Algorithms Seminar 2002–2004,
F. Chyzak (ed.), INRIA, (2005), pp. 77–81.

Available online at the URL
`http://algo.inria.fr/seminars/`.

Airy Phenomena and the Number of Sparsely Connected Graphs

Bruno Salvy
Algorithms Project, INRIA (France)

December 15, 2003

Summary by Vlady Ravelomanana

Abstract

The enumeration of connected graphs by excess (number of edges minus number of vertices) is a well-understood problem usually dealt by means of combinatorial decompositions or indirect formal series manipulations. In their work [3], Philippe Flajolet, Bruno Salvy, and Gilles Schaeffer derive such enumerative results mainly using analytic methods. In particular, they exhibit strong connections between Airy functions and the complete asymptotic expansions of the number of connected graphs of fixed excesses.

1. Introduction

It was Sir Edward M. Wright (1906–2005), of Hardy and Wright fame,[1] who initiated the enumeration of labelled connected graphs by number of vertices and edges [7, 8, 9]. The enumeration of graphs according to these two parameters has a long history which goes back to Cayley and whose main steps can be summarized as follows:

Author	Year	Results
Cayley	1889	number of unrooted trees
Rényi	1960	number of unicyclic graphs
Wright	1977	number of general connected graphs

These combinatorial problems are closely related to the theory of random graphs [2, 1, 4]. The starting point of Flajolet, Salvy and Schaeffer's work is the divergent series of connected labelled graphs

$$C(z,q) = \log\left(1 + \sum_{n=1}^{\infty}(1+q)^{\binom{n}{2}}\frac{z^n}{n!}\right). \tag{1}$$

(Throughout this abstract the variable z marks the number of vertices and the variable q reflects the number of edges.) Though Equation (1) can be viewed as an application of symbolic methods in combinatorial analysis [6], it is worthnoting that this series diverges for any $q > 0$. However, the paper [3] shows how to work with (1) mainly using analytical tools from asymptotic analysis. This abstract is divided into three parts as follows: (i) integral representation of (1), (ii) asymptotic expansions via standard saddle-point method and (iii) double saddle-point expansions and Airy functions.

[1]Sir Edward Wright was knighted in 1977 and a building in the University of Aberdeen beared in his very life. While his research career was striking long (1930–1980) and fruitful, he was also an excellent university admnistrator.

2. Formal Expressions and Integral Representations

Denote by $G(z,q)$ the bivariate EGF of labelled graphs (connected or not) then classically we obtain in $\mathbb{C}[[z,q]]$

$$G(z,q) = \sum_{n=0}^{\infty} (1+q)^{\binom{n}{2}} \frac{z^n}{n!}. \tag{2}$$

Thus, the EGF of connected graphs is given by

$$C(z,q) = \log\left(G(z,q)\right) = z + q\frac{z^2}{2!} + (3q^2+q^3)\frac{z^3}{3!} + (16q^3+15q^4+6q^5+q^6)\frac{z^4}{4!} + \cdots, \tag{3}$$

which is valid in $\mathbb{C}[[z,q]]$. Denote by $C_{n,n+\ell}$ the number of connected labelled graphs built with n nodes and $n+\ell$ edges. Let W_ℓ be the exponential generating function (EGF) of connected labelled graphs with ℓ edges more than vertices. Therefore, $W_\ell(z) = \sum_n C_{n,n+\ell}\frac{z^n}{n!}$ and in $\mathbb{C}[[z,q]]$ we have

$$\begin{aligned} Q(z,q) &:= \sum_{n,\ell} C_{n,n+\ell}(-q)^{\ell+1}\frac{z^n}{n!} = -qC(-z/q,-q) \\ &= W_{-1}(z) - qW_0(z) + q^2W_1(z) + \cdots = -q\log\left(\sum_{n=0}^{\infty}(1-q)^{\binom{n}{2}}\frac{(-zq^{-1})^n}{n!}\right). \end{aligned} \tag{4}$$

Observe that the *negatively signed* variable q represents an essential trick in the authors approach. In fact, the series (in z) $G(z,q)$ diverges as soon as $q>0$ is fixed but as pinpointed by the authors, this function can acquire *bona fide* analytic sense. In fact, if q is fixed and satisfies $0<q<2$ (so that $|1-q|<1$) by considering a weighting π that assigns to a graph g the weight $\pi(g) := (-q)^{(\#\text{edges}(g) - \#\text{vertices}(g))}$, Flajolet *et al.* introduced two analytic objects

$$\mathcal{H}(z,q) := \sum_{g \text{ graph}} \pi(g)\frac{z^{|g|}}{|g|!}, \qquad \mathcal{Q}(z,q) := -q \sum_{g \text{ connected graph}} \pi(g)\frac{z^{|g|}}{|g|!}. \tag{5}$$

Another analytic ingredient derives from integral representations:

Lemma 1. *If $V(z) = \sum_n v_n z^n$ with $\sum_n |v_n| < \infty$ and if w is a real number s.t. $w \in (0,1)$ then*

$$\sum_n w^{n^2/2} v_n = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} V(e^{ix\sqrt{\log w^{-1}}}) e^{-x^2/2}\, dx. \tag{6}$$

This Lemma is proved in [3, Lemma 1] by means of classical Fourier integral combined with interchange of summation and integration in infinite sequences. The main interest of (6) is that such tricky integral representation *linearizes* the *quadratic forms* present in the *exponents* of the concerned EGFs. Using this, the authors of [3] obtained from (5) that:

Lemma 2. *The generating function of connected graphs admits for $q \in (0,1)$ the integral representation*

$$\mathcal{Q}(z,q) = -q\log\left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\left(-\frac{x^2}{2} - z\frac{(1-q)^{-1/2}}{q}e^{ix\lambda(q)}\right)dx\right), \tag{7}$$

where

$$\lambda \equiv \lambda(q) := \sqrt{\log\frac{1}{1-q}}.$$

Now, interchange of limits and coefficients operators is obtained with the help of a result due to Gessel and Wang (who gave relations between inversions in rooted trees and connected graphs) combined with [3, Lemma 3 – p. 9]. More specifically, we have the following Lemma:

Lemma 3. *Assume that for each z with $|z| < 1/e$ and as $q \to 0^+$ the bivariate generating function $\mathcal{Q}(z,q)$ satisfies an asymptotic expansions,*

$$\mathcal{Q}(z,q) \sim \sum_{\ell \geq 0} \mathcal{W}_\ell(z)(-q)^\ell, \qquad |z| < e^{-1}, \qquad q \to 0^+, \tag{8}$$

for a sequence of functions $\mathcal{W}_\ell(z)$. Then, for each ℓ, the equality $\mathcal{W}_\ell(z) = W_\ell(z)$ holds where W_ℓ is given by $W_\ell(z) := \sum_n C_{n,n+\ell} \frac{z^n}{n!}$.

As a consequence, this allows to identify Q and $\mathcal{Q}$.

3. Single Saddle-Point Analysis and Asymptotic Expansions

The integral representation (7) serves as starting point to estimate $Q(z,q)$ as $q \to 0^+$. Recall the following results:

Theorem 1. *Let $t := T(z) = z\exp(T(z))$ where $T(z)$ is the generating function of rooted trees. Then, (i) $W_{-1}(z) = t - \frac{t^2}{2}$, (ii) $W_0(z) = \frac{1}{2}\log\frac{1}{1-t} - \frac{t}{2} - \frac{t^2}{4}$ and (iii) for $k \geq 1$, there exist polynomials A_k such that $W_k(z) = \frac{A_k(t)}{(1-t)^{3k}}$.*

The objective is to prove Theorem 1 by analysis of the integral representation (7) when t is restricted in some fixed interval $(0,a)$ with $a < 1$. In fact, when $q \to 0^+$ by setting $x\lambda = w$ the integrand rewrites as

$$\exp\left(-\frac{1}{q}\left(\frac{w^2}{2} + ze^{iw}\right)\right) \times \exp\left(\frac{w^2}{2}(1/q - 1/\lambda^2) + ze^{iw}\frac{1-(1-q)^{-1/2}}{q}\right) \tag{9}$$

and we see that the integrand can vary abruptly due to the factor $1/q$ above. The saddle-points of the first factor are located at points τ such that $\tau + ize^{i\tau} = 0$. Therefore, as $|z| < 1/e$, $\tau = -it = -iT(z)$. The saddle-point method consists in shifting the line of integration parallel to itself so that it crosses the point τ. Setting $w = u - it$ and replacing the contour on $(-\infty, \infty)$, we get

$$Q(z,q) = \left(t - \frac{t^2}{2}\right) + \left(1 - \frac{q}{\lambda^2}\right) - \frac{q}{\lambda\sqrt{2\pi}}\int_{-\infty}^{+\infty} \exp\left(-\frac{\left(\frac{u^2}{2} + t\left(e^{iu} - 1 - iu\right)\right)}{q}\right) h(u)\, du \tag{10}$$

where $h(u) = \exp\left((u^2/2 - uit)(q^{-1} - \lambda^{-2}) + te^{iu}(1-(1-q)^{-1/2})/q\right)$. Next, the kernel of the saddle-point integral is reduced to standard *quadratic form*. This is done using change of variable defined by $y^2 = f(u)$ with $f(u) := u^2/2 + t(e^{iu} - 1 - iu)$. This leads to another expression of the integral above, viz.

$$I = \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{q}\left(\frac{u^2}{2} + t\left(e^{iu} - 1 - iu\right)\right)\right) h(u)\, du = \int_{-\infty}^{+\infty} e^{-y^2/q} H(y)\, dy, \quad H(y) := h(u(y))\frac{du}{dy}. \tag{11}$$

Now, the expansion of H as a power series in y can be integrated term by term with the Gaussian kernel $\exp(-y^2/q)$ (the validity of this term by term integration is proved in [3]). Expansions at any finite order with respect to q are legitimated and lead to finite order expansions (as $q \to 0^+$) as follows:

$$Q(z,q) \sim \left(T(z) - \frac{T(z)^2}{2}\right) - \left(\frac{1}{2}\log\frac{1}{1-T(z)} - \frac{T(z)}{2} - \frac{T^2(z)}{4}\right) q + \sum_{k\geq 2} \frac{A_{k-1}(T(z))}{(1-T(z))^{3k}}(-q)^k. \tag{12}$$

4. Coalescing Saddle-Points and Airy Functions

The aim of this section is to summarize the proof that the asymptotic number of connected graphs $C_{n,n+k}$ (for $k \geq 2$) can be expressed with $A_k(1)$ and the derivatives $A_k^{(j)}(1)$ ($j \geq 1$). Moreover, in their turn, the $A_k(1)$ (given also in [4]) can be expressed in terms of the Airy 'Ai' function [5]. Among other results, the authors of [3] obtained the following:

Theorem 2. *For any fixed value of k, the asymptotic number of connected graphs with n vertices and $n+k$ edges satisfied*

$$C_{n,n+k} = A_k(1)\sqrt{\pi}\left(\frac{n}{e}\right)^n \left(\frac{n}{2}\right)^{\frac{3k-1}{2}} \left(\frac{1}{\Gamma(3k/2)} + \frac{\frac{A_k'(1)}{A_k(1)} - k}{\Gamma(3k/2-1/2)}\sqrt{\frac{2}{n}} + O\left(\frac{1}{n}\right)\right). \tag{13}$$

The generating series of the dominant coefficients $A_k(1)$ has an asymptotic series

(14)

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

where $c_k = (6k)!/((3k)!\,(2k)!\,3^{2k}2^{5k})$.

Observe first that the *single saddle-point* analysis of the previous section leads to an expansion valid for $t = T(z)$ in any closed interval included in $[0,1)$ and such an expansion becomes meaningless as t approaches 1. Thus, appropriate tools are needed. In fact as t approaches 1, *two coalescing saddle points* arise. The main steps of the analysis consist in (i) localizing the dominant saddle points, (ii) normalizing the integrand by means of a *cubic* change of variables, (iii) integrating formally term by term, (iv) analyzing the remander of the obtained expansions to legitimate the formal result, (v) reorganizing correctly such expansions.

After rescaling (by the change of variable $\alpha = q/\theta^3$ with $\theta = 1 - t \equiv 1 - T(z)$), the starting point is now the following integral representation:

$$I := \int_{-\infty}^{+\infty} e^{-f(u)/q} h(u) du, \text{ with } f(u) = \frac{u^2}{2} + (1-\theta)(e^{iu} - 1 - iu). \tag{15}$$

Note that $f(u)$ is now locally cubic near $\theta = 0$. Solving $f'(u) = 0$ for $u \neq 0$ reveals this time a (purely imaginary) saddle point $\rho = -2i\theta(1 + 1/3\theta + \ldots)$. The classical method of *coalescent saddle-points* is then useful to asymptotically estimate the integral defined in (15). Namely, it is convenient to introduce the cubic change of variable $f(u) = P(v)$ with $P(v) = f(\rho)/\theta^3(2v^3 + 3\theta v^2)$ where the polynomial P is s.t. P' has two roots at 0 and $-\theta$ and $P(0) = 0$, $P(-\theta) = f(\rho)$. Thus, P and f behave similarly in the neighborhood of their two central saddle points. The integral now admits the exact expression

$$I = -\int_{e^{-i\pi/3}\infty}^{e^{i\pi/3}\infty} e^{-P(v)/q} G(v)\, dv, \qquad G(v) = h(u(v))\frac{du}{dv}. \tag{16}$$

Then, like in the previous section the next step consists in expanding G as a power series in v: $G(v,\alpha,\theta)=\sum_k g_k(\alpha,\theta)v^k$ and integrating term by term. This process leads to

$$I \sim \sum_{k=0}^{\infty} g_k(\alpha,\theta)\theta^{k+1} R_k\left(\frac{-\theta^3}{f(\rho)}\alpha\right), \tag{17}$$

The exponential generating series of the $R_k(x)$ is given by

$$R(z) := \sum R_k(x)\frac{z^k}{k!} = \int_{e^{-i\pi/3}\infty}^{e^{i\pi/3}\infty} e^{1/x(2v^3+3v^2)+zv}\,dv. \tag{18}$$

Comparing this with the classical integral representation of the Airy function leads to

$$R(z) = 2\pi i\left(\frac{x}{6}\right)^{1/3}\exp\left(-\frac{z}{2}+\frac{1}{2x}-\frac{1}{2x}\left(1-\frac{2}{3}zx\right)^{3/2}\right)e^{\frac{2}{3}y^{3/2}}\,\mathrm{Ai}(y), \tag{19}$$

where $y=\left(1-\frac{2}{3}zx\right)\left(\frac{3}{4x}\right)^{2/3}$. After some analysis, it is shown that

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

The proof of [3, Theorem 2] is then completed by collecting the powers of θ in (17).

This short note is a summary of Ph. Flajolet, B. Salvy, and G. Schaeffer's article [3].

Bibliography

[1] Bollobás (Béla). – *Random graphs.* – Academic Press, 1985.

[2] Erdős (Paul) and Rényi (Alfred). – On random graphs. *Publicationes Mathematicae Debrecen*, vol. 6, 1959, pp. 290–297.

[3] Flajolet (Philippe), Salvy (Bruno), and Schaeffer (Gilles). – Airy phenomena and analytic combinatorics of connected graphs. *The Electronic Journal of Combinatorics*, vol. 11, n° R34, 2004. – Downloadable at `http://www.combinatorics.org/Volume_11/PDF/v11i1r34.pdf`.

[4] Janson (Svante), Knuth (Donald E.), Łuczak (Tomasz), and Pittel (Boris). – The birth of the giant component. *Random Structures Algorithms*, vol. 4, n° 3, 1993, pp. 231–358. – With an introduction by the editors.

[5] Whittaker (E. R.) and Watson (G. N.). – *A Course of Modern Analysis.* – Cambridge University Press, 1927. Reprinted 1973 (fourth edition).

[6] Wilf (Herbert). – *Generatingfunctionology.* – Academic Press, 1990. Avalaible freely online at `http://www.math.upenn.edu/~wilf/gfology.pdf`.

[7] Wright (Edward Maitland). – The number of connected sparsely edged graphs. *Journal of Graph Theory*, vol. 1, 1977, pp. 317–330.

[8] Wright (Edward Maitland). – The number of connected sparsely edged graphs ii: Smooth graphs and blocks. *Journal of Graph Theory*, vol. 2, 1978, pp. 299–305.

[9] Wright (Edward Maitland). – The number of connected sparsely edged graphs iii: Asymptotic results. *Journal of Graph Theory*, vol. 4, 1980, pp. 393–407.