Airy Phenomena and Analytic Combinatorics of Connected Graphs

Philippe Flajolet
Algorithms Project, INRIA Rocquencourt, 78153 Le Chesnay (France).
Philippe.Flajolet@inria.fr.

Bruno Salvy
Algorithms Project, INRIA Rocquencourt, 78153 Le Chesnay (France).
Bruno.Salvy@inria.fr.

Gilles Schaeffer
LIX – CNRS, École polytechnique, 91128 Palaiseau (France).
Gilles.Schaeffer@lix.polytechnique.fr

Submitted: Oct 11, 2002 & Mar 12, 2004; Accepted: Apr 7, 2004; Published: May 27, 2004.
MR Subject Classifications: 05A15, 05A16, 05C30, 05C40, 05C80

Abstract

Until now, the enumeration of connected graphs has been dealt with by probabilistic methods, by special combinatorial decompositions or by somewhat indirect formal series manipulations. We show here that it is possible to make analytic sense of the divergent series that expresses the generating function of connected graphs. As a consequence, it becomes possible to derive analytically known enumeration results using only first principles of combinatorial analysis and straight asymptotic analysis—specifically, the saddle-point method. In this perspective, the enumeration of connected graphs by excess (of number of edges over number of vertices) derives from a simple saddle-point analysis. Furthermore, a refined analysis based on coalescent saddle points yields complete asymptotic expansions for the number of graphs of fixed excess, through an explicit connection with Airy functions.

Introduction

E. M. Wright, of Hardy and Wright fame, initiated the enumeration of labelled connected graphs by number of vertices and edges in a well-known series of articles [34, 35, 36]. In particular, he discovered that the generating functions of graphs with a fixed excess of number of edges over number of vertices has a rational expression in terms of the tree function $T(z)$. Wright's approach is based on the fact that deletion of an edge in a connected graph leads to either one or two connected graphs with smaller excess. This decomposition translates into a quadratic differential recurrence from which Wright was able to deduce general structural results, especially as regards dominant asymptotics.

The problem of enumerating connected graphs by excess is obviously related to the question of connectivity in random graphs and so, not unnaturally, it has also been approached repeatedly through the probabilistic method. It is for instance of special importance in the emergence of the "giant component" under Erdős and Rényi's model. Bollobás's book [4, Ch. 6] contains an account of various aspects of the question examined from the probabilistic angle. The "giant paper" on the "giant component" by Janson, Knuth, Łuczak and Pittel [18] devotes some 25 pages to generating function evaluations before going into the actual physics of the random graph phase transition. Finally, the enumerative results valid asymptotically over the widest range of the parameters are those of Bender, Canfield, and McKay in [3].

In contrast, our approach here is completely straightforward. It starts from the bivariate generating function of connected graphs

$$C(z,q) = \log\left(1 + \sum_{n\geq 1}(1+q)^{n(n-1)/2}\frac{z^n}{n!}\right), \tag{1}$$

that is viewed nowadays as a direct instance of the classical "exponential formula" of combinatorial analysis; see for instance [14, 33]. (The formula was published by Riddel and Uhlenbeck [26] in 1953.) We show that this series that strongly diverges for any $q>0$ can in fact be represented by an integral that gives it *bona fide* analytic meaning for small $q<0$. In a way, this amounts to assigning *negative* weights (or probabilities) to edges, contrary to what is done commonly in probabilistic or enumerative treatments of the question like [3, 4]. Then, standard methods of asymptotic analysis, especially the saddle-point technique, apply. Thus, in a logical sense, the enumeration of graphs by excess "only" requires the exponential formula and basic asymptotics. Together with the article of Janson *et al.* [18], the present article is one of the very few approaches that treats connectivity of graphs starting from first principles. As opposed to [18], our approach is purely analytic and hopefully a little more transparent from a logical standpoint. It is also a curious fact that asymptotic analysis is used here to establish an exact enumerative result.

Our principal result is a purely analytic proof of a theorem, known from earlier works of Wright[1] [34] and of Janson *et al.* [18]. A main character throughout the article is the "tree function" that is defined by

$$T(z) = ze^{T(z)}, \qquad T(z) = \sum_{n\geq 1} n^{n-1}\frac{z^n}{n!}, \tag{2}$$

and is otherwise known to enumerate rooted labelled trees. For any connected graph with

[1] Wright's results were to some extent anticipated by Temperley [31] whose insightful short note of 1959 seems to rely partly on heuristic arguments.

k edges and n vertices, the quantity $k-n$ is always at least -1 and is called the *excess*[2] of the graph. Our goal is a characterization of the (exponential) generating functions (GFs) of graphs of any fixed excess.

Theorem 1. *(i) The generating function of unrooted trees (graphs with excess -1) is*

$$W_{-1}(z) = T(z) - \frac{1}{2}T^2(z). \tag{3}$$

(ii) The GF of connected graphs with excess 0 (unicyclic graphs) is

$$W_0(z) = \frac{1}{2}\log\frac{1}{1-T(z)} - \frac{1}{2}T(z) - \frac{1}{4}T^2(z). \tag{4}$$

(iii) The GF of connected graphs with excess $k \geq 1$ is a rational function of $T(z)$: there exist polynomials A_k, such that

$$W_k(z) = \frac{A_k(T(z))}{(1-T(z))^{3k}}. \tag{5}$$

Part (*i*) is commonly attributed to Cayley and several of his contemporaries (see [18, p. 240] for a discussion), while Part (*ii*) is due to Rényi; Equation (5) of Part (*iii*) is Wright's main result. Observe that Wright had to resort to an "external argument" based on special multigraphs [34, Sec. 7] in order to obtain the rationality of the $W_k(z)$ in terms of $T(z)$.

For completeness, we recall that the generating functions provided by Theorem 1 are equivalent to explicit forms for the quantities[3] $C_{n,n+\ell} = n![z^n]W_\ell(z)$, as results from the standard expansions,

$$\frac{1}{1-T(z)} = 1 + \sum_{n\geq 1} n^n \frac{z^n}{n!}, \qquad \log\frac{1}{1-T(z)} = \sum_{n\geq 0} Q_n n^{n-1}\frac{z^n}{n!}, \tag{6}$$

where Q_n is the Ramanujan Q-function (see [8, 19] and references therein):

$$Q_n = 1 + \frac{n-1}{n} + \frac{(n-1)(n-2)}{n^2} + \cdots .$$

From either the explicit forms deriving from (6) or from the known expansion [21] of the tree function at its dominant singularity $\exp(-1)$, namely,

$$T(z) = 1 - \sqrt{2}\sqrt{1-ez} + \frac{2}{3}(1-ez) - \frac{11}{36}\sqrt{2}(1-ez)^{3/2} + \cdots , \qquad z \to \exp(-1),$$

the following holds.

[2]Our notion of excess is consistent with the one of Janson *et al.* [18, p. 240]. Our W_k coincide with those of Wright [34, p. 318] and are equal to the $\widehat{C}_k$ in the notations of [18].

[3]As usual, we denote by $[z^n]f(z)$ the nth coefficient in the series $f(z)$.

Corollary 1. *The asymptotic form of the graph counts $C_{n,n+k} = n![z^n]W_k(z)$ for fixed $k \geq 2$ is*

$$A_k(1)\sqrt{\pi}\left(\frac{n}{e}\right)^n\left(\frac{n}{2}\right)^{\frac{3k-1}{2}}\left(\frac{1}{\Gamma\left(\frac{3k}{2}\right)}+\frac{\frac{A_k'(1)}{A_k(1)}-k}{\Gamma\left(\frac{3k-1}{2}\right)}\sqrt{\frac{2}{n}}+O\left(\frac{1}{n}\right)\right), \tag{7}$$

and more generally, lower order terms depend on the derivatives $A_k^{(j)}(1)$.

The numerical coefficients $A_k(1)$ are crucial to dominant asymptotics. They are expressed in terms of the Airy Ai function, which is another central character in our analysis. This function is classically defined by [32]

$$\mathrm{Ai}\,(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i(zt+t^3/3)}dt. \tag{8}$$

It satisfies the linear differential equation $y'' - zy = 0$, with limiting condition $\mathrm{Ai}\,(z) \to 0$ as $z \to +\infty$. Its asymptotic expansion at infinity is of a hypergeometric form given by [1, Entry 10.4.59]

$$\mathrm{Ai}\,(z) \sim \frac{1}{2}\pi^{-1/2}z^{-1/4}e^{-\zeta/3}\sum_{k=0}^{\infty} c_k(-\zeta)^{-k},$$

where $\zeta = 2z^{3/2}$ and

$$c_k = \frac{\Gamma(3k+\frac{1}{2})}{18^k k!\Gamma(k+\frac{1}{2})} = \frac{(6k)!}{(3k)!(2k)!3^{2k}2^{5k}}. \tag{9}$$

Our analysis allows us to characterize the coefficients $A_k(1), A_k'(1), \ldots$.

Theorem 2. (i) *The generating series of the dominant coefficients $A_k(1)$ is expressible as a formal power series and as an asymptotic series,*

$$\begin{aligned}\sum_{k=1}^{\infty} A_k(1)(-x)^k &= \log\left(1+\sum_{k=1}^{\infty} c_k(-x)^k\right)\\ &\sim \log\left(2\sqrt{\pi}(2x)^{-1/6}e^{1/(3x)}\,\mathrm{Ai}\left((2x)^{-2/3}\right)\right), \qquad x \to 0,\end{aligned} \tag{10}$$

where the c_k's are defined by (9).

(ii) *For $j \geq 1$, the generating series of the derivatives $A_k^{(j)}(1)$ can be expressed in terms of the Airy function. More precisely, let $S(x)$ be the asymptotic expansion*

$$S(x) = 1 + \frac{95}{288}x + \cdots \sim -\frac{2}{x}\left(1+(2x)^{1/3}\frac{\mathrm{Ai}'((2x)^{-2/3})}{\mathrm{Ai}((2x)^{-2/3})}\right), \quad x \to 0,$$

then, as formal power series, one as

$$\sum_{k\geq 1} A_k^{(j)}(1)x^k = \mathcal{A}^{(j)}(x, S(x)),$$

where $\mathcal{A}^{(j)}(x,s)$ is a polynomial of degree j in s with coefficients that are polynomials in x and x^{-1}. These polynomials can be determined effectively from Equation (47) below. (See Appendix II for a table.)

The dominant coefficients $A_k(1)$ are thus in essence the coefficients of the expansion of log(Ai) at infinity. The coefficients $A_k^{(j)}(1)$ intervene as subdominant terms in the expansion of $W_k(z)$ and their knowledge provides a full asymptotic expansion extending (7).

Part (i), the form (10) of the driving coefficients $A_k(1)$, is given explicitly by Janson *et al.* in [18] where the authors built upon earlier results of the "Russian school", most notably Bagaev and Voblyi. In view of Corollary 1 and the accompanying remarks, this form characterizes the dominant asymptotics of the number of graphs of some fixed excess k. Part (ii) of Theorem 2 then provides a "correction series" that describes precisely what goes on in successive subdominant asymptotic terms.

Wright had in fact obtained earlier a recursive determination of the $A_k(1)$ but he does not appear to have obtained the relation (10). We now know that the $W_k(z)$ and the $A_k(1)$ intervene in a number of closely interrelated problems and a variant of the sequence $\{A_k(1)\}$, called the "Wright-Louchard-Takács sequence", appears in [12]. Indeed, the $W_k(z)$ and the $A_k(1)$ surface in such diverse problems as: parking and linear probing hashing [12, 20], Brownian excursion area and area below Dyck path [10, 22], area below the Poisson excursion [27], inversions in trees [13, 23], path length in trees of various sorts [28, 29, 30], and naturally the enumeration of connected graphs [9, 18, 34, 35, 36]. See [12, 20] for a combinatorial perspective on the relationship between these problems.

The present article fits in a more global endeavour aimed at finding simple reasons for the occurrence of the Airy function in so many problems of analytic combinatorics. One good reason is the connection with coalescing saddles as exemplified by [2] in the case of random maps that are random planar graphs of a specific type. We propose to examine in future works the extent to which it can be applied to other graph models, to uniform estimates, and to phase transitions that arise in hashing and random allocation problems [12, 20].

Finally, statistical physics is lurking in the background. In a partly heuristic, but insightful paper [31], Temperley developed formulæ that correspond to a primitive form of Theorem 1. In [24], Monasson proposed to approach connectivity of the random graph via the replica method. It would be of obvious interest to confront the rigourous approach developed here with the powerful (but yet unrigourous) replica method. The present paper may hopefully contribute to the debate.

Plan of the article. The article is entirely based on an integral representation for the divergent series in (1). In other words, the generating function of graphs can be viewed as the asymptotic expansion of a *bona fide* analytic object. This is described in Section 1, where a combinatorial bijection due to Gessel and Wang is used to dispose of some of the divergence issues. A straightforward application of the saddle-point method for the asymptotics of integrals then yields Equations (3), (4) and (5) in Section 2. Consequently, Wright's representation (Part (iii) of Theorem 1) appears to *coincide* with a standard saddle-point expansion. The dominant coefficients $A_k(1)$ as well as their their subdominant companions, $A'_k(1)$, etc., are a bit more recondite. In Section 3, we show that relevant information can be gathered by a method of coalescent saddle points, which gives Theorem 2. Thus, it appears that very detailed formal expansions found by Janson *et al.* [18] are in fact precisely double saddle-point expansions. The expressions obtained involve hypergeometric functions that are reducible to the Airy function.

1 Principles of an Analytic Approach

We discuss here the principles on which our proof of Theorem 1 is built. It is based on assigning a complex-analytic meaning to strongly divergent series that express graphical enumerations and to correlative series rearrangements. Accordingly, special attention is required in distinguishing carefully between formal objects and their analytic counterparts.

1.1 Formal Expressions

In what follows, for K a field and $z_1, z_2, \ldots$ a finite collection of indeterminates, we let $K[[z_1, z_2, \ldots]]$ denote the ring of formal power series with indeterminates $z_1, z_2, \ldots$ and coefficients in K.

Let $G_{n,k}$ be the number of labelled graphs with n vertices and k edges and $C_{n,k}$ the number of those that are connected. The (formal) generating functions are defined as objects of $\mathbb{C}[[z,q]]$ by

$$G(z,q) := \sum_{n,k} G_{n,k} q^k \frac{z^n}{n!}, \qquad C(z,q) := \sum_{n,k} C_{n,k} q^k \frac{z^n}{n!}.$$

A graph is determined by the selection of edges amongst all possible pairs of points, implying the identity in $\mathbb{C}[[z,q]]$,

$$G(z,q) = \sum_{n\geq 0} (1+q)^{n(n-1)/2} \frac{z^n}{n!}. \tag{11}$$

On the other hand, a graph is a set of connected components, which, by the classical exponential formula, implies the relation $G(z,q) = \exp(C(z,q))$ and consequently

$$\begin{aligned} C(z,q) &= \log(G(z,q)) \\ &= z + q\frac{z^2}{2!} + (3q^2+q^3)\frac{z^3}{3!} + (16q^3+15q^4+6q^5+q^6)\frac{z^4}{4!} + \cdots, \end{aligned} \tag{12}$$

valid again in $\mathbb{C}[[z,q]]$.

Consider next the GF of connected graphs counted according to excess and to size. An essential component of our approach is to record excess and do so by a *negatively signed* variable. Then, in $\mathbb{C}[[z,q]]$ we have

$$\begin{aligned} Q(z,q) &:= \sum_{n,\ell} C_{n,n+\ell}(-q)^{\ell+1}\frac{z^n}{n!} = -qC(-z/q,-q) \\ &= z + \frac{z^2}{2!} + (3-q)\frac{z^3}{3!} + (16-15q+6q^2-q^3)\frac{z^4}{4!} + \cdots \\ &= W_{-1}(z) - qW_0(z) + q^2W_1(z) - \cdots, \end{aligned}$$

where each $W_\ell(z) \in \mathbb{C}[[z]]$ is by definition the generating function of connected graphs with excess ℓ:

$$W_\ell(z) := \sum_n C_{n,n+\ell} \frac{z^n}{n!}. \tag{13}$$

For instance $W_{-1}(z)$ is the GF of unrooted trees, $W_0(z)$ the GF of unicyclic components, and so on.

Now, the exponential formula (12) in conjunction with (11) permits us to express $Q(z,q)$ formally as

$$Q(z,q) = -q \log\left(\sum_{n \geq 0} (1-q)^{n(n-1)/2} \frac{(-zq^{-1})^n}{n!} \right). \tag{14}$$

In this formula, the right-hand side is to be taken as an element of $\mathbb{C}(q)[[z]]$ (that is, the ring of formal power series in z, with coefficients that are rational functions in q) and the reorganization of the series takes place in that domain, according to the formula

$$\log(1+u) = \frac{u}{1} - \frac{u^2}{2} + \frac{u^3}{3} - \cdots .$$

Massive algebraic cancellations in the coefficient field $\mathbb{C}(q)$ take place when the series is reorganized and this seems to be the cause of many analytic hardships.

1.2 Integral Representations

The basic analytic representation derives from the following simple lemma that we state in its bare-bones version.

Lemma 1. *Let v_n be a sequence with generating function $V(z) = \sum_n v_n z^n$ satisfying $\sum_n |v_n| < \infty$ and let w be a real number with $w \in (0,1)$. Then,*

$$\sum_n w^{n^2/2} v_n = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} V(e^{ix\sqrt{\log w^{-1}}}) e^{-x^2/2}\, dx. \tag{15}$$

Proof. For finite sequences v_n, the lemma directly results from the classical Fourier integral

$$e^{-t^2/2} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{ixt} e^{-x^2/2}\, dx.$$

For infinite sequences, it results from interchange of summation and integration. □

The importance of (15) for the analysis of q-series comes from the fact that *the integral representation linearizes the quadratic forms present in the exponents.* (See for instance [11] for a combinatorial application to chord systems.)

The graph generating function $G(z,q)$ specified by (11) and viewed as a function of its two parameters z, q diverges wildly as soon as q is positive. However, it acquires a *bona fide* analytic meaning if it is considered as a series in z with q a fixed parameter, provided $|1+q| < 1$. In that case, it becomes an entire function of z. Given this, we may legitimately expect $Q(z,q)$ to make analytic sense when q is restricted to the disk centred at $+1$ with radius 1. Precisely, we fix q as a *real valued* parameter such that $|1-q| < 1$, (i.e., $0 < q < 2$) and consider the weighting π that assigns to a graph g the weight $\pi(g) := (-q)^{e(g)-|g|}$ where $e(g)$ is the number of edges and $|g|$ is the number of vertices of g. We introduce the two analytic objects

$$\mathcal{H}(z,q) := \sum_{g \text{ graph}} \pi(g) \frac{z^{|g|}}{|g|!}, \qquad \mathcal{Q}(z,q) := -q \sum_{g \text{ connected graph}} \pi(g) \frac{z^{|g|}}{|g|!}.$$

The function $\mathcal{H}$ is an entire function of z for q in the given range, since it is directly related to G by $\mathcal{H}(z,q) = G(-z/q, -q)$. The exponential connection between $\mathcal{H}$ and $\mathcal{Q}$, namely $\mathcal{Q} = -q \log \mathcal{H}$, holds. Observe also that $\mathcal{Q}$ is an analytic function of z for $|z|$ sufficiently small, since $\mathcal{H}(0,q) = 1$.

Application of Lemma 1 now yields the following basic integral representation.

Lemma 2. *The generating function of connected graphs counted by excess and weighted with negative weights admits for $q \in (0,1)$ the integral representation*

$$\mathcal{Q}(z,q) = -q \log \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2} - z \frac{(1-q)^{-1/2}}{q} e^{ix\lambda(q)} \right) dx \right), \tag{16}$$

where

$$\lambda \equiv \lambda(q) := \sqrt{\log \frac{1}{1-q}} = \sqrt{q} \left(1 + \frac{1}{4} q + \frac{13}{96} q^2 + \cdots \right).$$

This representation is central to our treatment.

Proof. Start from Equation (14), set $w = 1-q$ and apply Lemma 1. The restriction to $q \in (0,1)$ ensures the well definedness of λ. ☐

1.3 Interchange of Limits and Coefficient Operators

We will prove later, as a by-product of the analysis, that there exists a family $\mathcal{W}_\ell(z)$ of analytic functions each having radius of convergence e^{-1} such that the analytic $\mathcal{Q}(z,q)$ satisfies when $|z| < e^{-1}$ as $q \to 0^+$:

$$\mathcal{Q}(z,q) \underset{q \to 0^+}{\sim} \sum_{\ell \geq 0} \mathcal{W}_{\ell-1}(z)(-q)^\ell. \tag{17}$$

This is to be understood in the usual sense of Poincaré [25], namely for $|z| < e^{-1}$ and every L,

$$\mathcal{Q}(z,q) - \sum_{\ell=0}^{L} \mathcal{W}_{\ell-1}(z)(-q)^\ell = O(q^{L+1}), \qquad q \to 0$$

with the coefficient hidden in the O-notation possibly dependent on z. The $\mathcal{W}_\ell$ found in the process satisfy precisely the conditions of Theorem 1 and the following sections will show how to establish (17) from first principles.

If granted the asymptotic expansion (17), the proof is complete once we establish that $W_\ell(z) = \mathcal{W}_\ell(z)$. Now, the *algebraic* quantities $W_\ell(z)$ are such that, by definition,

$$n![z^n]W_\ell(z) = C_{n,n+\ell} = (-1)^{\ell+1}[q^{\ell+1}]\left(n![z^n]Q(z,q)\right),$$

with $Q(z,q)$ as specified by the formal relation (14), and $[z^n], [q^\ell]$ representing here *formal* coefficient extraction in $\mathbb{C}[[z,q]]$. On the other hand, the *analytic* quantities $\mathcal{W}_\ell(z)$ are defined as

$$\mathcal{W}_\ell(z) = (-1)^{\ell+1}[q^{\ell+1}]\mathcal{Q}(z,q),$$

where the notation $[q^\ell]$ means now extraction of the coefficient of q^ℓ in the *asymptotic* expansion of $\mathcal{Q}(z,q)$ as a function of q with $q \to 0^+$. Thus, under the assumption (17), Theorem 1 only depends on the validity of the *interchange of coefficient operators*:

$$[q^\ell]\left([z^n]Q(z,q)\right) \stackrel{?}{=} [z^n]\left([q^\ell]\mathcal{Q}(z,q)\right).$$

Naturally, the divergent character of the underlying series renders this interchange non-obvious.

The basic ingredient is a lemma that grants conditionally such an interchange of limits and coefficient operators.

Lemma 3 (Interchange of limits and coefficient operators). (i) *Let $s_n(u)$ be polynomials with* nonnegative *coefficients and assume that the series*

$$S(z,u) = \sum_{n\geq 0} s_n(u)z^n$$

converges for $|z| < r$ (for some $r > 0$) and $|u| < 1$. Assume that there exists a function $f(z) = \sum_n f_n z^n$ analytic in $|z| < r$ such that

$$\lim_{u\to 1^-} S(z,u) = f(z) \qquad \textit{pointwise for any } z,\ |z| < r.$$

Then for all $n \geq 0$,

$$s_n(1) \equiv \lim_{u\to 1^-}[z^n]S(z,u) = [z^n]\lim_{u\to 1^-} S(z,u) \equiv f_n.$$

(ii) *Additionally, assume that there exist functions $g_0(z) = f(z)$, $g_1(z), \ldots$ analytic in $|z| < r$ such that for $u \to 1^-$, one has*

$$S(z,u) \underset{u\to 1^-}{\sim} \sum_{k=0}^{\infty}(u-1)^k g_k(z).$$

Then, all the derivatives of the polynomials s_n at 1 are given by

$$\left.\frac{d^k}{du^k}s_n(u)\right|_{u=1} = k![z^n]g_k(z), \qquad k,n \geq 0.$$

Proof. (i) For any fixed u, write $S_u(z) = S(z,u)$ and consider the family of analytic functions $\{S_u(z)\}$ (in the variable z) indexed by u that ranges between 0 and 1 while tending to 1. Inside the disk $|z| < r$, the convergence $S_u(z) \to f(z)$ as $u \to 1$ is dominated by $f(|z|)$: for positive z, this results plainly from the positivity of the $s_n(u)$, and for arbitrary z, from the triangular inequality. In particular, the convergence is dominated by a constant $M(r')$ in an arbitrarily chosen sub-disk $|z| \le r'$ with $0 < r' < r$. By a classical result of the theory of analytic functions, bounded pointwise convergence on compact sets implies uniform convergence of $S_u(z)$ to $f(z)$, together will all derivatives. (See the discussion of *normal families* of functions in [15, Ch. 12] or [17, Ch. 15], especially pp. 246–247.) By extraction of coefficients, assertion *(i)* follows.

(ii) The proof follows by induction on k. Assume that $S(z,u)$ now satisfies the stronger conditions of (ii) and that the conclusion is met up to $k-1$. Set

$$T(z,u) = \frac{1}{(u-1)^k}\left(S(z,u) - \sum_{i=0}^{k-1} g_i(z)(u-1)^i\right).$$

Then, by assumption, $T(z,u)$ admits a shifted asymptotic expansion of the same type as $S(z,u)$, and in particular, it converges to the limit $g_k(z)$ as $u \to 1^-$. Moreover, one has

$$T(z,u) = \sum_{n\ge 0} t_n(u) z^n \quad \text{with} \quad t_n(u) = \frac{1}{(u-1)^k}\left(s_n(u) - \sum_{i=0}^{k-1} s_n^{(i)}(1)\frac{(u-1)^i}{i!}\right),$$

where the $t_n(u)$ are polynomials in u. Now, if the $s_n(u)$ have nonnegative coefficients, then so do the $t_n(u)$. This results directly from rewriting $t_n(u)$ as $\Delta^k s_n(u)$, where Δ is the difference operator

$$\Delta p(u) := \frac{p(u) - p(1)}{u-1},$$

which preserves posivity of coefficients. Thus, part (i) of the statement applies to the function $T(z,u)$, giving

$$\frac{1}{k!} s_n^{(k)}(1) = \lim_{u\to 1} t_n(u) = [z^n] g_k(z),$$

so that the conclusion is satisfied for k. $\square$

In summary, Lemma 3 asserts that, under suitable conditions,

$$[(u-1)^k]\left([z^n]S^{\text{ana}}(z,u)\right) = [z^n]\left([(u-1)^k]S^{\text{asy}}(z,u)\right).$$

There, the notations stress the fact that an object $S(z,u)$ is taken either as an analytic function S^{ana} at $(0,0)$ or as the corresponding asymptotic expansion S^{asy} as $u \to 1$. The coefficient notations $[(u-1)^k]$ are to be interpreted accordingly.

1.4 Positivity of the Graphical Divergent Expansions

Finally, as we show now, the function $\mathcal{Q}(z,q)$ satisfies the conditions of Lemma 3. This corresponds to a supplementary positivity property, itself established by a specific *external* argument based on depth-first search traversal of graphs and inversions in trees [13].

Lemma 4. *Assume that, pointwise for each z with $|z| < e^{-1}$, and as $q \to 0^+$, the bivariate generating function of connected graphs $\mathcal{Q}(z,q)$ satisfies an asymptotic expansion,*

$$\mathcal{Q}(z,q) \underset{q\to 0^+}{\sim} \sum_{\ell\geq 0} \mathcal{W}_{\ell-1}(z)(-q)^{\ell}, \qquad |z| < e^{-1}, \tag{18}$$

for a sequence of functions $\mathcal{W}_\ell(z)$. Then, for each ℓ, the identity $\mathcal{W}_\ell(z) = W_\ell(z)$ holds, where $W_\ell(z)$ is defined algebraically by (13)*.*

Proof. Let $U_{n,\ell}$ denote the number of unrooted labelled trees with n vertices and ℓ inversions[4] and let

$$U_n(u) := \sum_{\ell\geq 0} U_{n,\ell}u^{\ell} \qquad \text{and} \qquad U(z,u) := \sum_{n\geq 0} U_n(u)\frac{z^n}{n!},$$

be the corresponding generating functions. A tree with n vertices has at most $\binom{n-1}{2}$ inversions so that the $U_n(u)$ are polynomials; the polynomials also have positive coefficients given their combinatorial origin. Moreover $U(z,u)$ is analytic for $|z| < e^{-1}$ and $|u| < 1$ since the number of unrooted trees is *a priori* bounded from above by n^n. (A tree is specified by n daughter-to-mother links).

There is an elegant relation between inversions in unrooted trees and connected graphs discovered by Ira Gessel and Da Lun Wang [13] who proved combinatorially the formal power series relation $U(z,1+q) = Q(z,-q)$. At the level of coefficients, this relation reads

$$U_n(1+q) = -q \sum_{\substack{g \text{ connected graph} \\ |g|=n}} \pi(g).$$

In essence, a connected graph may be considered as rooted at 1. From this root node, a depth first search traversal (with a suitable ordering on successor nodes) gives rise to a tree together with additional return edges that have to correspond to inversions. Conversely, each inversion in a tree may or may not be "activated" depending on the particular graph under consideration, and this fact is seen to be reflected by the relation $U(z,1+q) = Q(z,-q)$. (The negative argument $-q$ in Q is there since we have adopted a negative variable to mark excess.) As a consequence, we have the fundamental relation

$$Q(z,q) = U(z,1-q). \tag{19}$$

[4] An inversion is a pair of vertices (i,j) such that $1 < i < j$ and j is on the branch from i to 1.

By direct combinatorics, $U(z,u)$ has positive coefficients at $(0,0)$ and is bivariate analytic in $|z|<e^{-1}, |u|<1$. Now an asymptotic expansion as $q \to 0^+$,

$$\mathcal{Q}(z,q) \underset{q\to 0^+}{\sim} \sum_{\ell\geq 0} \mathcal{W}_{\ell-1}(z)(-q)^\ell,$$

can be recast via the relation (19) as an expansion of $U(z,u)$ as $u \to 1^-$,

$$U(z,u) \underset{u\to 1^-}{\sim} \sum_{\ell\geq 0} \mathcal{W}_{\ell-1}(z)(u-1)^\ell.$$

Thus, thanks to this combinatorial bijection, we have reached a stage where the necessary positivity condition for application of Lemma 3 holds. Part (i) of Lemma 3 gives us already $W_{-1}(z) = \mathcal{W}_{-1}(z)$ through $[z^n]W_{-1}(z) = [z^n]\mathcal{W}_{-1}(z)$, this without any requirement other than (18). More generally, the equality $W_\ell(z) = \mathcal{W}_\ell(z)$ follows from Part (ii) of Lemma 3 via the identities $[z^n]W_\ell(z) = [z^n]\mathcal{W}_\ell(z)$ valid for all n. □

The discussion above allows us to identify Q and $\mathcal{Q}$. Accordingly, we shall use the notation Q in the rest of this article.

2 Single Saddle-Point Analysis

We now proceed with the estimation of $Q(z,q)$ as $q \to 0^+$, starting from the integral representation (16) of Lemma 2. As will appear shortly, the "tree function" $T(z)$ of (2) is essential in our developments, and we shall accordingly adopt $t = T(z)$ as the main parameter (so that $z = te^{-t}$). In this section, the objective is to prove Wright's expansion (Theorem 1) by an analysis of the integral representation (16) when t is restricted to some fixed interval $(0,a)$ with $a<1$. Precisely, the *single saddle-point* analysis of this section is summarized by an expansion (30) of the form

$$-\frac{1}{q}Q(z,q) + \frac{1}{q}W_{-1}(z) - W_0(z) \sim \sum_{k\geq 1} A_k(t)\left(-\frac{q}{(1-t)^3}\right)^k \qquad (q\to 0^+), \tag{20}$$

where we have set $t = T(z)$.

Our treatment relies on the saddle-point method. Recall that a saddle-point of a function is a zero of its derivative. Locally, the geometry of the modulus of the function resembles a saddle-point. Many integrals of analytic functions have an asymptotic form which is governed by local expansions in the neighbourhood of saddle-points. See [7, 16, 25] for introductions.

The analysis proceeds in four steps: first a modification of the integration contour in the representation of Lemma 2 in order to obtain a saddle-point representation; second, a standard change of variables in order to normalize the saddle-point integrand; third, formal termwise integration; fourth, an analysis of the remainder of the expansion in order to prove that the formal result is indeed an asymptotic expansion of the integral.

2.1 Saddle-point representation

When $q \to 0^+$, the integrand in (16) oscillates more and more wildly, this because of the term $e^{ix\lambda}/q$ it contains. The tactics consist in disposing of the oscillation by shifting the integration contour so as to cross a saddle point. First, we set $x\lambda = w$, which transforms the integral into

$$Q(z,q) = -q \log \frac{1}{\lambda\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{w^2}{2\lambda^2} - z\frac{(1-q)^{-1/2}}{q} e^{iw}\right) dw.$$

The integrand rewrites as

$$\exp\left(-\frac{1}{q}\left(\frac{w^2}{2} + ze^{iw}\right)\right) \cdot \exp\left(\frac{w^2}{2}(q^{-1} - \lambda^{-2}) + ze^{iw}\frac{1-(1-q)^{-1/2}}{q}\right). \tag{21}$$

In this product, the first factor captures the dominant part of the integrand, while the second one acts as a small perturbation since it tends to a finite limit as $q \to 0^+$.

The saddle points ζ of the first factor are located at points ζ such that

$$\frac{d}{dw}\left(\frac{w^2}{2} + ze^{iw}\right)_{w=\zeta} \equiv \zeta + ize^{i\zeta} = 0. \tag{22}$$

We recall first some basic facts concerning the function $T(z)$ which is defined as the solution analytic at 0 of $T = ze^T$. On the boundary $|z| = e^{-1}$ of its disk of convergence, $T(z)$ has a unique singularity $z_0 = e^{-1}$ and $T(z_0) = 1$; moreover as z runs through the real segment $(0, e^{-1})$, $T(z)$ increases from 0 to 1. We recognize in (22) the equation satisfied by $T(z)$, so that, as long as $|z| < e^{-1}$, we can take $t = T(z)$ and obtain a saddle point

$$\zeta = -it = -iT(z).$$

In subsequent computations, $t = T(z)$ is taken as the independent variable (rather than z itself), and is restricted when the need arises to be a real quantity in $(0,1)$. (Analytic continuation makes it possible to extend the domain of validity of end formulæ, if needed.)

The saddle-point method now suggests shifting the line of integration parallel to itself so that it crosses the point ζ. This does not change the value of the integral by virtue of Cauchy's theorem and the fact that the integrand is small as $\Re(w) \to \pm\infty$. Thus, using $t = T(z)$ as a parameter, setting $w = u - it$ and replacing the integration contour on $(-\infty, +\infty)$ yields

$$\begin{aligned} Q(z,q) &= \left(t - \frac{t^2}{2}\right) + \left(1 - \frac{q}{\lambda^2}\right)\frac{t^2}{2} - q \log \frac{I}{\lambda\sqrt{2\pi}}, \\ I &:= \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{q}\left(\frac{u^2}{2} + t(e^{iu} - 1 - iu)\right)\right) h(u)\, du \end{aligned} \tag{23}$$

with

$$h(u) = \exp\left(\left(\frac{u^2}{2} - uit\right)(q^{-1} - \lambda^{-2}) + te^{iu}\frac{1-(1-q)^{-1/2}}{q}\right). \tag{24}$$

The new integral form (23) "explains" the rôle of the tree function in the problem. In effect, it will turn out that the first term in (23) dominates as $q \to 0$, so that it provides the enumeration of unrooted trees, *i.e.*, Part (i) of Theorem 1.

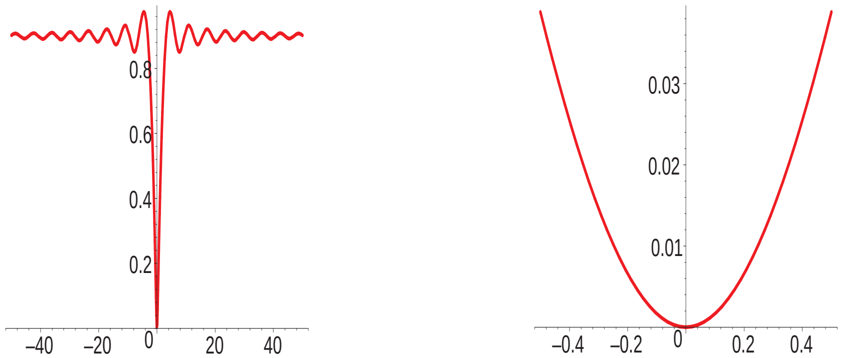

Figure 1: The steepest descent line $\Im(f(u)) = 0$ for $t = 0.9$: general aspect (left) and blow up near 0 (right).

2.2 Change of variable

First, we reduce the kernel of the saddle-point integral to standard *quadratic* form. The corresponding change of variable is defined by the equation

$$y^2 = f(u) \qquad \text{where} \quad f(u) := \frac{u^2}{2} + t(e^{iu} - 1 - iu). \tag{25}$$

We opt to perform the change of variable in such a way that y varies continuously on the real line from $-\infty$ to $+\infty$. Given the geometry of $f(u)$, this corresponds to taking the integral in (23) along the contour depicted in Figure 1—in fact a steepest descent line[5] connecting $-\infty + it$ to $+\infty + it$. (It can be seen that this steepest descent line is the curve $y = t + W(-te^{-t}\sin(x)/x)$, where W is Lambert's function, defined by $W(s)\exp(W(s)) = s$.) The value of the integral remains unaffected by virtue of analyticity and Cauchy's theorem.

The expression of the integral is changed into

$$I = \int_{-\infty}^{+\infty} e^{-y^2/q} H(y)\, dy, \qquad H(y) := h(u(y))\frac{du}{dy}. \tag{26}$$

2.3 Term by term integration

The next step is to expand H as a power series in y and integrate termwise against the Gaussian kernel $\exp(-y^2/q)$. The validity of this process will be proved later. The net result of this formal manipulation is to effect a linear transformation $\mathfrak{L}$ on y-expansions: odd powers of y disappear while even powers are transformed by

$$\mathfrak{L}(y^{2k}) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-y^2/q} y^{2k}\, dy = \frac{1 \cdot 3 \cdots (2k-1)}{2^k} q^{k+1/2}.$$

[5]One of the ways of conducting the saddle-point analysis of $\int e^{-f/q}$ is to cross the saddle-point following steepest descent lines, along which the imaginary part of f is constant, see [25, p. 136] and [7, chap. 5].

In order to compute the series expansion of H, we examine the formulæ induced by the change of variables. The quantity $f(u)/u^2$ is an entire function of u, and

$$f(u) = (1-t)\frac{u^2}{2}\left(1 + \frac{2tu}{1-t}\frac{e^{iu} - 1 - iu + u^2/2}{u^3}\right).$$

The Taylor expansion of its square-root thus has the form

$$y(u) = (1-t)^{1/2}\frac{u}{\sqrt{2}}\sum_{k\geq 0} a_k(t)\left(\frac{u}{1-t}\right)^k,$$

with $a_0(t) = 1$ and $a_k(t)$ a polynomial of degree k in t. The choice of determination is dictated by the direction in which the saddle-point is crossed.

Thus $y(1-t)^{-3/2}$ has a Taylor expansion in powers of $u/(1-t)$ with coefficients that are polynomials in t. Then by reversion of formal power series, the expansion of $u(y)$ is of the form

$$u(y) = (1-t)\sum_{k\geq 1} b_k(t)\frac{y^k}{(1-t)^{3k/2}}, \tag{27}$$

with $b_1(t) = \sqrt{2}$ and coefficients $b_k(t)$ that are again polynomial in t, and the expansion has a positive radius of convergence. Composing expansions then yields

$$H(y) = \frac{\sqrt{2}\,h(0)}{(1-t)^{1/2}}\left(1 + \sum_{k\geq 1} c_k(t,q)\frac{y^k}{(1-t)^{3k/2}}\right), \tag{28}$$

where the coefficients $c_k(t,q)$ are computable polynomials in t which are analytic with respect to q for $|q| < 1$ as follows from the definition of h.

Formally applying the $\mathfrak{L}$ transform to expansion (28) suggests the following divergent asymptotic expansion for Q:

$$\begin{aligned} Q(z,q) \sim \left(t - \frac{t^2}{2}\right) + \left(1 - \frac{q}{\lambda^2}\right)\frac{t^2}{2} - \frac{q}{2}\log\frac{1}{1-t} - \frac{q}{2}\log\frac{q}{\lambda^2} \\ - t(1-(1-q)^{-1/2}) - q\log\left(1 + \sum_{k\geq 1} c_{2k}(t,q)\frac{1\cdot 3\cdots(2k-1)}{2^k}\frac{q^k}{(1-t)^{3k}}\right). \end{aligned} \tag{29}$$

The validity of this asymptotic expansion is proved in §2.4 below. In particular, the dominant term of the integral I provides the enumeration of unicyclic graphs. Expansions at any finite order with respect to q are legitimate and yield finite order expansions (for $q \to 0^+$) of the bivariate generating function of connected graphs counted by size and excess:

$$\begin{aligned} Q(z,q) \underset{q\to 0^+}{\sim} \left(T(z) - \frac{T(z)^2}{2}\right) - \left(\frac{1}{2}\log\frac{1}{1-T(z)} - \frac{T(z)}{2} - \frac{T^2(z)}{4}\right)q \\ + \sum_{k\geq 2}\frac{A_{k-1}(T(z))}{(1-T(z))^{3k-3}}(-q)^k, \end{aligned} \tag{30}$$

where the A_k's are polynomials in $T(z)$.

This gives the results of Cayley and Rényi, as announced in (3), (4), as well as Wright's result as stated in (5).

In addition, this derivation provides a mechanical way to determine the $\mathcal{W}_k(z)$ by a simple process: (i) compute $u(y)$ by Eq. (25) and (27); (ii) determine the compound expansion (28); (iii) integrate termwise by (29) and conclude by expanding the logarithm like in (30). Barely a dozen instructions in a computer algebra system are needed to implement the algorithm. The computation yields in particular $A_1(t) = t^4(6-t)/(24(1-t)^3)$; a table of the first ten A_k (as a function of $\theta = 1-t$) is given in the appendix.

2.4 Analysis

To complete the proof of Theorem 1, we now legitimate term by term integration, thereby establishing that the right-hand side of (30) is an asymptotic expansion of Q for fixed real $t = T(z)$ in $(0,1)$. This is a variant of the classical Laplace method, where the c_k's depend on q, see also [25, p. 376].

Define the function $H_n(q,y)$ by (cf. (28))

$$H(y) = \frac{\sqrt{2}\,h(0)}{(1-t)^{1/2}}\left(1 + \sum_{k=1}^{n-1}\frac{c_k(t,q)y^k}{(1-t)^{3k/2}} + H_n(q,y)\right). \tag{31}$$

For fixed $n > 0$, integration termwise of the inner polynomial in y yields the initial part of the formal expansion of the integral I. To show that this process leads to an *asymptotic* expansion of I when $q \to 0$, it is sufficient to show that

$$\int_{-\infty}^{\infty} e^{-y^2/q} H_n(q,y)\,dy = O(q^{\frac{n+1}{2}}). \tag{32}$$

Since H and the coefficients c_k, $k = 1, \ldots, n-1$ are analytic with respect to q for $|q| < 1$, so is H_n. Moreover, H_n is analytic with respect to y in some neighbourhood of 0 since H is. The bound (32) is obtained using different tools inside and outside the disc of convergence of u defined by (25), whose radius we denote by R.

(a) Consider some $r_0 < R$. For real y with $|y| \geq r_0$, we have

$$|h(u(y))| \leq \exp(Cy^2)$$

where C does not depend on q, in view of the definition of h and the fact that $|y| \sim \sqrt{2}|u|$ for large y. From the change of variable (25), we also get for $|y| \to \infty$

$$\left|\frac{du}{dy}\right| = \left|\frac{2y}{f'(u)}\right| \sim \sqrt{2}|y|$$

so that by continuity there exists C' independent of q such that $|H(y)| \leq \exp(C'y^2)$ for $|y| \geq r_0$. Since H and H_n differ only by a polynomial, such a bound also holds for H_n. Therefore the portion of the integral (32) corresponding to $|y| \geq r_0$ is $O(\exp(-r_0^2/q))$ when $q \to 0$.

(*b*) The bound for $|y| \le r_0$ is obtained by first bounding the coefficients $d_k = c_k/(1-t)^{3k/2}$ from (28) uniformly with respect to q and then using a simple argument of majorizing series. Indeed, these coefficients are expressed by the Cauchy integral

$$d_k = \frac{1}{2i\pi} \oint \frac{F(q,y)}{y^{k+1}}\, dy,$$

where F is directly related to H and is an analytic function of q and y in $|q| < 1$ and $|y| < R$. A valid contour of integration is a circle of radius $r_0 + \delta < R$, on which $|F|$ is uniformly bounded with respect to q (for $|q| < 1/2$) by continuity. This shows that $d_k \le C(r_0+\delta)^{-k}$ for some C that does not depend on q and k. From there follows that $|H_n(q,y)|$ is bounded by $C' r_0^{-n} |y|^n$ for some C' that does not depend on q.

From these two bounds, (32) follows, which concludes the proof of Theorem 1.

3 Coalescing Saddle Points and the Airy Connection

In this section, we construct the generating function of the constants $A_k(1)$ that gives the dominant asymptotics of the number of connected graphs of a fixed excess (Theorem 2, Part (*i*) and Corollary 1). At the same time, we obtain an access to the successive "correction series" (Theorem 2, Part (*ii*)).

The *single saddle-point* analysis of the previous section is summarized by expansion (20) that we now recall *verbatim*

$$-\frac{1}{q}Q(z,q) + \frac{1}{q}W_{-1}(z) - W_0(z) \sim \sum_{k\ge 1} A_k(t)\left(-\frac{q}{(1-t)^3}\right)^k \qquad (q \to 0^+), \tag{33}$$

where $t = T(z)$. Such an expansion holds for t in any closed subinterval of $[0,1)$, for instance $t \in [0, \frac{3}{4}]$. However, the expression becomes obviously meaningless, should t approach 1. Accordingly, the proof of (33) given in Section 2 gives access to successive *lower order terms* of the polynomials $A_k(t)$ near $t = 0$.

Our purpose is to capture the behaviour of the polynomials $A_k(t)$ as t approaches 1. This is achieved by the rescaling

$$\alpha = \frac{q}{(1-t)^3},$$

which turns the right-hand side of (33) into a generating series of the polynomials $A_k(t)$. In this section, this generating series is recovered as an asymptotic expansion as α tends to 0. The coefficients in this asymptotic expansion will provide the desired information on the $A_k(t)$, since in the rescaled expansion it will be possible to let t tend to 1.

The salient new fact here is the need to take into account two saddle points that coalesce as $t \to 1$. The treatment offered here follows in the steps of the single saddle-point analysis of Section 2, with a different change of variables:

— an analysis of the location of dominant saddle points in §3.1;

— a *cubic* change of variables in order to normalize the saddle-point integrand (§3.2);

— formal termwise integration (§3.3);

— an analysis of the remainder of the expansion in order to prove that the formal result is indeed an asymptotic expansion of the integral (§3.4);

— reorganization of the expansion, when t tends to 1 (§3.5).

3.1 Saddle points

The starting point is the integral representation (23):

$$I := \int_{-\infty}^{+\infty} e^{-f(u)/q} h(u)\, du, \quad \text{with} \quad f(u) = \frac{u^2}{2} + (1-\theta)\left(e^{iu} - 1 - iu\right). \tag{34}$$

where we have simply set

$$\theta = 1 - t \equiv 1 - T(z).$$

The quantity h preserves its former meaning from (24), but now with $q = \alpha\theta^3$.

The main estimate of the previous section is of a form (33) that ceases to be valid when θ approaches 0 (*i.e.*, $t \to 1$). One reason is that $f(u)$ becomes locally cubic at $\theta = 0$ instead of being quadratic when $\theta \neq 0$. Solving $f'(u) = 0$ for $u \neq 0$ in the neighbourhood of the origin reveals a "shadow" saddle point ρ that is purely imaginary and satisfies the expansion

$$\rho = -2i\theta\left(1 + \frac{1}{3}\theta + \frac{2}{9}\theta^2 + \frac{22}{135}\theta^3 + \cdots\right). \tag{35}$$

Non-uniformity arises precisely from the coalescence of the two nearby saddle points at 0 and ρ, as $t \to 1$. By construction, $f(0) = 0$, while the value of f at the other saddle point is

$$f(\rho) = -\frac{2}{3}\theta^3 - \frac{2}{3}\theta^4 - \frac{28}{45}\theta^5 - \cdots.$$

We make use of several expansions related to ρ, as well as the position of saddle points other than 0 and ρ. A convenient expression for such quantities is provided through the use of the indexed Lambert $\mathbf{W}$ function [6]. This function is solution of $ye^y = z$. This is a simple variant of the equation (2) defining the tree function $T(z)$ and one has $T(z) = -\mathbf{W}(-z)$. The $\mathbf{W}$ function is multivalued and an index is used to distinguish between different branches: the branch $\mathbf{W}_k$ is defined by

$$\Im \mathbf{W}_k(z) \in [(2k-1)\pi, (2k+1)\pi], \qquad z \to +\infty.$$

Thus the series expansion of $T(z)$ in (2) coincides with that of $-\mathbf{W}_0(-z)$ near the origin, while

$$\rho = i(\mathbf{W}_{-1}(-z) - \mathbf{W}_0(-z)) = i\left(t + \mathbf{W}_{-1}(-te^{-t})\right) = i\left(1 - \theta + \mathbf{W}_{-1}((\theta-1)e^{\theta-1})\right).$$

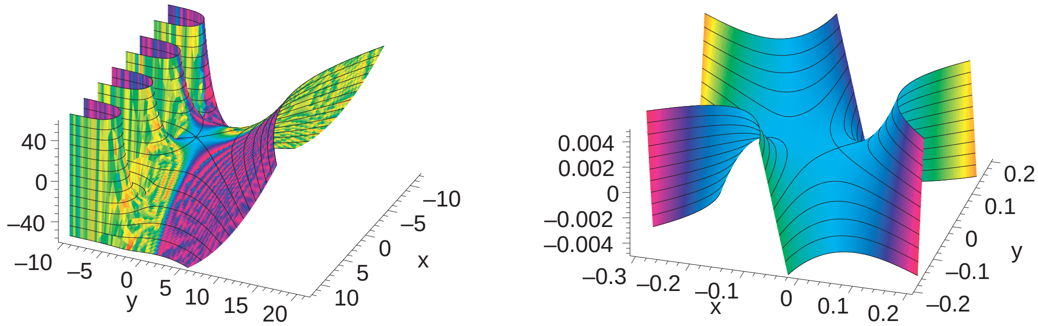

Figure 2: The landscapes of $-\Re f(u)$ and $-\Re P(v)$.

Of course, the $\mathbf{W}_j$ bear no relation to Wright's generating functions W_k or $\mathcal{W}_k$.

Figure 2 illustrates the landscape of $f(u)$ (the next closest saddle points are two symmetrical points σ and $-\bar{\sigma}$, that are given by $\sigma = i(1-\theta+\mathbf{W}_{-2}((\theta-1)e^{\theta-1}))$.)

3.2 Change of variable

In order to estimate asymptotically the integral in (34), the classical method of *coalescent saddle-points* due to Chester, Friedmann and Ursell (see in particular [25, p. 352–356]) is used. Accordingly, we introduce the *cubic* change of variable

$$f(u) = P(v) \qquad \text{where} \quad P(v) = \frac{f(\rho)}{\theta^3}(2v^3+3\theta v^2). \tag{36}$$

The polynomial P is such that P' has two roots at 0 and $-\theta$, $P(0)=0$ and $P(-\theta)=f(\rho)$. Thus P and f behave similarly in the neighbourhood of their two central saddle points, and one expects the change of variable to be conformal in this neighbourhood. Indeed, as illustrated by Figure 2, it is only when approaching the next saddle points σ and $-\bar{\sigma}$ of f that the two landscapes start to diverge qualitatively. Numerical experiments indicate that the change of variables is one-to-one in the domain $0<\theta<1$ and $|u|<|\sigma(0)| = |1+\mathbf{W}_{-2}(-e^{-1})| \approx 7.748360311$. However, in our proofs it will be sufficient to make use of the following instantiation of the general purpose result of [5, Th. 1].

Lemma 5. *There exists $\theta_0>0$ and $r_u>0$ such that the change of variable (36) is one-to-one for (θ,u) in the domain $|\theta|<\theta_0$, $|u|<r_u$.*

Since f does not possess any saddle point outside 0 on the real axis, the change of variable is also one-to-one on the whole domain of integration. The new contour of integration is obtained by following consistently the proper branch of the cubic (36): for real u with large absolute value, f is positive; since $f(\rho)<0$, this forces $\lim\arg(v) = \pm\pi/3$; for small u, $f(u)\sim\theta u^2/2$ so that the contour is vertical in the neighbourhood of $v=0$; finally, $v=-\theta$ corresponding to $u=\rho$ fixes the orientation on the contour. The integral (34) thus admits the exact expression:

$$I = -\int_{e^{-i\pi/3}\infty}^{e^{i\pi/3}\infty} e^{-P(v)/q}G(v)\,dv, \qquad G(v) = h(u(v))\frac{du}{dv}. \tag{37}$$

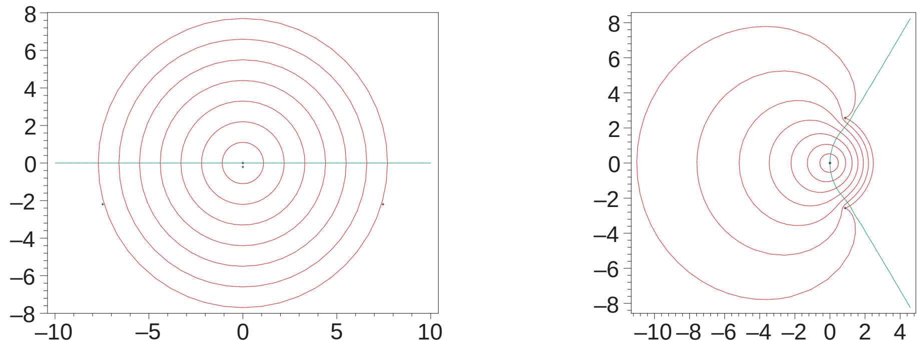

Figure 3: Circles and real axis in the u-plane (left) and their images in the v-plane (right) for $\theta = 1/10$.

Figure 3 displays the images of circles and of the real axis by the cubic change of variable (36). Small crosses on the left correspond to the next saddle points σ and $-\overline{\sigma}$, that are mapped to cusps on the right in the v-plane.

3.3 Term by term integration

Like in the single saddle-point analysis, the second step after the previous changes of variables consists in expanding H as a power series in v

$$G(v, \alpha, \theta) = \sum_{k \geq 0} g_k(\alpha, \theta) v^k, \tag{38}$$

and integrating termwise. The validity of the process is proved later. This formal manipulation reduces to a linear transformation $\mathfrak{M}$ on v-expansions defined by

$$\mathfrak{M}[\phi] := \int_{e^{-i\pi/3}\infty}^{e^{i\pi/3}\infty} e^{-P(v)/q} \phi(v)\, dv. \tag{39}$$

The transformation involves the basic integrals

$$R_k(x) := \int_{e^{-i\pi/3}\infty}^{e^{i\pi/3}\infty} e^{x^{-1}(2v^3+3v^2)} v^k\, dv. \tag{40}$$

Locally, a series expansion for the change of variables is obtained as in the previous section by taking square roots and inverting power series. This yields

$$u = 2i \left(\sqrt{-\frac{3f(\rho)}{2\theta^3}}\, v + \left(\frac{2\sqrt{-\frac{3f(\rho)}{2\theta^3}}}{\theta} - 3\frac{f(\rho)}{\theta^3} + 3\frac{f(\rho)}{\theta^4} \right) \frac{v^2}{6} + \dots \right). \tag{41}$$

Substituting this expansion in $G(u,\alpha,\theta)$ yields (38) with first terms:

$$G(v) = 2ih(0)\sqrt{-\frac{3f(\rho)}{2\theta^3}}\Bigg(1+ \\ (\frac{2}{3\theta}+\sqrt{-\frac{3f(\rho)}{2\theta^3}}(\frac{2}{3}-\frac{2}{3\theta}+2(\theta-1)(\frac{1}{\lambda^2}-\frac{1}{q\sqrt{1-q}})))v+O(v^2)\Bigg).$$

The net result of formally integrating this expansion termwise (*i. e.* applying the $\mathfrak{M}$ transform) is an asymptotic expansion of the fundamental integral I, namely

$$I \underset{\alpha\to 0}{\sim} -\sum_{k\geq 0} g_k(\alpha,\theta)\theta^{k+1}R_k\left(\frac{-\theta^3}{f(\rho)}\alpha\right). \tag{42}$$

That this expansion holds is justified in §3.4 below. This expansion must be viewed as an asymptotic expansion with respect to α, with θ a parameter. Indeed, the coefficients $g_k(\alpha,\theta)$ have finite limits when $\alpha\to 0$, and we show next that the $R_k(x)$ form an asymptotic scale, since they behave essentially like $x^{k/2}$ when $x\to 0$.

The Airy Connection

The asymptotic expansion (42) involves the basic quantities $R_k(x)$ of which the character as $x\to 0$ is needed.

It proves convenient to work with the exponential generating series of the $R_k(x)$, namely

$$R(z) := \sum_{k\geq 0} R_k(x)\frac{z^k}{k!} = \int_{e^{-i\pi/3}\infty}^{e^{i\pi/3}\infty} e^{x^{-1}(2v^3+3v^2)+zv}\,dv.$$

This is to be compared to the classical integral representation of the Airy function,

$$\mathrm{Ai}\,(y) = \frac{1}{2\pi i}\int_{e^{-i\pi/3}\infty}^{e^{i\pi/3}\infty} e^{w^3/3-yw}\,dw.$$

From there, it is easily recognized that the generating series $R(z)$ admits the following expression:

$$R(z) = 2\pi i\left(\frac{x}{6}\right)^{1/3}\exp\left(-\frac{z}{2}+\frac{1}{2x}-\frac{1}{2x}(1-\frac{2}{3}zx)^{3/2}\right)e^{\frac{2}{3}y^{3/2}}\,\mathrm{Ai}\,(y), \tag{43}$$

where $y=(1-\frac{2}{3}zx)\left(\frac{3}{4x}\right)^{2/3}$.

In order to prove that the R_k form an asymptotic scale (as $x\to 0+$), we now show that $R_k(x)$ grows like $x^{\lfloor\frac{k+1}{2}\rfloor+1/2}$. In the product above, the coefficient of z^i in the expansion of the first exponential is of order $x^{\lfloor\frac{i+1}{2}\rfloor}$: the argument of the exponential behaves like $-xz^2/12-x^2z^3/108$ and the powers of this binomial dominate the coefficients of z^i

in the exponential. On the other hand, the function $\phi(y) := \exp(2y^{3/2}/3)\,\mathrm{Ai}\,(y)$ has a Taylor expansion of the form

$$\phi(y) = \sum_{j\geq 0} \frac{\phi^{(j)}(\zeta)}{j!}\left(\frac{2x\zeta z}{3}\right)^j, \quad (z \to 0), \qquad \zeta = \left(\frac{3}{4x}\right)^{2/3}.$$

Now, using the classical asymptotic expansion of the Airy function, it follows that $\phi(u)$ behaves like $u^{-1/4}$ when u tends to $+\infty$, and its jth derivative is of order $u^{-j-1/4}$. Therefore, when $x \to 0+$, $\phi^{(j)}(\zeta)$ is of order $x^{2j/3+1/6}$. Combining this with the factor $(2x\zeta/3)^j$ shows that the coefficient of z^j in this Taylor expansion is of order $x^{j+1/6}$. It follows that the coefficient R_k of z^k in the product is dominated by the term coming from $\phi^{(0)}(\zeta)$ and is asymptotic to

$$R_k(x) \underset{x\to 0+}{\sim} i\sqrt{\frac{\pi}{3}} c_k x^{\lfloor \frac{k+1}{2}\rfloor + 1/2}, \quad \text{where} \quad c_{2k} = \frac{(-1)^k(2k)!}{12^k k!},\ c_{2k+1} = \frac{(-1)^k(2k+3)!}{36(k+1)!12^k}.$$

Also, we note that expanding the generating series (43) in Taylor series around $z = 0$ yields explicit analytic expressions for the integrals R_k in terms of the Airy function and its derivative, for instance,

$$\begin{aligned} R(z) = \pi i \zeta^{-1/2} e^{1/2x}\,\mathrm{Ai}\,(\zeta)\Bigg(1 &- \left(1 + \frac{\Sigma(\zeta)}{\zeta^{1/2}}\right)\frac{z}{2} + \left(1 + \frac{\Sigma(\zeta)}{\zeta^{1/2}}\right)\frac{z^2}{4} \\ &- \left(\frac{1}{\zeta^{3/2}} + 4 + 4\frac{\Sigma(\zeta)}{\zeta^{1/2}}\right)\frac{z^3}{48} + \left(\frac{2}{\zeta^{3/2}} + 4 + (\frac{1}{\zeta^2} + \frac{4}{\zeta^{1/2}})\Sigma(\zeta)\right)\frac{z^4}{192} + O(z^5)\Bigg), \end{aligned}$$

where

$$\zeta = \left(\frac{3}{4x}\right)^{2/3} \quad \text{and} \quad \Sigma(y) = \frac{\mathrm{Ai}'(y)}{\mathrm{Ai}\,(y)}.$$

3.4 Analysis

So far, we have developed the formal asymptotic expansion (42), by integrating termwise, and without analytic justification. We now show that the series (42) is indeed asymptotic to the integral I. The proof is similar to that of §2.4, but is technically more demanding because of uniformity issues.

By Lemma 5, we know that the cubic change of variables (36) is one-to-one for $|\theta| < \theta_0$ and either $|u| < r_u$ or u real. Set $r_v = \max_{|u|=r_u} |v(u)|$ and $\alpha_0 = \theta_0^{-3}$. We shall make use of the following lemma.

Lemma 6. *For $|\theta| < \theta_0$, $|\alpha| < \alpha_0$, $|v| < r_v$, the function $G(v, \alpha, \theta)$ is an analytic function of its three arguments.*

Proof. The change of variables is analytic and does not involve α. By definition, $G(v, \alpha, \theta) = h(u(v))du/dv$. The expression of h shows that it is analytic in u, θ and $q = \alpha\theta^3$ provided $|q| < 1$, which concludes the proof. $\square$

We now define $G_n(\alpha,\theta,v)$ as the remainder in expansion (38) (note the parallel with (31)),

$$G(v) = \sum_{k=0}^{n-1} g_k(\alpha,\theta)v^k + v^n G_n(\alpha,\theta,v). \tag{44}$$

The proof that (42) is asymptotic to I depends on the following lemma.

Lemma 7. *Let θ be $[0,\theta_0)$ and $\alpha > 0$, then*

$$I_n(\alpha,\theta) := \int_{e^{-i\pi/3}\infty}^{e^{i\pi/3}\infty} e^{-P(v)/q} v^n G_n(\alpha,\theta,v)\, dv = \theta^n O(\alpha^{n/2}), \tag{45}$$

where the $O(.)$ error term is uniform with respect to θ.

Proof. As in the single saddle-point analysis, the proof of the bound on I_n is obtained by bounding $|G_n(\alpha,\theta,v)|$ in two different regions. The contour of integration is the line $\Im(P(v)) = 0$. We deform this contour into one made of three pieces: a vertical segment from $-ir_v/2$ to $ir_v/2$; arcs of the circle $|v| = r_v/2$ from the endpoints of this segment to the original contour (that we denote by ξ and $\overline{\xi}$); the rest of this contour to $\exp(\pm i\pi/3)\infty$. In a "small v region" (comprising the first two pieces), we compute a bound on the coefficients g_k and then use a majorizing series argument. For larger v we consider the behaviour of $G(\alpha,\theta,v)$ when $|v|$ is large but remains on the contour of (45).

(a) *Large v region.* For $|v| \geq r_v/2$, using the change of variables (36) and a reasoning similar to that of Section 2.4, we get that $|G_n(\alpha,\theta,v)| < \exp(CP(v))$, where C is a positive constant. Injecting this bound into

$$\int_{\xi}^{+\infty} e^{-P(v)/q} G_n(v)\, dv,$$

leads to a bound $\exp(-C'\xi/\alpha\theta^3)$ for the remaining part of the integral, which is sufficient for this domain.

(b) *Small v region.* The coefficients g_k are given by Cauchy's formula

$$g_k(\alpha,\theta) = \frac{1}{2i\pi} \oint \frac{G(\alpha,\theta,v)}{v^{k+1}}\, dv,$$

where the contour is for instance a circle centered at the origin with radius $r \leq r_v$. For $|\alpha| < \alpha_0$ and $\theta < \theta_0$, G being analytic is uniformly bounded and thus there is a constant M such that $|g_k(\alpha,\theta)| \leq M/r^k$. From there it follows that

$$|G_n(\alpha,\theta,v)| \leq M\frac{r^{-n+1}}{r-|v|}, \tag{46}$$

as long as $|v| < r$.

On the vertical segment $(-ir_v/2, ir_v/2)$, the integral I_n is, thanks to (46), bounded by

$$\left|2\int_{-r_v/2\theta}^{r_v/2\theta} e^{3\frac{f(\rho)}{\alpha\theta^3}w^2}\theta^{n+1}w^n M r_v^{-n}\,dw\right| \leq 2Mr_v^{-n}\theta^{n+1}\left(\frac{-\alpha\theta^3}{3f(\rho)}\right)^{\frac{n+1}{2}}\Gamma\left(\frac{n+1}{2}\right).$$

By conjugacy, it suffices to discuss the arc situated in the upper half-plane. There, a direct computation leads to the bound

$$B = \left|\int_{\theta_0}^{\pi/2}\exp\left(-\frac{f(\rho)}{4\alpha\theta^6}(r_v^3\cos(3\phi)+3\theta r_v^2\cos(2\phi))\right)\left(\frac{r_v}{2}\right)^{n+1}Mr_v^{-n}\,d\phi\right|.$$

At this stage, we observe that the intersection of the circle $|v| = r_v/2$ and the curve $\Im(P(v)) = 0$ has an argument θ_0 which is properly contained in the interval $(\pi/3, \pi/2)$. Consequently, the first cosine in B is negative, while second one is bounded from above by $-1/2$. It follows that

$$B \leq 2^{-n}M\exp\left(\frac{Kf(\rho)}{\alpha\theta^5}\right),$$

for some positive constant K.

□

As a consequence of the previous lemma, the coefficients $g_k(\alpha,\theta)$ in (38) are analytic for $|\theta| < \theta_0$ and $|\alpha| < \alpha_0$. The uniform bound on I_n then legitimates expanding the coefficients in (42) with respect to α and reorganizing the truncated series.

3.5 Final reorganizations

We now conclude the proof of Theorem 2 by showing how the asymptotic expansion can be reorganized by collecting powers of θ.

In the previous section, the expansion (42) has been established. Since both the coefficients g_k and the scale functions $R_k(-\alpha\theta^3/f(\rho))$ are analytic with respect to θ in the neighbourhood of the origin, these quantities can be expanded as Taylor series in powers of θ. The presence of the factor θ^{k+1} implies that only a finite number of terms in (42) contribute to each term of the expansion of I in powers of θ. Thus, we get:

$$I = 2^{4/3}e^{-1/2}e^{\frac{1}{3\alpha}}\alpha^{1/3}\,\mathrm{Ai}\left((2\alpha)^{-2/3}\right)\times \\ \left(\theta + \left(\frac{1}{3\alpha}-\frac{1}{2}+2^{1/3}\alpha^{1/3}\frac{\mathrm{Ai}'((2\alpha)^{-2/3})}{\mathrm{Ai}\left((2\alpha)^{-2/3}\right)}(\frac{1}{3\alpha}-\frac{7}{6})\right)\theta^2+\dots\right).$$

A simpler view of this expression is achieved by introducing

$$S := \frac{{}_2F_0\left(\begin{matrix}\frac{5}{6},\frac{7}{6}\\ -\end{matrix}\middle| -\frac{3\alpha}{2}\right)}{{}_2F_0\left(\begin{matrix}\frac{1}{6},\frac{5}{6}\\ -\end{matrix}\middle| -\frac{3\alpha}{2}\right)} = -\frac{2}{\alpha}\left(1+(2\alpha)^{1/3}\frac{\mathrm{Ai}'((2\alpha)^{-2/3})}{\mathrm{Ai}((2\alpha)^{-2/3})}\right) = 1+\frac{95}{288}\alpha+O(\alpha^2),$$

where ${}_2F_0$ is the classical hypergeometric series

$$ {}_2F_0\left(\begin{matrix} a,b \\ - \end{matrix} \middle| z\right) = \sum_{n\geq 0} \frac{(a)_n (b)_n}{n!} z^n, \qquad (\lambda)_n := \lambda \cdots (\lambda + n - 1). $$

Then, this expansion rewrites

$$ I = 2^{4/3} e^{-1/2} e^{\frac{1}{3\alpha}} \alpha^{1/3} \operatorname{Ai}\left((2\alpha)^{-2/3}\right) \left(\theta + \left(\frac{2}{3} + \left(-\frac{1}{6} + \frac{7}{12}\,\alpha\right) S\right) \theta^2 \right. $$
$$ \left. + \left(-\frac{1}{360} \frac{(105\alpha^2 - 54\,\alpha + 20)\,S}{\alpha} - \frac{1}{180} \frac{17\,\alpha - 10}{\alpha}\right) \theta^3 + \dots \right) $$

We know that $Q(z,q)$ and I are related via

$$ Q(z,q) \sim \left(t - \frac{t^2}{2}\right) + \left(1 - \frac{\alpha\theta^3}{\lambda(\alpha\theta^3)^2}\right) \frac{t^2}{2} - \alpha\theta^3 \log\left(\frac{I}{\lambda(\alpha\theta^3)\sqrt{2\pi}}\right). \tag{47} $$

At this stage, it suffices to compose expansions in a routine way in order to obtain the final expansion of $Q(z, \alpha\theta^3)$. This yields

$$ Q(z,q) - (t - t^2/2) + \left(\frac{1}{2} \log \frac{1}{1-t} - \frac{t}{2} - \frac{t^2}{4}\right) q = -\alpha\theta^3 \ln\left({}_2F_0\left(\begin{matrix} \frac{5}{6}, \frac{1}{6} \\ - \end{matrix} \middle| -\frac{3\alpha}{2}\right)\right) $$
$$ - \frac{\theta^4}{12} (2\alpha + \alpha(7\alpha - 2)S) $$
$$ + \frac{\theta^5}{360} \left(\frac{5}{4}\alpha(7\alpha-2)^2 S^2 + (245\alpha^2 - 94\alpha + 20)S + 114\alpha - 20\right) $$
$$ - \frac{\theta^6}{20160} \left(\frac{35}{9}\alpha(7\alpha-2)^3 S^3 + \frac{14}{3}(7\alpha-2)(245\alpha^2 - 94\alpha + 20)S^2 \right. $$
$$ \left. + \frac{56}{9\alpha}(2019\alpha^3 - 840\alpha^2 + 144\alpha - 40)S - \frac{56}{9\alpha}(733\alpha^3 - 854\alpha^2 + 164\alpha - 40)\right) + \cdots, $$

For $j \geq 0$, the coefficient of θ^{3+j} in this expansion is precisely the generating function of the $A_k^{(j)}(1)$. The first five generating functions are given in Appendix II. As a final check, expanding the coefficient of θ^4 in the above expansion of Q with respect to α, one gets

$$ -\frac{19}{24}\alpha^2 + \frac{65}{48}\alpha^3 - \frac{1945}{384}\alpha^4 + \frac{21295}{768}\alpha^5 - \frac{603965}{3072}\alpha^6 + \frac{10454075}{6144}\alpha^7 + O\left(\alpha^8\right), $$

from which it is easy to recognize the coefficients of θ^2 in the polynomials A_k of Appendix I, that were previously obtained by the single saddle-point expansion.

Acknowledgement. We are thankful to Gaël Rémond for suggesting that a positivity property might help in Section 1. This work was supported in part by the ALCOM-FT project (contract IST-1999-14186) and by the IHRP Programme (grant HPRN-CT-2001-00272: *Algebraic Combinatorics in Europe*) of the European Union.

We would like to extend special thanks to an anonymous referee for a very thorough scrutiny of the paper and for numerous suggestions that (hopefully) resulted in a better organization of this article.

References

[1] Abramowitz, M., and Stegun, I. A., Eds. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Dover Publications Inc., New York, 1992. Reprint of the 1972 edition.

[2] Banderier, C., Flajolet, P., Schaeffer, G., and Soria, M. Random maps, coalescing saddles, singularity analysis, and Airy phenomena. *Random Structures & Algorithms 19*, 3/4 (2001), 194–246.

[3] Bender, E. A., Canfield, E. R., and McKay, B. D. The asymptotic number of labeled graphs with given number of vertices and edges. *Random Structures & Algorithms 1*, 2 (1990), 127–169.

[4] Bollobás, B. *Random Graphs*. Academic Press, 1985.

[5] Chester, C., Friedman, B., and Ursell, F. An extension of the method of steepest descents. *Proceedings of the Cambridge Philosophical Society 53* (1957), 599–611.

[6] Corless, R. M., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. J., and Knuth, D. E. On the Lambert W function. *Advances in Computational Mathematics 5*, 4 (1996), 329–359.

[7] De Bruijn, N. G. *Asymptotic Methods in Analysis*. Dover, 1981. A reprint of the third North Holland edition, 1970 (first edition, 1958).

[8] Flajolet, P., Grabner, P. J., Kirschenhofer, P., and Prodinger, H. On Ramanujan's Q-function. *Journal of Computational and Applied Mathematics 58*, 1 (1995), 103–116.

[9] Flajolet, P., Knuth, D. E., and Pittel, B. The first cycles in an evolving graph. *Discrete Mathematics 75* (1989), 167–215.

[10] Flajolet, P., and Louchard, G. Analytic variations on the Airy distribution. *Algorithmica 31*, 3 (2001), 361–377.

[11] Flajolet, P., and Noy, M. Analytic combinatorics of chord diagrams. In *Formal Power Series and Algebraic Combinatorics* (2000), D. Krob, A. A. Mikhalev, and A. V. Mikhalev, Eds., pp. 191–201. Proceedings of the 12th International Conference, FPSAC'2000; June 2000, Moscow.

[12] Flajolet, P., Poblete, P., and Viola, A. On the analysis of linear probing hashing. *Algorithmica 22*, 4 (Dec. 1998), 490–515.

[13] Gessel, I., and Wang, D. L. Depth-first search as a combinatorial correspondence. *Journal of Combinatorial Theory,* Series A *26*, 3 (1979), 308–313.

[14] Harary, F., and Palmer, E. M. *Graphical Enumeration.* Academic Press, 1973.

[15] Henrici, P. *Applied and Computational Complex Analysis.* John Wiley, New York, 1977. 3 volumes.

[16] Henrici, P. *Applied and Computational Complex Analysis.* John Wiley, New York, 1977. 3 volumes.

[17] Hille, E. *Analytic function theory.* Ginn and Company, Boston, 1959. 2 volumes.

[18] Janson, S., Knuth, D. E., Łuczak, T., and Pittel, B. The birth of the giant component. *Random Structures & Algorithms 4*, 3 (1993), 233–358.

[19] Knuth, D. E. *The Art of Computer Programming*, 3rd ed., vol. 1: Fundamental Algorithms. Addison-Wesley, 1997.

[20] Knuth, D. E. Linear probing and graphs. *Algorithmica 22*, 4 (Dec. 1998), 561–568.

[21] Knuth, D. E., and Pittel, B. A recurrence related to trees. *Proceedings of the American Mathematical Society 105*, 2 (Feb. 1989), 335–349.

[22] Louchard, G. The Brownian excursion: a numerical analysis. *Computers and Mathematics with Applications 10*, 6 (1984), 413–417.

[23] Mallows, C. L., and Riordan, J. The inversion enumerator for labeled trees. *Bulletin of the American Mathematical Society 1968* (74), 92–94.

[24] Monasson, R. Statistical physics of the random graph model. In *Algorithms Seminar 97-98* (1998), B. Salvy, Ed., vol. 3504 of *Inria Research Report*, pp. 57–62. Summary by Philippe Flajolet.

[25] Olver, F. W. J. *Asymptotics and Special Functions.* Academic Press, 1974.

[26] Riddell, Jr., R. J., and Uhlenbeck, G. E. On the theory of the virial development of the equation of state of mono-atomic gases. *Journal of Chemical Physics 21* (1953), 2056–2064.

[27] Spencer, J. Enumerating graphs and Brownian motion. *Communications on Pure and Applied Mathematics 50* (1997), 293–296.

[28] Takács, L. A Bernoulli excursion and its various applications. *Advances in Applied Probability 23* (1991), 557–585.

[29] Takács, L. Conditional limit theorems for branching processes. *Journal of Applied Mathematics and Stochastic Analysis 4*, 4 (1991), 263–292.

[30] Takács, L. On a probability problem connected with railway traffic. *Journal of Applied Mathematics and Stochastic Analysis 4*, 1 (1991), 1–27.

[31] Temperley, H. N. V. On the enumeration of the Mayer cluster integrals. *Proc. Phys. Soc. Sect. B. 72* (1959), 1141–1144.

[32] Whittaker, E. T., and Watson, G. N. *A Course of Modern Analysis*, fourth ed. Cambridge University Press, 1927. Reprinted 1973.

[33] Wilf, H. S. *Generatingfunctionology.* Academic Press, 1990.

[34] Wright, E. M. The number of connected sparsely edged graphs. *Journal of Graph Theory 1* (1977), 317–330.

[35] Wright, E. M. The number of connected sparsely edged graphs. II. Smooth graphs. *Journal of Graph Theory 2* (1978), 299–305.

[36] Wright, E. M. The number of connected sparsely edged graphs. III. Asymptotic results. *Journal of Graph Theory 4* (1980), 393–407.

Appendix I: Polynomials A_k for $k = 1, \ldots, 10$

The numerator polynomials of the generating function (5) of connected graphs with excess k. In this list, $\theta = 1 - t$.

$$A_1 = \frac{1}{24}(5 - 19\theta + 26\theta^2 - 14\theta^3 + \theta^4 + \theta^5)$$

$$A_2 = \frac{1}{48}(15 - 65\theta + 108\theta^2 - 87\theta^3 + 42\theta^4 - 23\theta^5 + 12\theta^6 - \theta^7 - \theta^8)$$

$$\begin{aligned} A_3 = \frac{1}{5760}(&5525 - 291175\theta + 635330\theta^2 - 745600\theta^3 + 535740\theta^4 - 273780\theta^5 + 115040\theta^6 - 40200\theta^7 \\ &+ 17250\theta^8 - 8790\theta^9 + 780\theta^{10} + 760\theta^{11}) \end{aligned}$$

$$\begin{aligned} A_4 = \frac{1}{11520}(&50850 - 3194250\theta + 8616850\theta^2 - 13155250\theta^3 + 12771850\theta^4 - 8606120\theta^5 + 4415260\theta^6 - 1857860\theta^7 \\ &+ 669640\theta^8 - 219770\theta^9 + 65770\theta^{10} - 24810\theta^{11} + 12410\theta^{12} - 1140\theta^{13} - 1080\theta^{14}) \end{aligned}$$

$$\begin{aligned} A_5 = \frac{1}{2903040}(&78269625 - 5707469250\theta + 18335929500\theta^2 - 34319170900\theta^3 + 41952020950\theta^4 - 35962324230\theta^5 \\ &+ 23020806760\theta^6 - 11708714080\theta^7 + 4975132830\theta^8 - 1830034590\theta^9 + 601177020\theta^{10} - 180425700\theta^{11} + 51474010\theta^{12} \\ &- 13991530\theta^{13} + 4851840\theta^{14} - 2390440\theta^{15} + 224440\theta^{16} + 207120\theta^{17}) \end{aligned}$$

$$\begin{aligned} A_6 = \frac{1}{5806080}(&1189944000 - 98791008750\theta + 367787936250\theta^2 - 813474509750\theta^3 + 1197755834450\theta^4 \\ &- 1253454584550\theta^5 + 980040640250\theta^6 - 599982951190\theta^7 + 300815966010\theta^8 - 128224952010\theta^9 + 47813205590\theta^{10} \\ &- 15961635210\theta^{11} + 4860196070\theta^{12} - 1374594530\theta^{13} + 366876510\theta^{14} - 95322290\theta^{15} + 24215710\theta^{16} - 7898880\theta^{17} \\ &+ 3842520\theta^{18} - 366200\theta^{19} - 330000\theta^{20}) \end{aligned}$$

$$\begin{aligned} A_7 = \frac{1}{1393459200}(&2596113838125 - 241701146268750\theta + 1022318885542500\theta^2 - 2605794622935000\theta^3 \\ &+ 4485353216988000\theta^4 - 5554329073622000\theta^5 + 5170942668589600\theta^6 - 3757809635925200\theta^7 \\ &+ 2210369988465100\theta^8 - 1089182823566900\theta^9 + 463046429450440\theta^{10} - 173878558351520\theta^{11} \\ &+ 58778486002120\theta^{12} - 18166181256440\theta^{13} + 5202430882400\theta^{14} - 1397249117680\theta^{15} \\ &+ 356490564290\theta^{16} - 87521266990\theta^{17} + 21236878580\theta^{18} - 5125789000\theta^{19} + 1594511240\theta^{20} \\ &- 767254120\theta^{21} + 73908320\theta^{22} + 65189760\theta^{23}) \end{aligned}$$

$$\begin{aligned} A_8 = \frac{1}{2786918400}(&54927280170000 - 5666278194281250\theta + 26824692404396250\theta^2 - 77366822213126250\theta^3 \\ &+ 152382297023475750\theta^4 - 218156583852371500\theta^5 + 236573726054155000\theta^6 - 200727482616511200\theta^7 \\ &+ 137253126997243600\theta^8 - 77849078308877100\theta^9 + 37645270228995980\theta^{10} - 15897201695720300\theta^{11} \\ &+ 5979980926603380\theta^{12} - 2036617471852960\theta^{13} + 636438562166760\theta^{14} - 184559490851000\theta^{15} \\ &+ 50154335417480\theta^{16} - 12888752686210\theta^{17} + 3160124190570\theta^{18} - 746834528570\theta^{19} \\ &+ 172068129110\theta^{20} - 39597102660\theta^{21} + 9176259280\theta^{22} - 2746991960\theta^{23} + 1309275160\theta^{24} \\ &- 127127840\theta^{25} - 109979520\theta^{26}) \end{aligned}$$

$$\begin{aligned} A_9 = \frac{1}{367873228800}(&87498905321953125 - 9905013754058981250\theta + 51871787359475737500\theta^2 \\ &- 166939400227582867500\theta^3 + 370214294152803581250\theta^4 - 601974300849169652250\theta^5 \\ &+ 746889757209202915000\theta^6 - 728352769734378481000\theta^7 + 572521530879543699300\theta^8 \\ &- 371538886214523778300\theta^9 + 203811478351577107000\theta^{10} - 96659157670498208280\theta^{11} \\ &+ 40436790240804091860\theta^{12} - 15178822141481801220\theta^{13} + 5185720095623707200\theta^{14} \\ &- 1631682514348900640\theta^{15} + 477579026952905450\theta^{16} - 131143668988403370\theta^{17} \\ &+ 34042117499108060\theta^{18} - 8411416734923180\theta^{19} + 1991881935357930\theta^{20} \\ &- 455320478552610\theta^{21} + 101351099827560\theta^{22} - 22200372140760\theta^{23} \\ &+ 4895021570560\theta^{24} - 1097042118040\theta^{25} + 318050227440\theta^{26} - 150316250800\theta^{27} \\ &+ 14684029120\theta^{28} + 12481662720\theta^{29}) \end{aligned}$$

$$\begin{aligned} A_{10} = \frac{1}{735746457600}(&2372826356485200000 - 292406946001350468750\theta + 1678039108549792931250\theta^2 \\ &- 5960044156291372743750\theta^3 + 14695394664078877691250\theta^4 - 26763633242744257571250\theta^5 \\ &+ 37444459561690183598750\theta^6 - 41390270455117959394250\theta^7 + 36974188060114422097750\theta^8 \\ &- 27242206694526564621100\theta^9 + 16886849649201908907300\theta^{10} - 8981951090343948957900\theta^{11} \\ &+ 4177516517534052648780\theta^{12} - 1728391230053110836180\theta^{13} + 645664138316670991980\theta^{14} \\ &- 220548396362568493780\theta^{15} + 69625446183552011500\theta^{16} - 20498409715101617990\theta^{17} \\ &+ 5671697993833056410\theta^{18} - 1484772265270065150\theta^{19} + 369960911518060650\theta^{20} \\ &- 88225286357617770\theta^{21} + 20243582444191590\theta^{22} - 4494000904793730\theta^{23} + 971193953017310\theta^{24} \\ &- 205930131107360\theta^{25} + 43273279907440\theta^{26} - 9208367019280\theta^{27} + 2006247267000\theta^{28} \\ &- 565935053040\theta^{29} + 265458859440\theta^{30} - 26053859520\theta^{31} - 21793017600\theta^{32}). \end{aligned}$$

Appendix II: Polynomials $\mathcal{A}^{(j)}$ for $j = 1, \ldots, 5$

The generating functions $\sum_k A_k^{(j)} x^k$ providing the correction terms in formula (7) are given by the polynomials $\mathcal{A}^{(j)}(x, v)$ evaluated at

$$v = S(x) = -\frac{2}{x}\left(1 + (2x)^{1/3}\frac{\mathrm{Ai}'((2x)^{-2/3})}{\mathrm{Ai}((2x)^{-2/3})}\right).$$

$$\begin{aligned}
\mathcal{A}^{(1)} &= \frac{1}{12}\,(2x + x(7x-2)v) \\
\mathcal{A}^{(2)} &= \frac{1}{1440}\left(5x(-2+7x)^2v^2 + (980x^2 + 80 - 376x)v - 80 + 456x\right) \\
\mathcal{A}^{(3)} &= \frac{-1}{25920x}\left(5x^2(-2+7x)^3v^3 + 6(-2+7x)(245x^2+20-94x)v^2x + \right. \\
&\qquad \left. (-320 + 16152x^3 - 6720x^2 + 1152x)v + 320 - 5864x^3 + 6832x^2 - 1312x\right) \\
\mathcal{A}^{(4)} &= \frac{1}{43545600x^2}\left(525x^3(-2+7x)^4v^4 + 840(245x^2+20-94x)(-2+7x)^2v^3x^2\right. \\
&\qquad + 56(2800 - 22640x + 462735x^4 - 336540x^3 + 109668x^2)v^2x \\
&\qquad + (-62720x - 12969216x^3 + 27938400x^4 + 2989952x^2 + 89600 - 9632880x^5)v \\
&\qquad \left. +8786688x^3 - 19265760x^4 + 17920x - 1811712x^2 - 89600\right) \\
\mathcal{A}^{(5)} &= \frac{-1}{130636800x^3}\left(105x^4(-2+7x)^5v^5 + 210(245x^2+20-94x)(-2+7x)^3v^4x^3\right. \\
&\qquad + 28(-2+7x)(321405x^4 - 237360x^3 + 82788x^2 - 16960x + 2000)v^3x^2 \\
&\qquad - 4(-232960x + 33600 + 20242740x^4 + 1524992x^2 - 7258616x^3 - 27277500x^5 + 4214385x^6)v^2x \\
&\qquad + (10940288x^3 - 107520x + 94748736x^5 - 2745344x^2 - 42289728x^4 - 35840 - 67430160x^6)v \\
&\qquad \left. +35840 + 1589504x^2 + 23671872x^4 - 6668160x^3 + 197120x - 56486304x^5\right).
\end{aligned}$$