_0 the random variables $(Y_k)_{k=4,\ldots,k_0,\text{ even}}$ are asymptotically independent Poisson with means $((1-2t)^k/k)_{k=4,\ldots,k_0,\text{ even}}$. Furthermore, for any even $4 \leq k = o(\sqrt{n/t})$ we have $\mathbb{E}_{\mathcal{M}}(C_k) \sim (1-2t)^k/k$.*

Proof. To prove that $(Y_k)_{4\leq k\leq k_0,\text{ even}}$ are asymptotically independent Poisson variables, we will show that

$$\lim_{n\to\infty} \mathbb{E}\left[\prod_{4\leq k\leq k_0,\text{ even}} (Y_k(M))_{r_k}\right] - \prod_{4\leq k\leq k_0,\text{ even}} \left(\frac{(1-2t)^k}{k}\right)^{r_k} = 0; \quad (13.2.12)$$

then the assertion follows from Theorem 3.4.1. To establish (13.2.12), we need to expand the factorial moment. Let $r = r_4 + r_6 + \cdots + r_{k_0}$. Moreover, for $k = 1, \ldots, r$ we let $\zeta_k = 2\max\{l \geq 1 : \sum_{i=2}^{l} r_{2i} \leq k\} + 2$. Then

$$\mu = \mathbb{E}\left[\prod_{4\leq k\leq k_0,\text{ even}} (Y_k(M))_{r_k}\right]$$

is the expected number of tuples $(S_1, \ldots, S_r)$ of distinct isolated cycles in G_M such that the k-th cycle S_k has length ζ_k. Thus, we let $\mathcal{S}$ be the set of all tuples $(S_1, \ldots, S_r)$ of pairwise disjoint subsets of V such that $|S_k| = \zeta_k$. Further, we say that $(S_1, \ldots, S_r)$ in $\mathcal{S}$ is *valid* in G_M if each S_k is an isolated cycle of length ζ_k in G_M. Then

$$\mu = \sum_{(S_1,\ldots,S_r)\in\mathcal{S}} \Pr\left[(S_1, \ldots, S_r) \text{ is valid in } G_M\right]. \quad (13.2.13)$$

Now, $(S_1, \ldots, S_r) \in \mathcal{S}$ is valid in G_M iff $(S_1, \ldots, S_M)$ was valid in G_n and none of the additional edges $E(G_M)\setminus E(G_n)$ is incident with a vertex in $S = \bigcup_{k=1}^{r} S_k$. Therefore, Proposition 13.2.2 implies in combination with (13.2.13) that

$$\mu = \sum_{(S_1,\ldots,S_r)\in\mathcal{S}} \Pr\left[(S_1, \ldots, S_r) \text{ is valid in } G_n\right] \cdot (1-2t)^{|S|} + o(1). \quad (13.2.14)$$

Furthermore, as $Y_4(G_n), \dots, Y_{k_0}(G_n)$ are asymptotically independent Poisson by Proposition 13.2.1,

$$\begin{aligned}\sum_{(S_1,\dots,S_r)\in\mathcal{S}} &\Pr\left[(S_1,\dots,S_r) \text{ is valid in } G_n\right] \\ &= \mathbb{E}\left[\prod_{4\le k\le k_0,\ \text{even}} (Y_k(n))_{r_k}\right] \\ &= o(1) + \prod_{4\le k\le k_0,\ \text{even}} k^{-r_k}. \end{aligned} \tag{13.2.15}$$

Combining (13.2.14) and (13.2.15) and observing that $|S| = \sum_{4\le k\le k_0,\ \text{even}} kr_k$, we obtain (13.2.12).

To compute $\mathbb{E}_{\mathcal{M}}(Y_k(n))$ for even $4 \le k = o(\sqrt{n/t})$, we consider $G = G_{n/2}+\rho$, where ρ is a $G_{n/2}$-configuration chosen uniformly at random (cf. Proposition 13.1.7). Then every cycle C of length k in G corresponds to precisely k sequences $(v_1,\dots,v_k)$ in V^k such that $\{v_i, v_{i-1}\} \in E(G_{n/2})$ for even $2 \le i \le k$ (because there are exactly k ways to choose the first vertex $v_1 \in C$). Further, in total there are $\prod_{j=0}^{k/2-1}(n-2j)$ such sequences $(v_1,\dots,v_k)$, because for even $2 \le i \le k$ the vertex v_i is determined by v_{i-1} (and $G_{n/2}$). Thus, the set $\mathcal{C}_k$ of all possible cycles of length k has cardinality $k^{-1}\prod_{j=0}^{k/2-1}(n-2j)$. Moreover, for each $C \in \mathcal{C}_k$ we have

$$\Pr\left[C \text{ occurs in } \rho\right] \sim (n-k-1)!!/(n-1)!!.$$

For given that C occurs in ρ, by Corollary 13.1.2 there are $(1+o(1))\exp(-\frac{1}{2})(n-k-1)!!$ ways to choose a configuration on $V-C$, while the total number of $G_{n/2}$-configurations is $(1+o(1))\exp(-\frac{1}{2})(n-1)!!$. Hence, if $t = 0$, then

$$\mathbb{E}_{\mathcal{M}}(Y_k(n)) \sim \frac{(n-k-1)!!\cdot\prod_{j=0}^{k/2-1}(n-2j)}{k\cdot(n-1)!!} \sim k^{-1}. \tag{13.2.16}$$

Finally, if $0 < t < \frac{1}{2}$, then each isolated cycle C in G_M was already a cycle in G_n and remained isolated until step M. Hence, (13.2.16) implies in combination with Proposition 13.2.2 that $\mathbb{E}_{\mathcal{M}}(Y_k(M)) = (1+o(1))k^{-1}(1-2t)^k$.

□

13.3 Component structure

When $t > 0$ the random graph G_M with $M = n + tn$ consists of one giant component on $(1-o(1))n$ vertices and a number of "short" isolated cycles. More precisely the following holds.

Proposition 13.3.1. *Suppose that $n^{-1}\ln^4 n \le t \le \frac{1}{2} - \delta$ for an arbitrarily small but constant $\delta > 0$. Then G_M consists of precisely one component of order $(1-o(1))n$ and isolated cycles of length $O(t^{-1}\ln n)$* a.a.s.

To prove Proposition 13.3.1, we need an alternative description of the random graph G_M for $M > n$. An *M-configuration* is a set $R \subset V$ of cardinality

$|R| = 2(M-n)$ together with a map $\rho : R \to R$ such that $\rho(v) \neq v$ for all $v \in R$ and $\rho \circ \rho = \mathrm{id}$. In other words, ρ is a perfect matching of R. Moreover, given a state G_n of a min-min process at step n, we say that an M-configuration (R, ρ) is a (G_n, M)-*configuration* if $\{v, \rho(v)\} \notin E(G_n)$ for all $v \in R$. Hence, the matching ρ describes a set of edges that can be added to G_n in order to obtain the state G_M at step M.

Let G_* be a 2-regular bipartite graph. If (R, ρ) is a (G_*, M)-configuration, then let $P(G_*, R, \rho)$ denote the probability that $E(G_M) \setminus E(G_n) = \{\{v, \rho(v)\} : v \in R\}$ given that $G_n = G_*$. The following lemma establishes a connection between the uniform distribution on the set of all (G_*, M)-configurations and the distribution given by $P(G_*, R, \rho)$.

Lemma 13.3.2. *Let G_* be a 2-regular bipartite graph, and let (R, ρ), (R', ρ') be two (G_*, M)-configurations. Then there are constants $0 < c_1 \leq c_2$ such that*

$$c_1 P(G_*, R, \rho) \leq P(R', \rho') \leq c_2 P(G_*, R, \rho).$$

Note that M-configurations are considerably more convenient to work with than (G_n, M)-configurations. So the following lemma will be useful.

Lemma 13.3.3. *Suppose that $M = (1+t)n$, where $0 \leq t < \frac{1}{2}$. Let G_* be any 2-regular bipartite graph, and let $R \subset V$, $4 \leq r = |R| = 2(M-n)$. Let ρ be a random perfect matching of R. Then there is a constant $\varepsilon > 0$ such that the probability that (R, ρ) is a (G_*, M)-configuration is $\geq \varepsilon$.*

Proof of Proposition 13.3.1. Suppose that $n^{-1} \ln^4 n \leq t \leq \frac{1}{2} - \delta$ for some constant $\delta > 0$, and let $M = (1+t)n$. While our goal is to show that *a.a.s.* the components of G_M on $\leq \frac{n}{2}$ vertices are cycles of length $O(t^{-1} \ln n)$, Lemmas 13.3.2 and 13.3.3 imply that we just need to prove that this statement holds *a.a.s.* for $G_n + \rho$, where (R, ρ) is an M-configuration chosen uniformly at random and $G_n + \rho = \{\{v, \rho(v)\} : v \in R\}$.

We shall first prove that $G_n + \rho$ has no component of order bigger than $ct^{-1} \ln n$ and smaller than $\frac{n}{2}$ *a.a.s.*, where $c > 0$ is a sufficiently large constant. Thus, let $\mathcal{S}$ be a set of components of G_n, let S be the set of all vertices that belong to components in $\mathcal{S}$, and suppose that $ct^{-1} \ln n \leq s = |S| \leq \frac{n}{2}$. We are to bound the probability that ρ melts the components $\mathcal{S}$ into a single component of $G_n + \rho$.

Let $p_{t,s,l}$ be the probability that precisely l edges of ρ connect two vertices of S, while the $tn - l$ remaining edges of ρ connect two vertices of $V \setminus S$. Then

$$p_{t,s,l} = \frac{\binom{s}{2l}\binom{n-s}{2(tn-l)}(2l-1)!!(2(tn-l)-1)!!}{\binom{n}{2tn}(2tn-1)!!} \ ; \tag{13.3.1}$$

for the total number of M-configurations (R, ρ) equals $\binom{n}{2tn}(2tn-1)!!$, as there are $\binom{n}{2tn}$ ways to choose the set R, and then $(2tn-1)!!$ ways to choose the perfect matching ρ. Similarly, the number of pairs (R', ρ') (resp. (R'', ρ'')) such that $R' \subset S$, $|R'| = 2l$ (resp. $R'' \subset V \setminus S$, $|R''| = 2(tn-l)$) and ρ' is a perfect matching of R (resp. ρ'' of R'') equals $\binom{s}{2l}(2l-1)!!$ (resp. $\binom{n-s}{2(tn-l)}(2(tn-l)-1)!!$).

Applying (3.4.2) to (13.3.1), we get

$$\begin{aligned} p_{t,s,l} &= \frac{(2tn)!}{(2l)!(2tn-2l)!} \cdot \frac{(s)_{2l}(n-s)_{2(tn-l)}}{(n)_{2tn}} \cdot \Theta\left[\left(\frac{l}{tn}\right)^l \left(1-\frac{l}{tn}\right)^{tn-l}\right] \\ &\leq \Theta(1) \cdot \binom{2tn}{2l} \left(\frac{s}{n}\right)^{2l} \left(1-\frac{s}{n}\right)^{2(tn-l)} \left(\frac{l}{tn}\right)^l \left(1-\frac{l}{tn}\right)^{tn-l}. \quad (13.3.2) \end{aligned}$$

Furthermore, by the Chernoff bounds (3.4.6)

$$\sum_{l:|l-st|\geq \frac{1}{10}st} \binom{2tn}{2l} \left(\frac{s}{n}\right)^{2l} \left(1-\frac{s}{n}\right)^{2(tn-l)} \leq \exp(-\Omega(st)). \quad (13.3.3)$$

In addition, if $\gamma = l - st$ has absolute value $< \frac{1}{10}st$, then our assumption $s \leq n/2$ entails that

$$\left(\frac{l}{tn}\right)^l = \left(\frac{s+\gamma/t}{n}\right)^l \leq \left(\frac{2}{3}\right)^{st/2} \leq \exp(-\Omega(st)). \quad (13.3.4)$$

Hence, plugging (13.3.3) and (13.3.4) into (13.3.2), we conclude that

$$p_{t,s,l} \leq \exp(-\Omega(st)) \leq n^{-3}$$

for all l and s, provided that $st \geq c \ln n$ for a large enough constant $c > 0$. Therefore, we get

$$\Pr[S \text{ is a component of } G_n + \rho] \leq \sum_{l=0}^{tn} p_{t,s,l} \leq n^{-2}. \quad (13.3.5)$$

Finally, by Proposition 13.2.1 G_n has at most $K \leq \ln n$ components *a.a.s.* Thus, there are at most $2^K \leq n$ ways to choose a set $\mathcal{S}$ such that $ct^{-1} \ln n \leq s = |S| \leq \frac{n}{2}$. Consequently, due to the union bound (13.3.5) implies that $G_n + \rho$ has no component of order $ct^{-1} \ln n \leq s \leq \frac{n}{2}$ *a.a.s.*

To show that *a.a.s.* all components of $G_{(1+t)n}$ are cycles, let $r \geq 2$ be an integer. Let $S_1, \ldots, S_r$ be components of G_n that contain $s \leq ct^{-1} \ln n$ vertices in total. Then our assumption $t \gg n^{-1} \ln^3 n$ entails in combination with (13.3.2) that

$$\sum_{r-1\leq l\leq tn/2} p_{t,s,l} \leq \sum_{r-1\leq l\leq tn/2} \left(\frac{l}{tn}\right)^l \ll \left(\frac{r-1}{\ln^3 n}\right)^{r-1}, \quad (13.3.6)$$

$$\sum_{l>tn/2} p_{t,s,l} \leq \sum_{l>tn/2} \binom{2tn}{2l} \left(\frac{s}{n}\right)^{2l} \left(1-\frac{s}{n}\right)^{2(tn-l)} \quad (13.3.7)$$

$$\overset{(3.4.6)}{\leq} \exp(-\Omega(tn)) \leq \exp(-\ln^3 n). \quad (13.3.8)$$

Let K be the total number of components of G_n. Then $K \leq \ln n$ *a.a.s.* by Proposition 13.2.1, so that there are at most $K^r \leq (\ln n)^r$ ways to choose the components $S_1, \ldots, S_r$. Hence, (13.3.6) and (13.3.8) imply that the probability that G_M has a component of order $\leq ct^{-1} \ln n$ that consists of several cycles is at most

$$\Pr[K > \ln n] + \sum_{2\leq r\leq \ln n} (\ln n)^r \left(\left(\frac{r-1}{\ln^3 n}\right)^{r-1} + \exp(-\ln^3 n)\right) = o(1).$$

Furthermore, since in the previous paragraph we showed that $G_n + \rho$ has no component of order $ct^{-1}\ln n \le s \le \frac{n}{2}$ *a.a.s.*, we conclude that all components of order $\le \frac{n}{2}$ are cycles *a.a.s.*

Finally, since *a.a.s.* $G_n+\rho$ has no component of order $ct^{-1}\ln n \le s \le \frac{n}{2}$, and since the total number of components is $K \le \ln n$ *a.a.s.* by Proposition 13.2.1, at most $ct^{-1}\ln n \cdot K = o(n)$ vertices belong to components of order $\le \frac{n}{2}$. Therefore, *a.a.s.* there is a component of order $(1-o(1))n$. □

Proof of Lemma 13.3.2. Let G_* be a bipartite 2-regular graph, and let $m = M - n$. Remember that we are assuming that $m \le (\frac{1}{2} - \delta)n$ for a fixed $\delta > 0$. If (R, ρ) is a (G_*, M)-configuration, then we let $E(R, \rho) = \{\{v, \rho(v)\} : v \in R\}$.

To prove Lemma 13.3.2, we derive upper and lower bounds on $P(G_*, R, \rho)$ that hold uniformly for all (G_*, M)-configurations (R, ρ). Let $\sigma : \{1, \dots, m\} \to E(R, \rho)$ be a bijection. Moreover, let $\mathcal{E}_j(\sigma)$ denote the event that $E(G_{n+j}) \setminus E(G_n) = \{\sigma(1), \dots, \sigma(j)\}$ $(1 \le j \le m)$, and let $\mathcal{E}_0(\sigma)$ be the event that $G_n = G_*$. In addition, let $p_j(R, \rho, \sigma) = \Pr[\mathcal{E}_j(\sigma) | \mathcal{E}_i(\sigma) \text{ for } 0 \le i < j]$.

Suppose that $0 \le j < m$. Let X_j be the number of edges connecting two vertices of degree 2 in G_{n+j}, and let $T_j = \binom{n-2j}{2} - X_j$ denote the number of possible edges that can be added to G_{n+j} in order to obtain G_{n+j+1}. Let us call an edge $e = \{v, w\}$ G_{n+j}-*admissible* if $e \notin E(G_{n+j})$ and v, w are distinct vertices of degree 2 in G_{n+j}. Then the probability that a specific G_{n+j}-admissible edge e is chosen to obtain G_{n+j+1} from G_{n+j} equals T_j^{-1}. Furthermore, since G_{n+j} has exactly $n - 2j$ vertices of degree 2, T_j can be bounded uniformly as follows:

$$\binom{n-2j}{2} - 2n \le T_j \le \binom{n-2j}{2}. \tag{13.3.9}$$

Now, given that the events $\mathcal{E}_i(\sigma)$, $0 \le i \le j$, occur, the edge $\sigma(j)$ is G_{n+j}-admissible. Therefore, (13.3.9) entails that

$$\binom{n-2j}{2}^{-1} \le p_{j+1}(R, \rho, \sigma) \le \left[\binom{n-2j}{2} - 2n\right]^{-1}. \tag{13.3.10}$$

Further, with respect to the probability $p(R, \rho, \sigma) = \prod_{j=1}^{m} p_j(R, \rho, \sigma)$ that in *each* step $1 \le j \le m$ the edge $\sigma(j)$ is added to G_{n+j-1} given that $G_n = G_*$ the estimate (13.3.10) yields

$$\prod_{j=1}^{m} \binom{n-2j}{2}^{-1} \le p(R, \rho, \sigma) \le \prod_{j=1}^{m} \left[\binom{n-2j}{2} - 2n\right]^{-1}. \tag{13.3.11}$$

Since $m \le (\frac{1}{2} - \delta)n$, a straightforward computation shows that there is a constant $c_3 > 0$ such that $\prod_{j=1}^{m} \binom{n-2j+2}{2} / (\binom{n-2j+2}{2} - 2n) \le c_3$. Hence, letting

$$q = \prod_{j=1}^{m} \binom{n-2j+2}{2}^{-1},$$

due to (13.3.11) we get

$$q \le p(R, \rho, \sigma) \le c_3 q \quad \text{for all triples } (R, \rho, \sigma). \tag{13.3.12}$$

Finally, since $P(G_*, R, \rho) = \sum_\sigma p(R, \rho, \sigma)$, where the sum ranges over all maps $\sigma : \{1, \dots, m\} \to E(R, \rho)$, (13.3.12) implies

$$(m/2)!q \le p(G_*, R, \rho) \le c_3(m/2)!q.$$

As the neither the upper nor the lower bound depends on (R, ρ), we have established Lemma 13.3.2.

□

Proof of Lemma 13.3.3. Let e be the number of edges $\{v, w\}$ of G_* such that $v, w \in R$, and let Z be the number of edges $\{v, \rho(v)\} \in E(G_*)$. Then (R, ρ) is a (G_*, M)-configuration iff $Z = 0$. If $r \le 10$, then clearly $\Pr[Z = 0] > \varepsilon_1$ for a certain constant $\varepsilon_1 > 0$. Thus, assume that $r \ge 12$. Then by inclusion and exclusion

$$\Pr[Z = 0] \ge 1 - \mathbb{E}(Z) + \frac{1}{2}\mathbb{E}(Z(Z-1)) - \frac{1}{6}\mathbb{E}(Z(Z-1)(Z-2)). \quad (13.3.13)$$

Furthermore,

$$\begin{aligned} \mathbb{E}(Z) &= \frac{e}{r-1}, \\ \mathbb{E}(Z(Z-1)) &\ge \frac{e(e-3)}{(r-1)(r-3)}, \\ \mathbb{E}(Z(Z-1)(Z-2)) &\le \frac{e^3}{(r-1)(r-2)(r-3)}. \end{aligned} \quad (13.3.14)$$

Plugging (13.3.14) into (13.3.13) and setting $e = r$ to its maximum value, we conclude that

$$\Pr[Z = 0] \ge 1 - \frac{r}{2(r-1)} - \frac{r^3}{(r-1)(r-3)(r-5)} \ge 0.01,$$

because $r \ge 12$. Hence, letting $\varepsilon = \min\{0.01, \varepsilon_1\}$, we obtain the desired result.

□

13.4 Connectedness

In this section we will show that G_M is connected with positive probability as soon as $M = (1 + t)n$ for an arbitrarily small but fixed $t > 0$. In fact, $\Pr[G_M \text{ is connected}]$ lies strictly between 0 and 1 if $0 < t < \frac{1}{2}$. More precisely, the following holds.

Theorem 13.4.1. *Let $M = (1 + t)n$.*

(i) If $-1 \le t \le 0$, then G_M is disconnected a.a.s.

(ii) Suppose that $0 < t < 1/2$ remains fixed as $n \to \infty$. Then G_M a.a.s. *has a giant component consisting of $> \frac{n}{2}$ vertices. Furthermore, the number Y of components of order $< \frac{n}{2}$ is asymptotically Poisson with mean $\mu_t = \frac{1}{2}(-(1-2t)^2 - \ln(4(t-t^2)))$. That is,*

$$\lim_{n\to\infty} \Pr[Y = k] = \mu_t^k \exp(-\mu_t)/k! \quad \text{for any } k.$$

In particular, letting

$$\Psi(t) = \exp(-\mu_t) = 2\ \sqrt{t-t^2}\ \exp((2t-1)^2/2), \tag{13.4.1}$$

we have $\lim_{n\to\infty} \Pr(G_M \text{ is connected}) = \lim_{n\to\infty} \Pr(Y=0) = \Psi(t)$.

(iii) If $t \geq 1/2$, *then* G_M *is connected* a.a.s.

Proof. As we have seen in the introduction, G_M simply consists of isolated vertices and edges when $-1 < t < -1/2$, and it consists of isolated paths and of isolated cycles of even length when $-1/2 \leq t < 0$. Thus G_M is *a.a.s.*disconnected when $-1 \leq t < 0$.

Proposition 13.2.1 implies that $\lim_{n\to\infty} \Pr[G_n \text{ is connected}] = 0$. Indeed, given $\varepsilon > 0$, choose k_0 so large that $\sum_{k=1}^{k_0} (2k)^{-1} \geq -\ln(\varepsilon/2)$. Then by Proposition 13.2.1

$$\begin{aligned} \Pr[G_n \text{ is connected}] &\leq \Pr[C_{2k} = 0 \text{ for } k = 1, \ldots, k_0] \\ &\leq (1+o(1)) \exp\left[-\sum_{k=1}^{k_0} \frac{1}{2k}\right] \leq (1+o(1))\frac{\varepsilon}{2} < \varepsilon \end{aligned}$$

for all sufficiently large n. Hence, for all $t \leq 0$ we have

$$\lim_{n\to\infty} \Pr[G_M \text{ is connected}] = 0.$$

Let $0 < t < \frac{1}{2}$ be constant, and let $\varepsilon > 0$ be an arbitrarily small constant. Then there is a number k_0 such that $\sum_{k>k_0} (1-2t)^k/k < \varepsilon$, so that by Corollary 13.2.3 and Markov's inequality the probability p_{k_0} that G_M contains an isolated cycle of length $> k_0$ is $\leq \varepsilon$. Furthermore, by Corollary 13.2.3 the total number Y of cycles of length $4 \leq k \leq k_0$ (k even) is asymptotically Poisson with mean

$$\begin{aligned} \mathbb{E}(Y) &= \sum_{4\leq k\leq k_0,\ \text{even}} \frac{(1-2t)^k}{k} \leq \sum_{4\leq k,\ \text{even}} \frac{(1-2t)^k}{k} \\ &\overset{(13.5.5)}{=} \frac{1}{2}\left[-(1-2t)^2 - \ln(-4(t^2-t))\right]. \end{aligned}$$

Conversely, since $\sum_{k>k_0} (1-2t)^k/k < \varepsilon$, we have

$$\mathbb{E}(Y) \geq \frac{1}{2}\left[-(1-2t)^2 - \ln(-4(t^2-t))\right] - \varepsilon.$$

Therefore,

$$1 \leq \frac{\exp[-\mathbb{E}(Y)]}{2\exp\left(\frac{1}{2}(1-2t)^2\right)[t(1-t)]^{1/2}} \leq \exp(\varepsilon) \leq 1+2\varepsilon, \tag{13.4.2}$$

provided that ε is small enough. As Y is asymptotically Poisson, we have $\Pr[Y=0] \sim \exp(-\mathbb{E}(Y))$, so that (13.4.2) yields

$$|\Pr[Y=0] - \Psi(t)| \leq 2\varepsilon + o(1),$$

where Ψ denotes the function defined in (13.4.1). Hence, Proposition 13.3.1 implies that

$$|\Pr[G_M \text{ is connected}] - \Psi(t)| \le |\Pr[Y = 0] - \Psi(t)| + p_{k_0} + o(1) \le 3\varepsilon + o(1).$$

Thus, $\Pr[G_M \text{ is connected}]$ converges to $\Psi(t)$ as $n \to \infty$.

Finally, the probability that G_M is connected is an increasing function of t. Hence, as $\lim_{t\to\frac{1}{2}} \Psi(t) = 1$, we conclude that G_M is connected *a.a.s.* if $t \ge \frac{1}{2}$. □

13.5 Giant component

Let $X = X(M)$ be the number of vertices outside of the largest component of G_M with $M = n+tn$ for $t > 0$. Then Theorem 13.4.1 shows that $\Pr[X = 0] > 0$ if $t > 0$. The following theorem gives a much more precise result: we can determine the limiting distribution of X as $n \to \infty$ precisely.

Theorem 13.5.1. *Let $M = n+tn$ for a constant $0 < t < 1/2$. Then as $n \to \infty$, $X = X(M)$ converges in distribution to the distribution given by the probability generating function*

$$q(z) \;=\; \sum_{l=0}^{\infty} q_l z^l = \exp\left[\frac{1}{2}(1-2t)^2(1-z^2)\right]\sqrt{\frac{1-(1-2t)^2}{1-(1-2t)^2z^2}}\ .$$

As a consequence, for any positive integer l

$$\begin{aligned}
&\lim_{n\to\infty} \Pr(X = 2l) = q_{2l}\\
=\;& 2\sqrt{t(1-t)}\exp\left[\frac{(1-2t)^2}{2}\right](1-2t)^{2l}\sum_{0\le m\le l}\binom{2m}{m}\frac{(-1)^{l-m}}{2^{l+m}(l-m)!}\\
=\;& (1+O(1/l))\ 2\exp(2t(t-1))\sqrt{\frac{t(1-t)}{\pi l}}(1-2t)^{2l}, \qquad (13.5.1)
\end{aligned}$$

while $\lim_{n\to\infty}\Pr(X = 2l-1) = q_{2l-1} = 0$. *Furthermore,* a.a.s. *all components on $< \frac{n}{2}$ vertices are cycles of even lengths.*

Theorem 13.5.1 states the limiting distribution of $X(M)$ for $M = (1+t)n$ with $t > 0$ bounded away from 0. Our next goal is to investigate the number of vertices outside of the largest component if $t = t(n) = o(1)$ tends to 0 as $n \to \infty$. Note that G_n is 2-regular and thus consists of isolated cycles. If $M = (1+t)n$ for $0 < t = o(1)$, then some of these cycles melt together to form a component consisting of $n - \Theta(t^{-1})$ vertices. Thus, we will have $X = X(M) = \Theta(t^{-1})$ *a.a.s.* In fact, the following theorem gives the precise limiting distribution of tX as $n \to \infty$ in the case that $t \ge \ln^4(n)n^{-1}$ is "not too small".

Theorem 13.5.2. *Suppose that $M = (1+t)n$, where $t = t(n)$ with $n^{-1}\ln^4 n \le t = o(1)$. Then tX converges in distribution to a gamma distribution with both shape and scale parameter equal to $\frac{1}{2}$. That is,*

$$\lim_{n\to\infty}\Pr(tX \le b) = \frac{1}{\sqrt{\pi}}\int_0^{2b}\frac{\exp(-s)}{\sqrt{s}}ds. \qquad (13.5.2)$$

As a consequence, for any $x > 0$,

$$\begin{aligned}\lim_{n\to\infty} \Pr(tX \geq x) &= \frac{\exp(-2x)}{\sqrt{2\pi x}} \sum_{m=0}^{\infty} (-1)^m (2m-1)!! 2^{-2m} x^{-m} \\ &= (1 + O(1/x)) (2\pi x)^{-1/2} \exp(-2x). \end{aligned} \tag{13.5.3}$$

Note that *a.a.s.* all components of G_M except for component are isolated cycles, by Proposition 13.3.1. Therefore, in order to analyse the number X of vertices outside of the largest component we consider the number Y_k of isolated cycles of length k in G_M ($4 \leq k \leq n$, even). In addition, let $\gamma = \gamma(n) = ct^{-1} \ln n$ for a sufficiently large constant $c > 0$; then $\gamma = o(\sqrt{n/t})$, provided that $t \gg n^{-1} \ln^2 n$. Moreover, setting $Y = \sum_{4 \leq k \leq \gamma,\ \text{even}} kY_k$, *a.a.s.* we have $X = Y$ by Proposition 13.3.1.

To investigate Y, let $(Z_k)_{k=4,6,\ldots}$ be a family of mutually independent Poisson variables with means $\mathbb{E}(Z_k) = \lambda_k = (1-2t)^k/k$. Then the characteristic function of Z_k is

$$\mathbb{E}(\exp(\vec{i}yZ_k)) \stackrel{(3.4.3)}{=} \exp\left(\lambda_k(\exp(\vec{i}y) - 1)\right). \tag{13.5.4}$$

We set $Z = \sum_{k \geq 4,\ \text{even}} kZ_k$ and $\lambda = \sum_{k \geq 4,\ \text{even}} \lambda_k = \Phi(1-2t)$, where Φ denotes the power series

$$\Phi(z) = \sum_{4 \leq k \text{ even}} \frac{z^k}{k} = -\frac{1}{2}\left(z^2 + \ln(1-z^2)\right); \tag{13.5.5}$$

the second equality sign follows by just plugging in the Taylor expansion of $x \mapsto \ln(1+x)$.

In order to prove Theorems 13.5.1 and 13.5.2 we need the following results on the characteristic function of X and tX, which we will prove in Section 13.6.

Proposition 13.5.3. *If $0 < t < 1$ is independent of n, then the characteristic function $y \mapsto \mathbb{E}(\exp(\vec{i}yX))$ of X converges pointwise to the function $y \mapsto \psi(y) = \exp(\Phi((1-2t)\exp(\vec{i}y)) - \Phi(1-2t))$, where Φ is the function defined in (13.5.5).*

Proposition 13.5.4. *Suppose that $n^{-1}\ln^4 n \leq t = o(1)$. Then the characteristic function $y \mapsto \mathbb{E}(\exp(\vec{i}ytX))$ of tX converges pointwise to $y \mapsto (1 - \vec{i}y/2)^{-1/2}$.*

Proof of Theorem 13.5.1. Let $q(z)$ be the probability generating function given by

$$\begin{aligned} q(z) &= \exp(\Phi((1-2t)z) - \Phi(1-2t)) \\ &= \exp((1-2t)^2/2 - (1-2t)^2 z^2/2) \sqrt{\frac{1-(1-2t)^2}{1-(1-2t)^2 z^2}}\ . \end{aligned}$$

By Proposition 13.5.3, the characteristic function of X converges pointwise to the characteristic function $y \mapsto \psi(y) = q(\exp(\vec{i}y))$ of the probability distribution described by $q(z)$. Therefore, Theorem 3.4.2 implies that the asymptotic probability distribution of X is given by $q(z)$. That is, letting $q(z) = \sum_{l \geq 0} q_l z^l$ be the power series expansion of $q(z)$, we have $\lim_{n\to\infty} \Pr[X = l] = q_l$ for all l.

To analyse the coefficients q_l, let $c(t) = 2\sqrt{t(1-t)}\exp(\frac{1}{2}(1-2t)^2)$. Then we can rewrite $q(z)$ as

$$q(z) \quad = \quad c(t)\frac{\exp(-(1-2t)^2z^2/2)}{\sqrt{1-(1-2t)^2z^2}}.$$

Since $(1-z)^{-1/2} = \sum_{i\geq 0}\frac{1}{4^i}\binom{2i}{i}z^i$ and $\exp(-z/2) = \sum_{i\geq 0}\frac{(-1/2)^i}{i!}z^i$, we have

$$(1-z)^{-1/2}\exp(-z/2) \quad = \quad \sum_{l\geq 0}\left(\sum_{0\leq m\leq l}\binom{2m}{m}\frac{(-1)^{l-m}}{2^{l+m}(l-m)!}\right)z^l.$$

Thus, we obtain

$$q(z) \quad = \quad c(t)\sum_{l\geq 0}(1-2t)^{2l}\left(\sum_{0\leq m\leq l}\binom{2m}{m}\frac{(-1)^{l-m}}{2^{l+m}(l-m)!}\right)z^{2l}.$$

As a consequence,

$$\lim_{n\to\infty}\Pr(X=2l) \quad = \quad q_{2l} = c(t)(1-2t)^{2l}\left(\sum_{0\leq m\leq l}\binom{2m}{m}\frac{(-1)^{l-m}}{2^{l+m}(l-m)!}\right),$$

while $\lim_{n\to\infty}\Pr(X = 2l+1) = q_{2l+1} = 0$ for all l. Thus, we have established (13.5.1).

□

Proof of Theorem 13.5.2. Proposition 13.5.4 implies in combination with (3.4.5) and Theorem 3.4.2 that the asymptotic probability distribution of tX is a gamma distribution:

$$\begin{aligned}\lim_{n\to\infty}\Pr(a\leq tX\leq b) &= \int_a^b \frac{s^{-1/2}\exp(-2s)}{\Gamma(1/2)(1/2)^{1/2}}ds \\ &= \sqrt{\frac{2}{\pi}}\int_a^b\frac{\exp(-2s)}{\sqrt{s}}ds. \qquad (13.5.6)\end{aligned}$$

Thus, we have established (13.5.2). Furthermore, (13.5.6) implies that for any real $x \geq 0$

$$\lim_{n\to\infty}\Pr(tX\geq x) \quad = \quad \sqrt{\frac{2}{\pi}}\int_x^\infty\frac{\exp(-2s)}{\sqrt{s}}ds = \sqrt{\frac{2}{\pi}}\int_{\sqrt{4x}}^\infty\exp(-s^2/2)ds.$$

Integrating by parts repeatedly, we get

$$\int_a^\infty\exp(-s^2/2)ds = \frac{\exp(-a^2/2)}{a}\sum_{m=0}^\infty\frac{(-1)^m(2m-1)!!}{a^{2m}}$$

and hence

$$\begin{aligned}\lim_{n\to\infty}\Pr(tX\geq x) &= \sqrt{\frac{2}{\pi}}\int_{\sqrt{4x}}^\infty\exp(-s^2/2)ds \\ &= \frac{\exp(-2x)}{\sqrt{2\pi x}}\sum_{m=0}^\infty\frac{(-1)^m(2m-1)!!}{2^{2m}x^m}.\end{aligned}$$

In particular, we conclude that

$$\lim_{n\to\infty} \Pr(tX \geq x) = (1 + O(1/x))\exp(-2x)/\sqrt{2\pi x},$$

thereby proving (13.5.3). □

13.6 Characteristic functions

In this section we will prove Propositions 13.5.3 and 13.5.4 on characteristic functions.

Proof of Proposition 13.5.3. Suppose that $0 < t < \frac{1}{2}$ is independent of n. Let $y_0 \in R \setminus \{0\}$ be arbitrary but fixed, and $\alpha > 0$ be arbitrarily small but fixed. Then our goal is to show that there exists an $n_0 = n_0(\alpha, y_0)$ such that

$$|\mathbb{E}(\exp(\vec{i}y_0X)) - \psi(y_0)| \leq 3\alpha \quad \text{if } n \geq n_0. \tag{13.6.1}$$

Since the characteristic function of Z is

$$\begin{aligned}
\mathbb{E}(\exp(\vec{i}yZ)) &= \mathbb{E}\left[\exp\left(\sum_{4\leq k,\ \text{even}} \vec{i}kyZ_k\right)\right] \\
&= \prod_{4\leq k\ \text{even}} \mathbb{E}\left[\exp\left(\vec{i}kyZ_k\right)\right] \\
&\overset{(13.5.4)}{=} \prod_{4\leq k\ \text{even}} \exp\left[\lambda_k\left(\exp(\vec{i}ky) - 1\right)\right] \\
&= \exp\left[-\lambda + \sum_{k\geq 4,\ \text{even}} \lambda_k \exp(\vec{i}ky)\right] \\
&= \psi(y),
\end{aligned}$$

(13.6.1) is equivalent to

$$|\mathbb{E}(\exp(\vec{i}y_0X)) - \mathbb{E}(\exp(\vec{i}y_0Z))| \leq 3\alpha. \tag{13.6.2}$$

To establish (13.6.2), we choose a number $K = K(\alpha, y_0)$ such that

$$\sum_{k>K,\ \text{even}} \lambda_k = \sum_{k>K,\ \text{even}} (1-2t)^k/k < \alpha/2; \tag{13.6.3}$$

such a number K exists because $0 < t < \frac{1}{2}$ is constant. Set

$$Y' = \sum_{4\leq k\leq K,\ \text{even}} kY_k, \quad Z' = \sum_{4\leq k\leq K,\ \text{even}} kZ_k.$$

Since by Corollary 13.2.3 $(Y_k)_{4\leq k\leq K,\ \text{even}}$ are asymptotically independent Poisson variables with means $(\lambda_k)_{4\leq k\leq K,\ \text{even}}$, Y' converges to Z' in distribution. Therefore, Theorem 3.4.2 entails that the characteristic function of Y', i.e., $y \mapsto \mathbb{E}(\exp(\vec{i}yY'))$, converges pointwise to that of Z', i.e., $y \mapsto \mathbb{E}(\exp(\vec{i}yZ'))$.

In addition, we shall derive from (13.6.3) that Z' is a good approximation of Z, and from Corollary 13.2.3 and (13.6.3) that Y' approximates Y and hence X well. Due to (13.6.3), $\Pr[Z \neq Z'] \leq \sum_{k>K,\text{ even}} \Pr[Z_k > 0] \leq \sum_{k>K,\text{ even}} \lambda_k < \alpha$. Because of Proposition 13.1.7, Corollary 13.2.3, and (13.6.3), we have

$$\begin{aligned}\Pr[Y \neq Y'] \leq \Pr_{\mathcal{M}}(Y \neq Y') + o(1) &\leq o(1) + \sum_{k>K,\text{ even}} k\Pr_{\mathcal{M}}(Y_k > 0) \\ &\leq o(1) + \sum_{k>K,\text{ even}} \lambda_k \leq 2\alpha/3,\end{aligned}$$

provided that n is sufficiently large. Further, Proposition 13.3.1 entails that $X = Y$ *a.a.s.*, so that $\Pr[X \neq Y'] \leq \Pr[Y \neq Y'] + \Pr[X \neq Y] \leq \alpha$ if n is large enough.

Finally, applying Lemma 3.4.3, we conclude that

$$\begin{aligned}&|\mathbb{E}\left[\exp(\vec{i}y_0X)\right] - \mathbb{E}\left[\exp(\vec{i}y_0Z')\right]|F \\ \leq\ &|\mathbb{E}\left[\exp(\vec{i}y_0X) - \exp(\vec{i}y_0Y)\right]| + |\mathbb{E}\left[\exp(\vec{i}y_0Y')\right] - \mathbb{E}\left[\exp(\vec{i}y_0Z')\right]| \\ &+|\mathbb{E}\left[\exp(\vec{i}y_0Z) - \exp(\vec{i}y_0Z')\right]| \\ \leq\ &3\alpha,\end{aligned}$$

thereby establishing (13.6.2). □

In order to study $q_{2l} = \Pr(X = 2l)$ for large l, we apply the principles of singularity analysis [63] to $r(z) = (1-z)^{-1/2}\exp(-z/2)$. Let $r(z) = \sum_{l\geq 0} r_l z^l$ be the power series expansion of $r(z)$. Then the asymptotics of r_l result from a dominant singularity of $r(z)$. Since $\exp(-z/2)$ has no singularity, whereas $(1-z)^{-1/2}$ has a dominant singularity at $z = 1$, the asymptotic expansion of $r(z)$ near $z = 1$ is obtained by the analytic expansion of $\exp(-z/2)$ at $z = 1$, namely

$$\exp(-z/2) = \exp(-1/2)\sum_{i\geq 0}\frac{1}{2^i i!}(1-z)^i,$$

multiplied by $(1-z)^{-1/2}$, that is,

$$\begin{aligned}r(z) &= \exp(-1/2)\sum_{i\geq 0}\frac{1}{2^i i!}(1-z)^{i-1/2} \\ &= \exp(-1/2)(1-z)^{-1/2} + O((1-z)^{1/2}).\end{aligned}$$

Using the following expansions

$$(1-z)^{-1/2} = \sum_{i\geq 0}\frac{1}{4^i}\binom{2i}{i}z^i \quad\text{and}\quad (1-z)^{1/2} = -\sum_{i\geq 1}\frac{2}{i4^i}\binom{2i-2}{i-1}z^i,$$

we obtain

$$r_l = \exp(-1/2)\frac{1}{4^l}\binom{2l}{l}(1+O(1/l)) \overset{(3.4.1)}{=} \exp(-1/2)\frac{1}{\sqrt{\pi l}}(1+O(1/l)).$$

Therefore, we conclude that

$$\begin{aligned}\lim_{n\to\infty}\Pr(X=2l) &= \frac{c(t)}{\exp(1/2)\sqrt{\pi l}}(1-2t)^{2l}(1+O(1/l))\\ &= 2\exp(2t(t-1))\sqrt{\frac{t-t^2}{\pi l}}(1-2t)^{2l}(1+O(1/l)),\end{aligned}$$

thereby proving (13.5.1).

Proof of Proposition 13.5.4. Suppose that $n^{-1}\ln^4 n \le t = o(1)$. Let $y_0 \in R\backslash\{0\}$ and $\alpha > 0$ be given. Our aim is to show that

$$|(1-\vec{i}y_0/2)^{-1/2} - \mathbb{E}(\exp(\vec{i}ty_0X))| \le 8\alpha \tag{13.6.4}$$

if $n \ge n_0$ for a large enough $n_0 = n_0(\alpha, y_0)$. In order to establish (13.6.4), we first prove that $\mathbb{E}(\exp(\vec{i}ty_0Z))$ is close to $(1-\frac{\vec{i}}{2}y_0)^{-1/2}$ if $n \ge n_0$ is large enough. Then, we shall compare $\mathbb{E}(\exp(\vec{i}ty_0Z))$ and $\mathbb{E}(\exp(\vec{i}ty_0X))$.

We have

$$\begin{aligned}&\mathbb{E}\left[\exp(\vec{i}ty_0Z)\right] = \prod_{4\le k,\ \text{even}} \mathbb{E}\left[\exp(\vec{i}ty_0kZ_k)\right]\\ =\ &\exp\left[-\lambda + \sum_{4\le k,\ \text{even}} \lambda_k \exp(\vec{i}kty_0)\right]\\ =\ &\exp\left[\Phi((1-2t)\exp(\vec{i}ty_0)) - \Phi(1-2t)\right]\\ =\ &\exp\left[\frac{1}{2}(1-2t)^2\left(1-\exp(2\vec{i}ty_0)\right)\right]\left(\frac{1-(1-2t)^2}{1-(1-2t)^2\exp(2\vec{i}ty_0)}\right)^{1/2}\end{aligned} \tag{13.6.5}$$

Furthermore, since $t = o(1)$ as $n \to \infty$, we have that

$$\begin{aligned}\exp\left[\frac{1}{2}(1-2t)^2\left(1-\exp(2\vec{i}ty_0)\right)\right] &= \exp\left[\frac{1}{2}(1-2t)^2(2\vec{i}ty_0+O(t^2))\right]\\ &= \exp(o(1)) \sim 1,\end{aligned} \tag{13.6.6}$$

and that

$$\begin{aligned}\frac{1-(1-2t)^2}{1-(1-2t)^2\exp(2\vec{i}ty_0)} &= \frac{4t-O(t^2)}{4t-2\vec{i}ty_0+O(t^2)}\\ &\sim \frac{2}{2-\vec{i}y_0}.\end{aligned} \tag{13.6.7}$$

Plugging (13.6.6) and (13.6.7) into (13.6.5), we get that if $n \ge n_0$ for a sufficiently large $n_0 > 0$, then

$$|\mathbb{E}(\exp(\vec{i}ty_0Z)) - (1-\frac{\vec{i}}{2}y_0)^{-1/2}| \le \alpha. \tag{13.6.8}$$

In order to compare $\mathbb{E}(\exp(\vec{i}ty_0Z))$ and $\mathbb{E}(\exp(\vec{i}ty_0X))$, we shall approximate the random variable X by the number of vertices on isolated cycles of certain

lengths. If $k \geq \omega/t$ for some large but fixed ω, then by Corollary 13.2.3 the expected number of vertices on isolated cycles of lengths $k \geq \omega/t$ is approximately

$$\begin{aligned} \sum_{k \geq \omega/t, \text{ even}} k\lambda_k &\leq \sum_{k \geq \omega/t, \text{ even}} (1-2t)^k \\ &\leq \sum_{k \geq \omega/t, \text{ even}} \exp(-2tk) \\ &\leq \frac{2}{t}\exp(-\omega). \end{aligned} \tag{13.6.9}$$

Hence, cycles of length $\geq \omega/t$ contribute little to tX if ω is large. Furthermore, once more due to Corollary 13.2.3 the expected number of vertices on isolated cycles of lengths $k < \varepsilon/t$ is about

$$\sum_{k < \varepsilon/t, \text{ even}} k\lambda_k \leq \sum_{k < \varepsilon/t, \text{ even}} (1-2t)^k \leq \frac{\varepsilon}{t}. \tag{13.6.10}$$

Thus, also the contribution of cycles of lengths $< \varepsilon/t$ to tX becomes negligible as $\varepsilon > 0$ gets small. More specifically, choosing $\varepsilon = \varepsilon(y_0, \alpha)$ small enough and $\omega = \omega(y_0, \alpha)$ large enough such that

$$32|y_0|(\varepsilon + \exp(-\omega)) \leq \alpha^2, \tag{13.6.11}$$

we will approximate X by $Y'' = \sum_{\varepsilon/t \leq k < \omega/t, \text{ even}} kY_k$, i.e., in terms of the number of vertices on isolated cycles of lengths $\varepsilon/t \leq k < \omega/t$.

While in the proof of Proposition 13.5.3 we used the fact that the number of isolated cycles of constant length is asymptotically Poisson, we now need to deal with cycles of lengths $\varepsilon/t \leq k < \omega/t$; that is, k grows as a functions of n. In effect, the mean λ_k of Y_k tends to 0 as $n \to \infty$, whence the statement that Y_k is asymptotically Poisson is void (though true). Nonetheless, to compare $\mathbb{E}(\exp(\vec{i}yX))$ and $\mathbb{E}(\exp(\vec{i}yZ))$, we would like to approxmate X in terms of asymptotically independent Poisson variables. Therefore, we partition the interval $[\varepsilon t^{-1}, \omega t^{-1})$ into K pieces $I_j = [\xi_j, \eta_j)$ of equal length δt^{-1}; here $K = K(y_0, \alpha)$ is chosen large enough so that $\delta = (\omega - \varepsilon)/K$ satisfies

$$16|y_0|\delta(\ln\omega - \ln\varepsilon) \leq \alpha^2. \tag{13.6.12}$$

Now, we let J_j be the set of all even integers in I_j, and we define $\mathcal{X}_j = \sum_{k \in J_j} Y_j$ to be the number of cycles of G_M whose length lies in J_j. In addition, set $\Lambda_j = \sum_{k \in J_j} \lambda_k$. At the end of this section we will prove the following proposition, which shows that the random variables $\mathcal{X}_1, \ldots, \mathcal{X}_K$ can indeed be used to approximate Y'' (and thus X) by mutually independent Poisson variables.

Proposition 13.6.1. *$(\mathcal{X}_j)_{1 \leq j \leq K}$ are asymptotically independent Poisson variables with means $(\Lambda_j)_{1 \leq j \leq K}$.*

Set $\mathcal{Z}_j = \sum_{k \in J_j} Z_k$ for $1 \leq j \leq K$. Then $\mathcal{Z}_1, \ldots, \mathcal{Z}_K$ are mutually independent Poisson variables with means $(\Lambda_j)_{1 \leq j \leq K}$. In addition to X, Y, Z, we

consider

$$\begin{aligned} Y' &= \sum_{j=1}^{K} \xi_j \mathcal{X}_j, \\ Y'' &= \sum_{\varepsilon/t \le k < \omega/t,\ \text{even}} kY_k = \sum_{j=1}^{K} \sum_{k \in J_j} kY_k, \\ Z' &= \sum_{j=1}^{K} \xi_j \mathcal{Z}_j, \\ Z'' &= \sum_{\varepsilon/t \le k < \omega/t,\ \text{even}} kZ_k = \sum_{k=2}^{K} \sum_{k \in J_j} kZ_k. \end{aligned}$$

Let us first compare Z' and Z''. Let $W = \sum_{\varepsilon/t \le k \le \omega/t,\ \text{even}} Z_k$. Since $k - \xi_j \le \delta/t$ for all $k \in J_j$,

$$Z'' - Z' \le \frac{\delta}{t} \cdot W. \tag{13.6.13}$$

Furthermore, as

$$\begin{aligned} \sum_{\varepsilon/t \le k \le \omega/t,\ \text{even}} \lambda_k &\le \sum_{\varepsilon/t \le k \le \omega/t,\ \text{even}} k^{-1} \\ &\le \ln(\omega/t) - \ln(\varepsilon/t) \le \ln(\omega) - \ln(\varepsilon), \end{aligned} \tag{13.6.14}$$

we get $\mathbb{E}(W) = \sum_{\varepsilon/t \le k \le \omega/t,\ \text{even}} \lambda_k \le \ln(\omega) - \ln(\varepsilon)$. Therefore, due to Markov's inequality,

$$\Pr\left[W > (\ln(\omega) - \ln(\varepsilon))/\alpha\right] \le \alpha.$$

Consequently,

$$\begin{aligned} \Pr\left[Z'' - Z' > \frac{\alpha}{4t|y_0|}\right] &\overset{(13.6.12)}{\le} \Pr\left[Z'' - Z' > \frac{\delta(\ln(\omega) - \ln(\varepsilon))}{t\alpha}\right] \\ &\overset{(13.6.13)}{\le} \Pr\left[W > \frac{\ln(\omega) - \ln(\varepsilon)}{\alpha}\right] \le \alpha. \end{aligned} \tag{13.6.15}$$

Furthermore, (13.6.9), (13.6.10), and (13.6.11) entail that

$$\mathbb{E}(Z - Z'') = \sum_{k < \varepsilon/t,\ \text{even}} k\lambda_k + \sum_{k \ge \omega/t,\ \text{even}} k\lambda_k \le \frac{\alpha^2}{4t|y_0|},$$

so that Markov's inequality yields

$$\Pr\left[Z - Z'' \ge \frac{\alpha}{4t|y_0|}\right] \le \alpha. \tag{13.6.16}$$

Finally, (13.6.15) and (13.6.16) imply that

$$\Pr\left[t \cdot |y_0| \cdot |Z' - Z| > \alpha\right] \le 2\alpha. \tag{13.6.17}$$

Now, we compare Y' and Y''. Let $U = \sum_{\varepsilon/t \le k < \omega/t,\ \text{even}} Y_k$ be the number of cycles of length between εt^{-1} and ωt^{-1}. Since $k - \xi_j \le \delta/t$ for all $k \in J_j$, we have $Y'' - Y' \le \frac{\delta}{t} \cdot U$. Moreover, by Corollary 13.2.3

$$\mathbb{E}_{\mathcal{M}}[U] \le (1+o(1)) \sum_{\varepsilon/t \le k < \omega/t,\ \text{even}} \lambda_k \overset{(13.6.14)}{\le} 2(\ln(\omega) - \ln(\varepsilon)),$$

so that Markov's inequality yields $\Pr_{\mathcal{M}}[U > 4(\ln(\omega) - \ln(\varepsilon))/\alpha] \le \alpha/2$. Hence,

$$\begin{aligned}
\Pr\left[Y'' - Y' > \frac{\alpha}{4t|y_0|}\right] &\le \Pr_{\mathcal{M}}\left[Y'' - Y' > \frac{\alpha}{4t|y_0|}\right] + o(1) \\
&\overset{(13.6.12)}{\le} \Pr_{\mathcal{M}}\left[Y'' - Y' > \frac{4\delta(\ln(\omega) - \ln(\varepsilon))}{t\alpha}\right] \\
&\le \Pr_{\mathcal{M}}\left[U > \frac{4(\ln(\omega) - \ln(\varepsilon))}{\alpha}\right] + o(1) \\
&\le \alpha. \qquad (13.6.18)
\end{aligned}$$

As a next step, we shall compare Y and $Y'' \le Y$. By Corollary 13.2.3 we have

$$\begin{aligned}
\mathbb{E}_{\mathcal{M}}(Y - Y'') &= \sum_{k < \varepsilon/t,\ \text{even}} k\mathbb{E}_{\mathcal{M}}(Y_k) + \sum_{k \ge \omega/t,\ \text{even}} k\mathbb{E}_{\mathcal{M}}(Y_k) \\
&\le (1+o(1)) \left[\sum_{k < \varepsilon/t,\ \text{even}} (1-2t)^k + \sum_{k \ge \omega/t,\ \text{even}} (1-2t)^k\right] \\
&\overset{(13.6.9),(13.6.10),(13.6.11)}{\le} \frac{\alpha^2}{8t|y_0|}.
\end{aligned}$$

Hence, by Markov's inequality

$$\Pr\left[|Y - Y''| \ge \frac{\alpha}{4t|y_0|}\right] \le \Pr_{\mathcal{M}}\left[|Y - Y''| \ge \frac{\alpha}{4t|y_0|}\right] + o(1) \le \alpha. \qquad (13.6.19)$$

Finally, by Proposition 13.3.1 $\Pr[X \ne Y] = o(1)$ as $n \to \infty$, so that

$$\Pr[X \ne Y] \le \alpha \qquad (13.6.20)$$

if $n \ge n_0$ for a large enough n_0. Thus, (13.6.18), (13.6.19), and (13.6.20) imply that

$$\Pr[t \cdot |y_0| \cdot |X - Y'| > \alpha] \le 3\alpha. \qquad (13.6.21)$$

Proposition 13.6.1 entails that ty_0Z' converges to ty_0Y' in distribution. Therefore, Theorem 3.4.2 yields that the characteristic function of Y', i.e., $y \mapsto \mathbb{E}(\exp(\vec{i}tyY'))$, converges pointwise to the characteristic function of Z', i.e., $y \mapsto \mathbb{E}(\exp(\vec{i}tyZ))$. That is,

$$|\mathbb{E}(\exp(\vec{i}ty_0Z')) - \mathbb{E}(\exp(\vec{i}ty_0Y'))| \le \alpha. \qquad (13.6.22)$$

Finally, (13.6.17), (13.6.21), and (13.6.22) in combination with Lemma 3.4.3 imply that

$$\begin{aligned}|\mathbb{E}\left[\exp(\vec{i}ty_0X)\right] - \mathbb{E}\left[\exp(\vec{i}ty_0Z)\right]| \quad &\leq \quad |\mathbb{E}\left[\exp(\vec{i}ty_0X) - \exp(\vec{i}ty_0Y')\right]| \\ &+|\mathbb{E}\left[\exp(\vec{i}ty_0Y')\right] - \mathbb{E}\left[\exp(\vec{i}ty_0Z')\right]| \\ &+|\mathbb{E}\left[\exp(\vec{i}ty_0Z) - \exp(\vec{i}ty_0Z')\right]| \leq 7\alpha.\end{aligned}$$

Hence, invoking (13.6.8), we conclude that $|\mathbb{E}\left[\exp(\vec{i}ty_0X)\right] - (1-\vec{i}y_0/2)^{-1/2}| \leq 8\alpha$ if $n \geq n_0$ for a large enough n_0, thereby completing the proof of (13.6.4). □

Proof of Proposition 13.6.1. By Proposition 13.1.7 it suffices to show that the random variables $\mathcal{X}_1, \ldots, \mathcal{X}_K$ are asymptotically independent Poisson variables *with respect to the measure* $\Pr_{\mathcal{M}}$. Moreover, if we fix a perfect matching G_* of $V = \{1, \ldots, n\}$, then the $\Pr_{\mathcal{M}}$-distribution of $\mathcal{X}_1, \ldots, \mathcal{X}_K$ coincides with the conditional $\Pr_{\mathcal{M}}$-distribution of $\mathcal{X}_1, \ldots, \mathcal{X}_K$ given that $G_{n/2} = G_*$. Therefore, letting $\Lambda_i = \mathbb{E}_{\mathcal{M}}(\mathcal{X}_i) = \mathbb{E}_{\mathcal{M}}(\mathcal{X}_i|G_{n/2} = G_*)$, we shall prove that for any fixed numbers $r_1, \ldots, r_K$

$$\mu = \mathbb{E}_{\mathcal{M}}\left[\prod_{i=1}^{K}(\mathcal{X}_i)_{r_i}\right] = \mathbb{E}_{\mathcal{M}}\left[\prod_{i=1}^{K}(\mathcal{X}_i)_{r_i}|G_{n/2} = G_*\right] \sim \Lambda = \prod_{i=1}^{K}\Lambda_i^{r_i}; \quad (13.6.23)$$

then the assertion follows from Theorem 3.4.1. If $C_1, \ldots, C_k$ are cycles on the vertex set V, then we let

$$\begin{aligned}&p(\{C_1, \ldots, C_k\}) \\ &= \Pr_{\mathcal{M}}\left[C_1, \ldots, C_k \text{ occur as isolated cycles in } G_M|G_{n/2} = G_*\right].\end{aligned}$$

To show (13.6.23), we employ the following lemma.

Lemma 13.6.2. *Let $C_1, \ldots, C_l$ be cycles of lengths $\varepsilon/t \leq \ell_1, \ldots, \ell_l \leq \omega/t$ such that $p(\{C_1, \ldots, C_l\}) > 0$. Then*

$$\frac{p(\{C_1, \ldots, C_l\})}{\prod_{j=1}^{l}p(C_j)} \sim \frac{\prod_{j=1}^{l}\prod_{k=0}^{\ell_j/2-1}(n-2k-1)}{\prod_{k=0}^{\ell/2-1}(n-2k-1)}.$$

Next we use the linearity of the expectation to expand μ into a sum over $\sum_{i=1}^{K} r_i$-tuples of cycles. More precisely, we will order the terms of this sum according to the lengths of the cycles. Thus, let $\mathcal{L}$ signify the set of all tuples $L = (L_1, \ldots, L_K)$, where each L_i is a tuple $L_i = (L_i^{(j)})_{1\leq j\leq r_i} \in J_i^{r_i}$. Then each $L \in \mathcal{L}$ corresponds to one possibility to specify the cycle lengths in an $\sum_{i=1}^{K} r_i$-tuple of cycles.

In addition, let $\mathcal{Q}$ denote the set of all cycles on the vertex set V that can occur in G_n given that $G_{n/2} = G_*$. Then for each $L \in \mathcal{L}$ we let $\mathcal{D}(L)$ be the set of all tuples $D = (\mathcal{D}_1, \ldots, \mathcal{D}_K)$, where each $\mathcal{D}_i$ is a tuple $(\mathcal{D}_i^{(j)})_{1\leq j\leq r_i}$ of cycles in $\mathcal{Q}$ such that the length of $\mathcal{D}_i^{(j)}$ equals $L_i^{(j)}$; here we do *not* require that the cycles $\mathcal{D}_i^{(j)}$ are distinct. Let $p(D) = p(\{D_i^{(j)} : 1 \leq i \leq K, 1 \leq j \leq r_i\})$ be the

probability that all cycles $D_i^{(j)}$ occur as isolated cycles in G_M. Finally, let $\mathcal{C}(L)$ be the set of all $C = (\mathcal{C}_1, \dots, \mathcal{C}_K) \in \mathcal{D}(L)$ with $\mathcal{C}_i = (\mathcal{C}_i^{(j)})_{j=1,\dots,r_i}$ such that the cycles $(\mathcal{C}_i^{(j)})_{1\le i\le K, 1\le j\le r_i}$ are pairwise vertex disjoint.

Now, μ equals the expected number of tuples $C \in \bigcup_{L\in\mathcal{L}} \mathcal{C}(L)$ such that the cycles in C occur as isolated cycles in G_M. Therefore, the linearity of the expectation yields

$$\mu = \sum_{L\in\mathcal{L}} \sum_{C\in\mathcal{C}(L)} p(C). \tag{13.6.24}$$

Moreover, expanding Λ using the linearity of the expectation, we obtain that

$$\Lambda = \sum_{L\in\mathcal{L}} \sum_{D\in\mathcal{D}(L)} \prod_{i=1}^{K} \prod_{j=1}^{r_i} p(D_i^{(j)}). \tag{13.6.25}$$

To compare (13.6.24) and (13.6.25), we shall compare each of the contributions

$$\mu_L = \sum_{C\in\mathcal{C}(L)} p(C), \quad \Lambda_L = \sum_{D\in\mathcal{D}(L)} \pi_L \quad \text{for } L \in \mathcal{L},$$

where $\pi_L = \prod_{i=1}^{K} \prod_{j=1}^{r_i} p(D_i^{(j)})$. Since $p(D_i^{(j)})$ depends only on the length $L_i^{(j)}$ of the cycle, π_L depends only on L but not on the choice of $D \in \mathcal{D}(L)$. Similarly, $\pi'_L = p(C)$ is the same for all $C \in \mathcal{C}(L)$. Hence,

$$\mu_L = \pi'_L \cdot |\mathcal{C}(L)|, \text{ and } \Lambda_L = \pi_L \cdot |\mathcal{D}(L)|, \tag{13.6.26}$$

Let $\ell = \sum_{i=1}^{K} \sum_{j=1}^{r_i} L_i^{(j)} = o(n)$.

To compare $|\mathcal{C}(L)|$ and $|\mathcal{D}(L)|$, we let $L = (L_i^{(j)})_{1\le i\le K, 1\le j\le r_i} \in \mathcal{L}$, and set $\ell = \sum_{i=1}^{K} \sum_{j=1}^{r_i} L_i^{(j)}$. Then we can construct an element $C = (C_i^{(j)})_{1\le i\le K, 1\le j\le r_i}$ in $\mathcal{C}(L)$ as follows. We choose a tuple $(v_1, \dots, v_\ell)$ of vertices in V as follows. If k is odd, then v_k is chosen arbitrarily from $V \setminus \{v_1, \dots, v_{k-1}\}$; thus, there are $n - k + 1$ ways to choose v_k. Furthermore, if k is even, then v_k is the neighbour of v_{k-1} in G_*, so that v_k is uniquely determined by v_{k-1}. Let $\mathcal{T}$ denote the set of all tuples $(v_1, \dots, v_\ell)$ that can be obtained by this construction. Then $|\mathcal{T}| = \prod_{j=0}^{\ell/2-1} (n - 2j)$.

Moreover, given the tuple $(v_1, \dots, v_\ell)$, we can construct the $\sum_{i=1}^{K} r_i$ cycles of a tuple $C \in \mathcal{C}(L)$ as follows. We turn the first $\sum_{j=1}^{r_1} L_1^{(j)}$ vertices in $(v_1, \dots, v_\ell)$ into cycles of lengths $L_1^{(j)}$ in the natural way: the first cycle starts at v_1, its last vertex is $v_{L_1^{(1)}}$, and its last edge is $\{v_1, v_{L_1^{(1)}}\}$; then, the second cycle contains the vertices $v_{L_1^{(1)}+1}, \dots, v_{L_1^{(1)}+L_1^{(2)}}$ etc. We construct the remaining cycles in C similarly, so that we obtain a map $(v_1, \dots, v_\ell) \mapsto C$ from $\mathcal{T}$ onto $\mathcal{C}(L)$.

However, this map is not one to one. Indeed, for each cycle $C_i^{(j)}$ of length $L_i^{(j)}$ there are precisely $L_i^{(j)}$ ways to list the vertices $w_1, \dots, w_{L_i^{(j)}}$ of $C_i^{(j)}$ such that $\{w_{2s-1}, w_{2s}\}$ in G_* for $1 \le s \le L_i^{(j)}$. Therefore, each $C \in \mathcal{C}(L)$ has precisely $\Gamma = \prod_{i=1}^{K} \prod_{j=1}^{r_i} L_i^{(j)}$ inverse images in $\mathcal{T}$. Consequently,

$$|\mathcal{C}(L)| = \frac{|\mathcal{T}|}{\Gamma} = \Gamma^{-1} \prod_{k=0}^{\ell/2-1} (n - 2k). \tag{13.6.27}$$

A similar counting argument shows that

$$|\mathcal{D}(L)| = \Gamma^{-1} \prod_{i=1}^{K} \prod_{j=1}^{r_i} \prod_{k=0}^{L_i^{(j)}/2-1} (n-2k). \tag{13.6.28}$$

Combining (13.6.27) and (13.6.28), we obtain

$$\frac{|\mathcal{C}(L)|}{|\mathcal{D}(L)|} \sim \frac{\prod_{k=0}^{\ell/2-1}(n-2k)}{\prod_{i=1}^{K}\prod_{j=1}^{r_i}\prod_{k=0}^{L_i^{(j)}/2-1}(n-2k)}. \tag{13.6.29}$$

Finally, combining Lemma 13.6.2 with (13.6.26) and (13.6.29), we conclude that

$$\begin{aligned}
\frac{\mu_L}{\Lambda_L} &= \frac{\pi'_L \cdot |\mathcal{C}(L)|}{\pi_L \cdot |\mathcal{D}(L)|} = \prod_{k=0}^{\ell/2-1} \frac{n-2k}{n-2k-1} \times \prod_{i=1}^{k}\prod_{j=1}^{r_i}\prod_{k=0}^{L_i^{(j)}/2-1} \frac{n-2k-1}{n-2k} \\
&= \prod_{k=0}^{\ell/2-1} \left(1 + \frac{1}{n-2k-1}\right) \times \prod_{i=1}^{k}\prod_{j=1}^{r_i}\prod_{k=0}^{L_i^{(j)}/2-1} \left(1 - \frac{1}{n-2k}\right) \\
&= \exp\left[O(\ell/n)\right] \sim 1,
\end{aligned}$$

whence $\mu_L \sim \Lambda_L$ for all $L \in \mathcal{L}$. Therefore, (13.6.24) and (13.6.25) yield

$$\mu \sim \sum_{L\in\mathcal{L}} \mu_L \sim \sum_{L\in\mathcal{L}} \Lambda_L = \Lambda,$$

so that we have established (13.6.23).

□

Proof of Lemma 13.6.2. Assume that $p(C_1, \ldots, C_l) > 0$. Let

$$\begin{aligned}
q(C_1, \ldots, C_l) &= \Pr_{\mathcal{M}}\left[C_1, \ldots, C_l \text{ occur in } G_n | G_{n/2} = G_*\right], \\
q(C_j) &= \Pr_{\mathcal{M}}\left[C_j \text{ occurs in } G_n | G_{n/2} = G_*\right].
\end{aligned}$$

Moreover, let $\ell_j = O(1/t) = o(\sqrt{n/t})$ denote the length of C_j. Then by Proposition 13.2.2 $p(C_j) \sim (1-2t)^{\ell_j} \cdot q(C_j)$. Therefore,

$$\prod_{j=1}^{l} p(C_j) \sim (1-2t)^{\ell} \cdot \prod_{j=1}^{l} q(C_j). \tag{13.6.30}$$

Similarly, as $\ell = \sum_{j=1}^{l} \ell_j \leq l\omega t^{-1} = o(\sqrt{n/t})$, Proposition 13.2.2 implies

$$p(C_1, \ldots, C_l) \sim (1-2t)^{\ell} \cdot q(C_1, \ldots, C_l). \tag{13.6.31}$$

It is enough to show that

$$\frac{q(C_1, \ldots, C_l)}{\prod_{j=1}^{l} q(C_j)} \sim \frac{\prod_{j=1}^{l}\prod_{k=0}^{\ell_j/2-1}(n-2k-1)}{\prod_{k=0}^{\ell/2-1}(n-2k-1)}. \tag{13.6.32}$$

For combining it with (13.6.30) and (13.6.31), we can conclude that

$$p(C_1, \ldots, C_l) \sim \prod_{j=1}^{l} p(C_j),$$

as desired.

To prove (13.6.32) we let $\rho \in \mathcal{K}(G_*)$ denote a uniformly distributed G_*-configuration, which is defined in Section 13.1. Given that $G_{n/2} = G_*$, G_n is distributed as $G_* + \rho$ (with respect to the measure $\Pr_{\mathcal{M}}$). Moreover, the cycles $C_1, \ldots, C_l$ are present in G_n if and only if there occur corresponding $E(G_*)$-cycles $\rho_1, \ldots, \rho_l$ of lengths $\ell_1, \ldots, \ell_l$ in ρ (cf. Section 13.1 for the definition). Therefore (13.6.32) is equivalent to

$$\frac{\Pr\left[\rho_1, \ldots, \rho_l \text{ occur in } \rho\right]}{\prod_{j=1}^{l} \Pr\left[\rho_j \text{ occurs in } \rho\right]} \sim \frac{\prod_{j=1}^{l} \prod_{k=0}^{\ell_j/2-1} (n-2k-1)}{\prod_{k=0}^{\ell/2-1} (n-2k-1)}. \tag{13.6.33}$$

Let $S \subset V$ be such that for each $s \in S$ there is a vertex $t \in S$ such that $\{s,t\} \in E(G_*)$. Then a (G_*, S)-*configuration* is a map $\sigma : S \to S$ that satisfies $\sigma \circ \sigma = \mathrm{id}$ and $\{s, \sigma(s)\} \notin E(G_*)$ for all $s \in S$. To prove (13.6.33), we observe that by Corollary 13.1.2 the number of (G_*, S)-configurations is

$$(\exp(-1/2) + o(1))(|S|-1)!! \text{ as } |S| \to \infty. \tag{13.6.34}$$

Due to (13.6.34), we can estimate $\Pr\left[\rho_j \text{ occurs in } \rho\right]$ as follows. Let ℓ_j be the length of the cycle C_j. Then the set S_j of vertices outside of C_j has cardinality $n - \ell_j = \Omega(n)$, so that there are $(\exp(-1/2) + o(1))(|S_j|-1)!!$ (S_j, G_*)-configurations. Hence, the number of configurations ρ in that ρ_j occurs is $(\exp(-1/2) + o(1))(|S_j|-1)!!$, while the total number of configurations is $(\exp(-1/2) + o(1))(n-1)!!$. Thus,

$$\begin{aligned} \Pr\left[\rho_j \text{ occurs in } \rho\right] &= \frac{(\exp(-1/2) + o(1))(|S_j|-1)!!}{(\exp(-1/2) + o(1))(n-1)!!} \\ &\sim \frac{(|S_j|-1)!!}{(n-1)!!} = \frac{(n-\ell_j-1)!!}{(n-1)!!}. \end{aligned} \tag{13.6.35}$$

Similarly, letting $S = \bigcap_{j=1}^{l} S_j$ and $\ell = \sum_{j=1}^{l} \ell_j$, we have

$$\Pr\left[\rho_1, \ldots, \rho_l \text{ occur in } \rho\right] = \frac{(\exp(-1/2) + o(1))(|S|-1)!!}{(\exp(-1/2) + o(1))(n-1)!!} \sim \frac{(n-\ell-1)!!}{(n-1)!!}. \tag{13.6.36}$$

Finally, (13.6.33) follows immediately from (13.6.35) and (13.6.36).

□

Bibliography

[1] M. Aigner and G. Ziegler. *Proofs form THE BOOK*. 3rd edition, Springer-Verlag, 2004.

[2] D. Aldous. The random walk construction of uniform spanning trees and uniform labelled trees. *SIAM J. Discrete Math.*, 3:450–465, 1990.

[3] D. Aldous and P. Diaconis. Shuffling cards and stopping times. *Amer. Math. Monthly*, 93:333–348, 1986.

[4] L. Alonso, J. L. Remy, and R. Schott. A linear-time algorithm for the generation of trees. *Algorithmica*, 17:162–182, 1997.

[5] K. Appel and W. Haken. Every planar map is four colorable. Part I. Discharging. *Illinois J. Math.*, 21:429–490, 1977.

[6] K. Appel and W. Haken. Every planar map is four colorable. Part II. Reducibility. *Illinois J. Math.*, 21:491–567, 1977.

[7] M. J. Atallah and J. B. Manning. Fast detection and display of symmetry in outerplanar graphs. *Discrete Appl. Math.*, 39:13–35, 1992.

[8] K. B. Athreya and P. E. Ney. *Branching processes*. Springer-Verlag, New York, 1972. Die Grundlehren der mathematischen Wissenschaften, Band 196.

[9] C. Banderier, P. Flajolet, G. Schaeffer, and M. Soria. Random maps, coalescing saddles, singularity analysis, and Airy phenomena. *Random Structures and Algorithms*, 19:194–246, 2001.

[10] H. Bass. The Ihara-Selberg zeta function of a tree lattice. *Internat. J. Math.*, 3:717 – 797, 1992.

[11] M. Behrisch, A. Coja-Oghlan, and M. Kang. The order of the giant component of random hypergraphs. Submitted. Preliminary version at `http://arxiv.org/abs/0706.0496`, 2007+.

[12] M. Behrisch, A. Coja-Oghlan, and M. Kang. Local limit theorems and the number of connected hypergraphs. Submitted. Preliminary version at `http://arxiv.org/abs/0706.0497`, 2007+.

[13] A. Bender, Z. Gao, and N. C. Wormald. The number of labeled 2-connected planar graphs. *Electronic Journal of Combinatorics*, 9(43), 2002.

[14] E. A. Bender and E. R. Canfield. The number of degree-restricted rooted maps on the sphere. *SIAM J. Discrete Math.*, 7:9–15, 1994.

[15] E. A. Bender and E. R. Canfield. The asymptotic number of labeled graphs with given degree sequences. *J. Combin. Theory Ser. A*, 24:296–307, 1978.

[16] E. A. Bender and L. B. Richmond. The asymptotic enumeration of rooted convex polyhedra. *Journal of Combinatorial Theory*, 36:276–283, 1984.

[17] E. A. Bender, E. R. Canfield, and B. D. McKay. The asymptotic number of labeled connected graphs with a given number of vertices and edges. *Random Structures and Algorithms*, 1:127–169, 1990.

[18] F. Bergeron, G. Labelle, and P. Leroux. *Combinatorial Species and Tree-like Structures.* Cambridge University Press, Cambridge, 1998.

[19] D. Bessis, C. Itzykson, and J. B. Zuber. Quantum field theory techniques in graphical enumeration. *Adv. in Appl. Math.*, 1:109–157, 1980.

[20] T. Biedl. Drawing outer-planar graphs in O($nlogn$) area. In *Proceedings of the tenth Annual Conference on Graph Drawing (GD'02)*, LNCS 2528, pages 54–65, 2002.

[21] M. Bodirsky and M. Kang. Generating random outerplanar graphs. *Combinatorics, Probability, and Computation*, 15(3):333–343, 2006.

[22] M. Bodirsky, C. Gröpl, and M. Kang. Generating labeled planar graphs uniformly at random. In *Proceedings of the Thirtieth International Colloquium on Automata, Languages and Programming (ICALP'03)*, pages 1095–1107, 2003.

[23] M. Bodirsky, C. Gröpl, and M. Kang. Generating unlabeled cubic planar graphs uniformly at random. In *Proceedings of European Conference on Combinatorics, Graph Theory, and Applications (EuroComb'03), ITI Series 2003-145*, pages 61–66, 2003. Journal version submitted.

[24] M. Bodirsky, O. Giménez, M. Kang, and M. Noy. On the number of series-parallel and outerplanar graphs. In *Proceedings of European Conference on Combinatorics, Graph Theory, and Applications (EuroComb'05), DMTCS Proceedings Series*, pages 383–388, 2005.

[25] M. Bodirsky, C. Gröpl, and M. Kang. Sampling unlabeled biconnected planar graphs. In *Proceedings of the 16th Annual International Symposium on Algorithms and Computation (ISAAC'05)*, volume LNCS 3827, pages 593–603, 2005.

[26] M. Bodirsky, É. Fusy, M. Kang, and S. Vigerske. Enumeration and asymptotic properties of unlabeled outerplanar graphs. *Submitted for publication*, 2006.

[27] M. Bodirsky, É. Fusy, M. Kang, and S. Vigerske. An unbiased pointing operator with applications to counting and sampling. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA'07)*, pages 356–365, 2007.

[28] M. Bodirsky, C. Gröpl, D. Johannsen, and M. Kang. A direct decomposition of 3-connected planar graphs. *Séminaire Lotharingien de Combinatoire (SLC)*, pages B54 Ak, 15 pp, 2007. Preliminary version in *the 17th International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC'05)*.

[29] M. Bodirsky, C. Gröpl, and M. Kang. Generating labeled planar graphs uniformly at random. *Theoretical Computer Science*, 379:377–386, 2007.

[30] M. Bodirsky, M. Kang, M. Löffler, and C. McDiarmid. Random cubic planar graphs. *Random Structures and Algorithms*, 30:78–94, 2007.

[31] M. Bodirsky, O. Giménez, M. Kang, and M. Noy. Enumeration and limit laws of series-parallel graphs. *To appear in European Journal of Combinatorics*, 2007+.

[32] H. L. Bodlaender and F. V. Fomin. Approximation of pathwidth of outerplanar graphs. In *Proceedings of the 27th International Workshop on Graph-Theoretic Concepts in Computer Science*, LNCS 2204, pages 166–176, 2001.

[33] B. Bollobás. *Random Graphs*. 2nd edition, Cambridge University Pres, 2001.

[34] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graph. *Europ. J. Combinatorics*, 1:311–316, 1980.

[35] B. Bollobás. The evolution of random graphs. *Trans. Am. Math. Soc.*, 286:257–274, 1984.

[36] B. Bollobás and A. Thomason. Threshold functions. *Combinatorica*, 7: 35–38, 1987.

[37] B. Bollobás and A. Thomason. Random graphs of small order. *Ann. Discrete Math.*, 28:47–97, 1985.

[38] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures and Algorithm*, 31:3–122, 2007.

[39] N. Bonichon, C. Gavoille, and N. Hanusse. An information-theoretic upper bound of planar graphs using triangulation. In *Proceedings of the 20th Annual Symposium on Theoretical Aspects of Computer Science (STACS'03)*, pages 499–510, 2003.

[40] J. Bouttier, P. Di Francesco, and E. Guitter. Census of planar maps: from the one-matrix model solution to a combinatorial proof. *Nuclear Phys. B*, 645:477–499, 2002.

[41] J. Bouttier, P. Di Francesco, and E. Guitter. Combinatorics of hard particles on planar graphs. *Nuclear Phys. B*, 655:313–341, 2003.

[42] C. Brézin, C. Itzykson, G. Parisi, and J.-B. Zuber. Planar diagrams. *Commun. Math. Phys.*, 59:35–51, 197.

[43] R. Bubley and M. Dyer. Path coupling: a technique for proving rapidly mixing in markov chains. In *Proceedings of the 38th IEEE Symposium on Foundations of Computer Science*, pages 223–231. IEEE Computer Science Press, 1997.

[44] P. Bürgisser, M. Clausen, and M. A. Shokrollahi. *Algebraic Complexity Theory*. Number 315 in Grundlehren der mathematischen Wissenschaften. Springer Verlag, 1997.

[45] F. Chung and L. Lu. Connected components in random graphs with given expected degree sequences. *Annals of Combinatorics*, 6:125–145, 2002.

[46] A. Coja-Oghlan and M. Kang. The evolution of the min-min random graph process. Submitted for publication, 2005.

[47] A. Denise and P. Zimmermann. Uniform random generation of decomposable structures using floating-point arithmetic. *Theoretical Computer Science*, 218:233–248, 1999.

[48] A. Denise, M. Vasconcellos, and D. Welsh. The random planar graph. *Congressus Numerantium*, 113:61–79, 1996.

[49] L. Devroye, P. Flajolet, F. Hurtado, M. Noy, and W. Steiger. Properties of random triagulations and trees. *Discrete Comput. Geom.*, 22:105–117, 1999.

[50] P. Di Francesco. Matrix model combinatorics: applications to folding and coloring. In *Random matrix models and their applications*, Math. Sci. Res. Inst. Publ., 40, Cambridge Univ. Press, Cambridge, pages 111–170, 2001.

[51] P. Di Francesco. 2D quantum gravity, matrix models and graph combinatorics. Available at `http://arxiv.org/abs/math-ph/0406013`, 2004.

[52] P. Diaconis. *Group Representations in Probability and Statistics*. Inst. Math. Stat., Hayward, 1988.

[53] P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of markov chains. *Ann. Appl. Probab.*, 1:36–61, 1991.

[54] R. Diestel. *Graph Theory*. Springer–Verlag, New York, 1997.

[55] P. Duchon, P. Flajolet, G. Louchard, and G. Schaeffer. Boltzmann samplers for the random generation of combinatorial structures. *Combinatorics, Probability and Computing*, 13(4-5):577–625, 2004.

[56] M. Dyer and C. Greenhill. Random walks on combinatorial objects. *Surveys in Combinatorics*, 1999 (eds. J. D. Lamb and D A. Preece), London Mathematical Society Lecture Note Series, Cambridge University Press, 1999.

[57] M. Dyer, A. Frieze, and R. Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *J. Assoc. Comput. Mach.*, 38:1–17, 1991.

[58] P. Epstein and J.-R. Sack. Generating triagulations at random. *ACM Trans. Model. and Comput. Simul.*, 4:267–278, 1994.

[59] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6: 290–297, 1959.

[60] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl*, 5:17–61, 1960.

[61] P. Erdős and A. Rényi. On the existence of a factor of degree one of a connected random graphs. *Acta Math. Acad. Sci. Hungar.*, 17:359–368, 1966.

[62] W. Feller. *An introduction to probability theory and its applications*. Wiley, 1967.

[63] P. Flajolet and R. Sedgewick. *Analytic combinatorics*. Book in preparation, the version of the 14th of February 2007. Available at `http://algo.inria.fr/flajolet/Publications`, 2007+.

[64] P. Flajolet, P. Zimmerman, and B. Van Cutsem. A calculus for the random generation of labelled combinatorial structures. *Theoretical Computer Science*, 132:1–35, 1994.

[65] D. Foata and D. Zeilberger. A combinatorial proof of Bass's evaluations of the ihara-selberg zeta function for graphs. *Transactions of the AMS*, 351-6:2257 – 2274, 1999.

[66] É. Fusy. Quadratic exact-size and linear approximate-size random sampling of planar graphs. In *International Conference on Analysis of Algorithms*, volume DMTCS Conference Volume AD (2005), pages 125–138, 2005.

[67] É. Fusy, D. Poulalhon, and G. Schaeffer. Dissections and trees: applications to optimal mesh encoding and random sampling. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'05)*, pages 690–699, 2005.

[68] A. Galluccio and M. Loebl. (p, q)-odd digraphs. *Journal of graph theory*, 23:175–184, 1996.

[69] S. Gerke and C. McDiarmid. On the number of edges in random planar graphs. *Combinatorics, Probability and Computing*, 13:165–183, 2004.

[70] S. Gerke, C. McDiarmid, A. Steger, and A. Weißl. Random planar graphs with n nodes and a fixed number of edges. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'05)*, pages 999 – 1007, 2005.

[71] O. Giménez and M. Noy. Asymptotic enumeration and limit laws of planar graphs. Available at `http://arxiv.org/abs/math.CO/0501269`, 2005.

[72] R. Govindran, M. A. Langston, and X. Yan. Approximation the pathwidth of outerplanar graphs. *Inform. Proc. Letters*, 68:17–23, 1998.

[73] C. Greenhill, A. Ruciński, and N. C. Wormald. Connectedness of the degree bounded star process. *Combinatorics, Probability & Computing*, 12:269–283, 2003.

[74] C. Greenhill, A. Ruciński, and N. C. Wormald. Random hypergraph processes with degree restrictions. *Graphs and Combinatorics*, 20:319–332, 2004.

[75] H. M. Hall. *The Theory of Groups*. Macmillan, 1959.

[76] F. Harary and E. Palmer. *Graphical Enumeration*. Academic Press, New York-London, 1973.

[77] J. Harer and D. Zagier. The Euler characteristic of the moduli space of curves. *Inventiones Mathematicae*, 85:457–485, 1986.

[78] J. E. Hopcroft and R. E. Tarjan. $a\,v^2$ algorithm for determining isomorphism of planar graphs. *Information Processing Letters*, 1(1):32–34, 1971.

[79] J. E. Hopcroft and J. K. Wong. Linear time algorithm for isomorphism of planar graphs. In *Proceedings of the sixth annual ACM Symposium on Theory of Computing*, pages 172–184, 1974.

[80] S. Janson. The growth of components in random graphs. *Random Struc. Alg.*, 1:343–356, 2000.

[81] S. Janson, D. Knuth, T. Łuczak, and B. Pittel. The birth of the giant component. *Random Struc. Alg.*, 4:233–358, 1993.

[82] S. Janson, T. Łuczak, and A. Ruciński. *Random Graphs*. Wiley, 2000.

[83] J. Jaworski and T. Łuczak. Cycles in a uniform graph process. *Combinatorics, Probability & Computing*, 1:223–239, 1992.

[84] M. Jerrum. Mathematical foundations of the Markov chain Monte Carlo method. *Probabilistic Methods for Algorithmic Discrete Mathematics*, Algorithms and combinatorics (eds. M. Habib, C. McDiarmid, J. Ramirez-Alfonsin and B. Reed), Springer, 1998.

[85] M. Kang and M. Loebl. The enumeration of planar graphs via Wick's theorem. Preliminary version at `http://arxiv.org/abs/math.CO/0605218`, 2006.

[86] M. Kang and T. G. Seierstad. The phase transition of the minimum degree random multi-graph process. To appear in *Random Structures and Algorithms*, 2007+.

[87] M. Kang and T. G. Seierstad. The critical phase for random graphs with a given degree sequence. To appear in *Combinatorics, Probability & Computing*, 2007+.

[88] M. Kang, Y. Koh, T. Łuczak, and S. Ree. The connectivity threshold for the min-degree random graph process. *Random Structures and Algorithms*, 29:105–120, 2006.

[89] R. Kannan, L. Lovász, and M. Simonovits. Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random Struct. Algorithms*, 11:1–50, 1997.

[90] M. Karoński and B. Pittel. Existence of a perfect maching in a random $(1 + e^{-1})$-out bipartite graph. *J. Combin. Th. Ser B*, 88:1–16, 2003.

[91] M. Kontsevich. Intersection theory on the moduli space of curves and the matrix airy function. *Commun. Math. Phys.*, 147:1–23, 1992.

[92] S. K. Lando and A. Zvonkin. *Graphs on Surfaces and Their Applications*. Springer-Verlag, 2004.

[93] W. Lenhart and G. Liotta. Proximity drawings of outerplanar graphs. In *Proceedings of the fifth Annual Conference on Graph Drawing (GD'96)*, LNCS 1190, pages 286–302, 1997.

[94] J. Leydold and P. F. Stadler. Minimal cyle bases of outerplanar graphs. *The Electronic Journal of Combinatorics*, 5, 1998.

[95] M. Löffler. Counting and uniform generation of labeled planar structures. Diplomarbeit, Humboldt-Universität zu Berlin, 2005.

[96] T. Łuczak. Component behavior near the critical point of the random graph process. *Random Struc. Alg.*, 1:287–310, 1990.

[97] T. Łuczak. Sparse random graphs with a given degree sequence. in *Random Graphs* **2** (eds. A. M. Frieze, T. Łuczak), 165–182, Wiley, New York, 1992.

[98] T. Łuczak, B. Pittel, and J. Wierman. The structure of a random graph near the point of the phase transition. *Trans. Amer. Math. Soc.*, 341: 721–748, 1994.

[99] C. McDiarmid, A. Steger, and D. Welsh. Random planar graphs. *Journal of Combinatorial Theory, Series B*, 93:187–205, 2005.

[100] S. L. Mitchell. Linear algorithms to recognize outerplanar and maximal outerplanar graphs. *Inform. Proc. Letters*, 9:229–232, 1979.

[101] B. Mohar and C. Thomassen. *Graphs on surfaces*. Johns Hopkins Studies in the Mathematical Sciences, 2001.

[102] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures and Algorithm*, 6:161–180, 1995.

[103] M. Molloy and B. Reed. The size of the giant component of a random graph with a given degree sequence. *Combinatorics, Probability and Computing*, 7:295–305, 1998.

[104] M. Mulase. Asymptotic analysis of a hermitian matrix integral. *Intern. J. Math.*, 6:881–892, 1995.

[105] M. Mulase and A. Waldron. Duality of orthogonal and symplectic matrix integrals and quaternionic feynman graphs. *Commun. Math. Phys.*, 240: 553–586, 2003.

[106] R. C. Mullin and P. J. Schellenberg. The enumeration of c-nets via quadrangulations. *Journal of Combinatorial Theory*, 4:259–276, 1968.

[107] M. Newman, S. Strogatz, and D. Watts. Random graphs with arbitrary degree distributions and their applications. Available at `http://arxiv.org/pdf/cond-mat/0007235`, 2001.

[108] A. Nijenhuis and H. S. Wilf. *Combinatorial algorithms.* Academic Press Inc., 1979.

[109] J. R. Norris. *Markov chains.* Cambridge University Press, 1998.

[110] H. Okamura and P. D. Seymour. Multicommodity flows in planar graphs. *J. Combin. Theory Ser. B*, 31:75–81, 1981.

[111] A. Okounkov. Random trees and moduli of curves. In *Asymptotic combinatorics with applications to mathematical physics (eds. A. M. Vershik)*, Springer. Lect. Notes Math. 1815, pages 89–126, 2003.

[112] D. Osthus, H. J. Prömel, and A. Taraz. On random planar graphs, the number of planar graphs and their triangulations. *Journal of Combinatorial Theory, Series B*, 88:119–134, 2003.

[113] R. Otter. The number of trees. *Ann. Math.*, 49:583–599, 1948.

[114] R. Pemantle. Uniform random spanning trees. Topics in contemporary probability and its applications (J.Snell ed.), Boca Raton, CRC Press, Probability and Stochastics Series, 1995.

[115] R. C. Penner. Perturbative series and the moduli space of riemann surfaces. *J. Differential Geom.*, 27:35–53, 1988.

[116] N. Pippenger and K. Schleich. Topological characteristics of random triangulated surfaces. *Random Structures and Algorithms*, 28:247–288, 2006.

[117] J. M. Plotkin and J. W. Rosenthal. How to obtain a singular expansion of a sequence from an analytic identity satisfied by its generating function. *Journal of the Australian Mathematical Society, Series A*, 56:131–143, 1994.

[118] G. Pólya. Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen. *Acta Mathematica*, 68:145–254, 1937.

[119] R. C. Read. On general dissections of a polygon. *Aequationes Mathematicae, University of Waterloo*, 18:370–388, 1978.

[120] J. Riordan. Forests of labeled trees. *J. Combin. Theory*, 5:90–103, 1968.

[121] H. D. Robalewska. Stochastic processes and random graphs. PhD thesis, University of Melbourne, 1996.

[122] H. D. Robalewska and N. C. Wormald. Random star processes. *Combinatorics, Probability & Computing*, 9:33–43, 2000.

[123] N. Robertson, D. P. Sanders, P. D. Seymour, and R. Thomas. The four colour theorem. *J. Combin. Theory Ser. B*, 70:2–44, 1997.

[124] A. Ruciński and N. C. Wormald. Connectedness of graphs generated by a random d-process. *J. Austral. Math. Soc.*, 72:67–85, 2002.

[125] A. Ruciński and N. C. Wormald. Random graph processes with degree restrictions. *Combinatorics, Probability & Computing*, 1:169–180, 1992.

[126] W. Rudin. *Real and complex analysis*. McGraw-Hill. 3rd ed., 1987.

[127] G. Schaeffer. *Conjugaison d'arbres et cartes combinatoires aléatoires*. PhD thesis, Université Bordeaux I, 1998.

[128] G. Schaeffer. Random sampling of large planar maps and convex polyhedra. In *Proceedings of the thirty-first annual ACM symposium on theory of computing (STOC'99)*, pages 760–769, 1999.

[129] E. Schröder. Vier combinatorische Problem. *Z. Math. Physik*, 15:361–376, 1870.

[130] S. Sherman. Combinatorial aspects of the Ising model of ferromagnetism I. *J. Math. Phys.*, 1:202–217, 1960.

[131] A. N. Shiryayev. *Probability*. Springer-Verlag, 1984.

[132] A. Sinclair. Improved bounds for mixing rates of markov chains and multicommodity flow. *Combinatorics, Probability and Computing*, 1:351–370, 1992.

[133] N. J. A. Sloane. The On-Line Encyclopedia of Integer Sequences. `http://www.research.att.com/~njas/sequences/`, Last updated 25 July 2007.

[134] L. Takacs. On the number of distinct forests. *SIAM J. Discrete Math.*, 3: 574–581, 1990.

[135] Y. Tomii. Gewurzelte unbeschriftete Outerplanare Graphen. Studienarbeit, Humboldt-Universität zu Berlin, 2005.

[136] B. A. Trakhtenbrot. Towards a theory of non-repeating contact schemes. *Trudi Mat. Inst. Akad. Nauk SSSR*, 51:226–269, 1958. [In Russian].

[137] W. Tutte. A census of planar triangulations. *Canad. J. Math.*, 14:21–38, 1962.

[138] W. T. Tutte. *Graph Theory*. Cambridge University Press, 1984.

[139] S. Vigerske. Asymptotic enumeration of unlabelled outerplanar graphs. Diplomarbeit, Humboldt-Universität zu Berlin, 2005.

[140] T. R. Walsh. Counting labeled three-connected and homeomorphically irreducible two-connected graphs. *J. Combin. Theory B*, 32:1–11, 1982.

[141] T. R. Walsh. Counting unlabelled three-connected and homeomorphically irreducible two-connected graphs. *J. Combin. Theory*, 32:12–32, 1982.

[142] T. R. Walsh. Counting nonisomorphic three-connected planar maps. *J. Combin. Theory*, 32:33–44, 1982.

[143] H. S. Wilf. The uniform selection of free trees. *J. Algorithms*, 2:204–207, 1981.

[144] H. S. Wilf. *Generatingfunctionology.* Academic Press, Boston, 1994.

[145] N. C. Wormald. The differential equation method for random graph processes and greedy algorithms. In *Lectures on Approximation and Randomized Algorithms (eds. M. Karoński, H-J. Prömel)*, pages 75–152, Warsaw, 1999. PWN.

[146] N. C. Wormald. A poly-time algorithm for 3-connected unlabelled planar graphs. Personal communication, 2002.

[147] N. C. Wormald. Models of random regular graphs. In: Lamb, J.D, Preece, D.A. (eds.): *Surveys in Combinatorics.* Cambridge University Press, 239–298, 1999.

[148] N. C. Wormald. Differential equations for random processes and random graphs. *Ann. Appl. Probab.*, 5(4):1217–1235, 1995.

[149] A. Zvonkin. Matrix integral and map enumeration:an accessible introduction. *Mathl. Comput. Modelling*, 26:281–304, 1997.

Selbständigkeitserklärung

Hiermit erkläre ich, dass

- ich die vorliegende Habilitationsschrift selbstständig ohne fremde Hilfe verfasst und nur die angegebene Literatur und Hilfsmittel verwendet habe,
- für mich weder ein früheres noch ein schwebendes Habilitationsverfahren existiert, und
- mir die Habilitationsordnung der Mathematisch-Naturwissenschaftlichen Fakultät II der Humboldt-Universität zu Berlin vom 17.01.2005 veröffentlicht in Amt-liches Mit-teilungs-blatt der Humboldt-Universität zu Berlin Nr. 23/2005, bekannt ist.

Berlin, den 31. Oktober 2006

Mihyun Kang