INFORMATIQUE THÉORIQUE ET APPLICATIONS

HELMUT PRODINGER

Comments on the analysis of parameters in a Random Graph Model

Informatique théorique et applications, tome 29, nº 3 (1995), p. 245-253

<http://www.numdam.org/item?id=ITA_1995__29_3_245_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 29, n° 3, 1995, p. 245 à 253)

COMMENTS ON THE ANALYSIS OF PARAMETERS IN A RANDOM GRAPH MODEL (*)

by Helmut PRODINGER ([1])

Communicated by J.-E. PIN

Abstract. – *Using generating functions and classical identities due to Euler and Gauss we can extend and simplify some of the results in the paper "Performance Considerations on a Random Graph Model for Parallel Processing", RAIRO Theoretical Informatics and Applications* **27** *(1993), 367-388 by Afrati and Stafylopatis.*

Résumé. – *Grâce aux fonctions génératrices et à des identités classiques dues à Euler et Gauss, il nous est possible d'étendre et de simplifier quelques uns des résultats de l'article « Performance Considerations on a Random Graph Model for Parallel Processing », RAIRO Theoretical Informatics and Applications* **27** *(1993), 367-388 par Afrati et Stafylopatis.*

1. INTRODUCTION

In the paper [1] a random (acyclic) graph model is considered. The nodes are $\{1, \ldots, n\}$; each new node k will have a directed edge to each of the earlier ones $\{1, \ldots, k-1\}$ with a fixed probability p and will be accordingly put into a certain level. The parameter of interest is the number of levels, and in particular its average value. This model being too complicated, the authors introduce two simplified models, serving as "lower bound" and "upper bound". We don't want to describe them here but rather give some comments about how to analyze them.

This note was written while the author was visiting the Algo project of INRIA, Rocquencourt and encountered an especially warm hospitality there. Interesting and valuable discussions with Philippe Flajolet are gratefully acknowledged.

(*) Received November 1993; revised July 1994; accepted October 1994.

([1]) Technical University of Vienna, Department of Algebra and Discrete Mathematics, Wiedner Hauptstrasse 8-10, A-1040 Vienna, Austria.

0988-3754/95/03/$ 4.00/© AFCET-Gauthier-Villars

In both instances the authors ask for explicit solutions of their recursions. We answer both questions in the affirmative. Also, we find surprising alternative representations of the appearing constants, due to theorems of Euler and Gauss, from the theory of partitions. Apart from the concrete results we feel that the presented analysis might be interesting in itself and useful for related questions.

For the necessary background we naturally refer to the paper [1] and only state the recursions that we are considering in the next two sections. We reverse the order and start we the easier "lower bound".

To simplify the authors' notation, we use $q := 1 - p$, where p is a probability, throughout this note. It is amazing that this "q", denoting a probability and the formal variable "q", used in so-called q-series [2] fit together so well!

Let us sketch the methods. The recursions will be solved by setting up *ordinary generating functions*. This gives immediately an expression for the generating function in the easy case, whereas in the difficult case the generating function must be extracted by iterations. In both instances we arrive at something like $1/(1 - z)^2$ multiplied by a function which is analytic in a larger area than $|z| < 1$. Thus the asymptotic behaviour of the coefficients of interest is given by n times a constant, which is just the value of the extra factor at $z = 1$. It is then possible to rewrite these constants as infinite products instead of infinite sums. The advantage of these representations is that their behaviour for $q \to 1$ is quite easy to obtain, as opposed to the sum representations.

For all these mathematical methods we refer without further comments to the brilliant survey [4].

Finally we want to give a flavour of the quantitative results that are obtained in the next two sections. Typically, we might expect $\alpha \cdot n$ levels for large n, where the constant α is depending on the probability p and of the model, α_1 referring to the lower bound model and α_2 referring to the upper bound model. For instance, in the symmetric case $p = q = \frac{1}{2}$ we find $\alpha_1 = 0.56546\ldots$ and $\alpha_2 = 0.60914\ldots$.

The behaviour of these constants for $p \to 0$ (or $q \to 1$) is of special interest, as it describes the behaviour of what is known as *sparse graphs*. It turns out that $\alpha_1 \sim \dfrac{p}{e-2}$, whereas $\alpha_2 \sim \sqrt{\dfrac{2p}{\pi}}$. More precise information is available, sharpening the results of [1]. So the two models show a different behaviour; one constant depending linearly, the other being like a square root of the (small) parameter p.

2. THE "UPPER BOUND"

The recursion of the upper bound model is (equation (13) of [1])

$$P(n, 1) = \sum_{k=1}^{n-1} P(n-k, 1)\, q^{\binom{k}{2}} (1 - q^k) \quad n \geq 2; \qquad P(1, 1) = 1. \quad (1)$$

With these quantities one is interested in the "mean length"

$$\overline{L_2(n)} := \sum_{k=1}^{n} P(k, 1). \qquad (2)$$

By additionally defining $P(0, 1) = 0$ we can extend the range of the summation in (1) from 1 to n.

Since the recursion has the flavour of a *convolution*, it is extremely natural to use *generating functions*: Set

$$f(z) := \sum_{n \geq 0} P(n, 1)\, z^n, \qquad (3)$$

then (1) translates into

$$f(z) - z = f(z) \cdot \sum_{k \geq 0} q^{\binom{k}{2}} (1 - q^k)\, z^k. \qquad (4)$$

Set

$$\psi(z) :\doteq \sum_{k \geq 0} q^{\binom{k+1}{2}} z^k, \qquad (5)$$

then we find

$$\sum_{k \geq 0} q^{\binom{k}{2}} (1 - q^k)\, z^k = 1 - (1 - z)\, \psi(z) \qquad (6)$$

and thus

$$f(z) = \frac{z}{(1 - z)\, \psi(z)}. \qquad (7)$$

We find the sought quantity $P(n, 1)$ as the coefficient of z^n in this (explicit!) function $f(z)$, which we denote, as usual, by $[z^n] f(z)$. Furthermore,

$$\overline{L_2(n)} = [z^n] \frac{1}{1-z} \cdot f(z) = [z^n] \frac{z}{(1-z)^2} \cdot \frac{1}{\psi(z)}. \tag{8}$$

Since $\psi(z)$ is analytic at $z = 1$, we see from (8) by *singularity analysis* [3] that

$$\overline{L_2(n)} = \frac{n}{\psi(1)} + \frac{\psi'(1)}{\psi(1)} + \text{exponentially small terms.} \tag{9}$$

The quantity $\psi(1)$ can be evaluated by a formula of Gauss [2, (2.2.13)]:

$$\psi(1) = \sum_{n \geq 0} q^{\binom{n+1}{2}} = \prod_{m \geq 1} \frac{1-q^{2m}}{1-q^{2m-1}}. \tag{10}$$

In [1] the behaviour of

$$\frac{\alpha_2}{p} := \frac{1}{(1-q)\,\psi(1)} = \prod_{m \geq 1} \frac{1-q^{2m+1}}{1-q^{2m}} \tag{11}$$

is considered for $q \to 1$. With the product representation this is very simple. We substitute $q = e^{-x}$, take the logarithm and call the resulting function $g(x)$:

$$g(x) = \sum_{m \geq 1} \log\left(1 - e^{-(2m+1)x}\right) - \sum_{m \geq 1} \log\left(1 - e^{-2mx}\right). \tag{12}$$

We compute its *Mellin transform,*

$$g^*(s) = -\Gamma(s)\,\zeta(s+1)\left(-1 + \zeta(s)\left(1 - 2^{1-s}\right)\right), \tag{13}$$

and find, by virtue of the *inversion formula,*

$$g(x) = \frac{1}{2\pi i} \int_{2-i\infty}^{2+i\infty} g^*(s)\, x^{-s}\, ds. \tag{14}$$

Now we shift the integral to the left and collect the residues of the integrand to get the desired asymptotic expansion of $g(x)$ for $x \to 0$. Then we can go back to $\exp(g(x))$ and replace the variable x by $-\log q$. All this can be done by MAPLE:

$$
\begin{aligned}
&\frac{1}{(1-q)\,\psi(1)}\\
&\quad=\sqrt{\frac{2}{-\pi\,\log q}}\cdot\left(1-\frac{3}{8}\log q+\frac{11}{384}\log^2 q+O\left(\log^3 q\right)\right)\\
&\quad=\sqrt{\frac{2}{\pi}}\,\frac{1}{\sqrt{1-q}}\left(1-\frac{1}{8}(1-q)+\frac{19}{284}(1-q)^2+O\left((1-q)^3\right)\right) \quad (15)
\end{aligned}
$$

This is of course a quantitative refinement of the statement that $\frac{\alpha_2}{1-q}$ goes to infinity.

THEOREM U: [Upper bound] *The mean length* $\overline{L_2(n)}$ *is the coefficient of* z^n *in*

$$
\frac{z}{(1-z)^2}\cdot\Big(\sum_{k\geq 0} q^{\binom{k+1}{2}}\,z^k\Big)^{-1};
$$

it is asymptotically equivalent to

$$
\overline{L_2(n)}\sim n\cdot\prod_{m\geq 1}\frac{1-q^{2m-1}}{1-q^{2m}}.
$$

The behaviour of the "inverse of the efficiency" for $q\to 1$ *is given by*

$$
\begin{aligned}
&\prod_{m\geq 1}\frac{1-q^{2m+1}}{1-q^{2m}}\\
&\quad\sim\sqrt{\frac{2}{\pi}}\,\frac{1}{\sqrt{1-q}}\left(1-\frac{1}{8}(1-q)+\frac{19}{284}(1-q)^2+O\left((1-q)^2\right)\right).
\end{aligned}
$$

3. THE "LOWER BOUND"

This time the recursion for the lower bound model looks like

$$
P(n,1)=\sum_{k=1}^{n-1}P(n-1,k)\,(1-q^k)+q^2\,P(n-1,1)
$$

$$
\text{for}\quad n\geq 3, \tag{16}
$$

$$P(n, j) = P(n-1, j-1)\, q^{j-1} p + P(n-1, j)\, q^{j+1} \qquad \text{for} \quad 2 \leq j \leq n-2, \tag{17}$$

with $P(n, n-1) = q^{\binom{n-1}{2}} p^{n-1}$ and $P(n, n) = q^{\binom{n}{2}}$ for $n \geq 1$. Note that (16) can be replaced by the condition $\sum_{j=1}^{n} P(n, j) = 1$ for all n.

We introduce for each j a generating function $h_j(z)$ by

$$h_j(z) := \sum_{n \geq j} P(n, j)\, z^n. \tag{18}$$

By a routine computation (16) and (17) are translated into

$$h_1(z) = \frac{1}{1 - pz - q^2 z} \Big(z - q^2 z^2 + z \sum_{k \geq 2} h_k(z)\,(1 - q^k)\Big) \tag{19}$$

and

$$h_j(z) = pq^{j-1} z h_{j-1}(z) + zq^{j+1} h_j(z) + q^{1+\binom{j}{2}} z^j (1 - zq^j) \qquad \text{for} \quad j \geq 2. \tag{20}$$

We are interested in the function $h_1(z)$ since it "contains" the interesting coefficients. Let us recall that the desired "mean length"

$$\overline{L_1(n)} = q^2 + \sum_{k=1}^{n} P(k, 1) - q^2 \sum_{k=1}^{n-1} P(k, 1) \tag{21}$$

can be obtained as

$$\overline{L_1(n)} = q^2 + [z^n]\, \frac{1}{1-z}\, h_1(z) - q^2\, [z^{n-1}]\, \frac{1}{1-z}\, h_1(z). \tag{22}$$

From (20) we see

$$h_j(z) = a_j\, h_{j-1}(z) + b_j,$$

$$\text{with} \quad a_j = \frac{pq^{j-1}}{1 - zq^{j+1}} \quad \text{and} \quad b_j = \frac{q^{1+\binom{j}{2}} z^j (1 - zq^j)}{1 - zq^{j+1}}. \tag{23}$$

We can iterate this and thus express each $h_j(z)$ by $h_1(z)$:

$$h_j(z) = A_j \sum_{k=2}^{j} \frac{b_k}{A_k} + A_j h_1(z), \qquad \text{where} \quad A_k = a_2 a_3 \ldots a_k. \tag{24}$$

We can now insert this into (19) and solve the linear equation for $h_1(z)$. The obtained solution is admittedly not very nice, but it is explicit. Define $d_j = A_j \sum_{k=2}^{j} b_k/A_k$, then we find in this way

$$h_1(z) = \frac{z - q^2 z^2 + z \sum_{k \geq 2} (1 - q^k) d_k}{1 - pz - q^2 z - z \sum_{k \geq 2} (1 - q^k) A_k}. \tag{25}$$

Let us now engage on asymptotics. In [1] it was implicitly proved that

$$h_j(z) \sim \frac{\pi(j)}{1 - z} \quad \text{as } z \to 1, \tag{26}$$

where

$$\left.\begin{aligned} \pi(j) &= \frac{p^{j-1} q^{\binom{j}{2}}}{(1 - q^3) \ldots (1 - q^{j+1})} \\ \text{and} \quad & \\ \frac{1}{\pi(1)} &= \sum_{j \geq 1} \frac{p^{j-1} q^{\binom{j}{2}}}{(1 - q^3) \ldots (1 - q^{j+1})}. \end{aligned}\right\} \tag{27}$$

We could refine this by setting

$$h_j(z) \sim \frac{\pi(j)}{1 - z} + \sigma(j), \tag{28}$$

insert it into (20), compare coefficients and thus express the numbers $\sigma(j)$ by the $\pi(j)$'s. We omit this since the formulae are not too nice.

But we can conclude, again by *singularity analysis* that

$$P(n, 1) = \pi(1) + \text{exponentially small terms} \tag{29}$$

and

$$\begin{aligned} \overline{L_1(n)} = & n \cdot (1 - q^2) \pi(1) + q^2 + \pi(1) \\ & + (1 - q^2) \sigma(1) + \text{exp. small terms}. \end{aligned} \tag{30}$$

Let us now analyse the constant $(1-q^2)\,\pi(1)$ which is also called α_1 in [1]. By some simple computations we find that

$$\frac{1}{\alpha_1}=\frac{1}{p}\sum_{j\geq 0}\frac{p^j\,q^{\binom{j}{2}}}{(1-q)\dots(1-q^j)}-1-\frac{2\,q}{p}. \tag{31}$$

This time we can use a formula of Euler [2, (2.2.26)]:

$$\sum_{j\geq 0}\frac{p^j\,q^{\binom{j}{2}}}{(1-q)\dots(1-q^j)}=\prod_{k\geq 0}(1+pq^k). \tag{32}$$

From this representation the behaviour of α_1/p (needed in [1]) for $q\to 1$ is very easy to obtain. We consider its reciprocal, which is

$$\prod_{k\geq 0}(1+(1-q)\,q^k)-1-q, \tag{33}$$

forget about the extra $-1-q$, take the logarithm, expand it as a Taylor series, interchange the order of summation and expand. With the help of MAPLE we get

$$\frac{\alpha_1}{p}=\frac{1}{e-2}-\frac{\frac{e}{4}-1}{(e-2)^2}(q-1)+O\left((q-1)^2\right), \tag{34}$$

but we could easily get as many terms as we please.

THEOREM L: [Lower bound] *The mean length $\overline{L_1(n)}$ is given by*

$$\overline{L_1(n)}=q^2+[z^n]\,\frac{1}{1-z}\,h_1(z)-q^2\,[z^{n-1}]\,\frac{1}{1-z}\,h_1(z),$$

where the function $h_1(z)$ is given by

$$h_1(z)=\frac{z-q^2\,z^2+z\sum_{k\geq 2}(1-q^k)\,d_k}{1-pz-q^2\,z-z\sum_{k\geq 2}(1-q^k)\,A_k}.$$

Here,

$$A_k=p^{k-1}\,q^{\binom{k}{2}}\Big/\prod_{j=3}^{k+1}(1-zq^j)$$

and

$$d_k = A_k \sum_{j=2}^{k} \frac{1}{A_j} \frac{q^{1+\binom{j}{2}} z^j (1 - zq^j)}{1 - zq^{j+1}}.$$

It is asymptotically equivalent to

$$\overline{L_1(n)} \sim n \cdot \left(\frac{1}{p} \prod_{k \geq 0} (1 + pq^k) - 1 - \frac{2q}{p} \right)^{-1}.$$

The behaviour of the "inverse of the efficiency" for $q \to 1$ is given by

$$\frac{1}{p} \left(\frac{1}{p} \prod_{k \geq 0} (1 + pq^k) - 1 - \frac{2q}{p} \right)^{-1}$$

$$\sim \frac{1}{e-2} - \frac{\frac{e}{4} - 1}{(e-2)^2} (q-1) + O((q-1)^2).$$

REFERENCES

1. F. Afrati and A. Stafylopatis, Performance Considerations on a Random Graph Model for Parallel Processing, *RAIRO Theoretical Informatics and Applications*, 1993, *27*, pp. 367-388.
2. G. E. Andrews, *The Theory of Partitions*, Addison-Wesley, Reading, MA, 1976.
3. P. Flajolet and A. Odlyzko, Singularity Analysis of Generating Functions, *SIAM J. Discrete Mathematics*, 1990, *3*, pp. 216-240.
4. P. Flajolet and J. Vitter, Analysis of Algorithms and Data Structures, *Handbook of Theoretical Computer Science, Vol. A: Algorithms and Complexity*. (J. van Leeuwen, ed.), 1990, pp. 431-524.