Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER Journal of Combinatorial Theory, Series A 107 (2004) 117–125
Journal of Combinatorial Theory Series A
http://www.elsevier.com/locate/jcta

Asymptotics of combinatorial structures with large smallest component

Edward A. Bender,[a] Atefeh Mashatan,[b] Daniel Panario,[c] and L. Bruce Richmond[b]

[a] Department of Mathematics, University of California, San Diego, La Jolla 92093, USA
[b] Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Canada N2L 3G1
[c] School of Mathematics and Statistics, Carleton University, Ottawa, Canada K1S 5B6

Received 5 May 2003

Abstract

We study the probability of connectedness for structures of size n when all components must have size at least m. In the border between almost certain connectedness and almost certain disconnectedness, we encounter a generalized Buchstab function of n/m.

Keywords: Combinatorial enumeration; Components; Connectedness

1. Results

We consider a class of decomposable combinatorial objects $\mathscr{A}$ and require that each object attains a unique decomposition over a sub-class of the original class $\mathscr{C}$, called irreducible or connected components. We examine structures such as general graphs or graphs with certain properties and their components, monic polynomials over finite fields viewed as products of irreducible cycles, permutations viewed as sets of cycles, to name a few.

There is a natural notion of size related to these combinatorial objects and their components. We let A_n be the number of structures of size n and C_n the number of those that are connected. We let $A(x)$ denote the generating function for the objects

E-mail addresses: ebender@ucsd.edu (E.A. Bender), amashatan@uwaterloo.ca (A. Mashatan), daniel@math.carleton.ca (D. Panario), lbrichmond@uwaterloo.ca (L.B. Richmond).

0097-3165/$ - see front matter
doi:10.1016/j.jcta.2004.04.001

and $C(x)$ denote the generating function for the connected components. Some combinatorial objects are labelled and some are unlabelled. In the case of labelled structures without restrictions we obtain $\mathscr{A}$ with performing the *set* operator on $\mathscr{C}$ and hence the exponential generating functions are related by (for example, see [5,10])

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

For unlabelled structures, $\mathscr{A}$ is obtained by performing the *multiset* operator on $\mathscr{C}$ and the ordinary generating functions are related by

$$A(x) = \exp\left(\sum_{k\geqslant 1} C(x^k)/k\right).$$

The prime example for labelled combinatorial structures are cycles and permutations. We have $n!$ different permutation on n element and hence the exponential generating function for permutations is

$$\sum_{n\geqslant 0} n!\frac{x^n}{n!} = \sum_{n\geqslant 0} x^n = \frac{1}{1-x}.$$

The number of cycles of size n is $(n-1)!$ and the exponential generating function for the cycles is

$$\sum_{n\geqslant 0} (n-1)!\frac{x^n}{n!} = \sum_{n\geqslant 0}\frac{x^n}{n} = \log\frac{1}{1-x}.$$

Hence we check that

$$\exp(C(x)) = \exp\left(\log\frac{1}{1-x}\right) = \frac{1}{1-z} = A(x).$$

The prime example of unlabelled combinatorial structures are monic polynomials over a finite field $\mathbb{F}_q$ with the generating function

$$\sum_{n\geqslant 0} q^n x^n = \frac{1}{1-qx},$$

where q is a power of a prime integer.

We let $A_{n,m}$ be the number of structures of size n whose smallest component has size at least m. For a general discussion over both labelled and unlabelled structures let

$$a_n = \begin{cases} \frac{A_n}{n!} & \text{if } \mathscr{A} \text{ is labelled,}\\ A_n & \text{if } \mathscr{A} \text{ is unlabelled,}\end{cases}$$

$$c_n = \begin{cases} \frac{C_n}{n!} & \text{if } \mathscr{C} \text{ is labelled,}\\ C_n & \text{if } \mathscr{C} \text{ is unlabelled,}\end{cases}$$

$$a_{n,m} = \begin{cases} \frac{A_{n,m}}{n!} & \text{if } \mathscr{A} \text{ is labelled,}\\ A_{n,m} & \text{if } \mathscr{A} \text{ is unlabelled.}\end{cases}$$

The probability of connectedness of objects of size n, $C_n/A_n = c_n/a_n$, was studied in [2]. In particular, it was shown that if $\rho = \lim_{n\to\infty} C_n/A_n$ exists then divergence of $C(R)$ implies $\rho = 0$ while convergence implies $0<\rho<1$ and any value in that range is possible, where $0<R<\infty$ is the radius of convergence of $C(x)$. Here we are interested in the probability that an object of size n whose smallest component has size at least m, is connected. Hence we study $C_n/A_{n,m} = c_n/a_{n,m}$ and let m tend to infinity along with n.

Buchstab [3] defined the following function, $\omega(u)$, for $u \geqslant 1$

$$\begin{cases} \omega(u) = u^{-1} & \text{if } 1 \leqslant u \leqslant 2; \\ \frac{d}{du}(u\omega(u)) = \omega(u-1) & \text{if } u \geqslant 2. \end{cases}$$

Here we need a generalization of this function. For each $K>0$ we define a generalized Buchstab function on $[1,\infty)$ by

$$\Omega_K(x) = \begin{cases} 1 & \text{if } 1 \leqslant x < 2, \\ 1 + K\int_2^x \frac{\Omega_K(t-1)}{t-1}\,dt & \text{if } x \geqslant 2. \end{cases}$$

We note that the standard Buchstab function is $\Omega_1(x)/x$.

Theorem 1.1. *Fix $\varepsilon>0$ sufficiently small.*

(a) *If $c_n \sim f(n)/nR^n$ where $f(n) = o(f^2(\alpha n))$ uniformly for $\varepsilon \leqslant \alpha \leqslant 1-\varepsilon$, then*

$$\lim_{n\to\infty} c_n/a_{n,m} = \begin{cases} 1 & \text{if } 1/2 < m/n \leqslant 1, \\ 0 & \text{if } \varepsilon \leqslant m/n \leqslant 1/2 - \varepsilon, \end{cases}$$

uniformly for $\varepsilon \leqslant m/n \leqslant 1$.

(b) *If $c_n \sim f(n)/nR^n$ where $f^2(\alpha n) = o(f(n))$ uniformly for $\varepsilon \leqslant \alpha \leqslant 1-\varepsilon$, then $c_n/a_{n,m} \sim 1$ uniformly for $\varepsilon \leqslant m/n \leqslant 1$.*

(c) *If $c_n \sim K/nR^n$, then $c_n/a_{n,m} \sim 1/\Omega_K(n/m)$ uniformly for $\varepsilon \leqslant m/n \leqslant 1$.*

Our proof is a modification of Buchstab's treatment for the smallest prime factor of the first n integers; however, he has $\log n$ and $\log m$ while we have n and m. We adapt Tenenbaum's presentation [11], however the fact that there can be many components of a given size (rather than a single prime) leads to significant modifications of his argument. Using other methods, Panario and Richmond [10] obtained (c); however their formula for Ω_K is more complicated except in the case $K = 1$ which they related to the Buchstab function.

2. Examples

The next three examples are from [5].

• *Permutations*: As mentioned previously the exponential generating function for cycles in permutations is $C(x) = \log\frac{1}{1-x}$. Therefore $C_n = (n-1)!$ and hence

$c_n = (n-1)!/n! = 1/n$. On the other hand the radius of convergence of the function $\log\frac{1}{1-x}$ is 1. Consequently part (c) of the theorem applies with $K = 1$. This gives rise to Ω_1.

• *Polynomials*: As mentioned before, the ordinary generating function for monic polynomials over a finite field $\mathbb{F}_q$ is $\frac{1}{1-qx}$ and we have the well-known approximation for C_n, the number of irreducible polynomials of degree n,

$$C_n = \frac{q^n}{n} + \mathcal{O}(q^{n/2}).$$

From this approximation we can find R, the radius of convergence of $C(x)$, as follows:

$$R = \limsup_{n\to\infty} |C_n|^{-1/n} = \limsup_{n\to\infty} \left|\frac{q^n}{n} + \mathcal{O}(q^{n/2})\right|^{-1/n} = q^{-1}.$$

From our definition $c_n = C_n = q^n/n + \mathcal{O}(q^{n/2})$ in the case of unlabelled objects. We now compute the term $1/(nR^n)$:

$$\frac{1}{nR^n} = \frac{(R^{-1})^n}{n} = \frac{q^n}{n}.$$

Hence $K = 1$ and part (c) of the theorem applies.

• *2-regular graphs*: The exponential generating function for labelled 2-regular graphs is

$$\frac{e^{-x/2-x^2/4}}{(1-x)^{1/2}} = \exp\left(\frac{1}{2}\log\frac{1}{1-x} - \frac{x}{2} - \frac{x^2}{4}\right).$$

Hence the exponential generating function for the components is

$$\left(\frac{1}{2}\log\frac{1}{1-x} - \frac{x}{2} - \frac{x^2}{4}\right).$$

Extracting coefficients yields in $C_1 = C_2 = 0$ and

$$C_n = \frac{(n-1)!}{2}.$$

Hence $c_n = \frac{1/2}{n}$ and this gives rise to $\Omega_{1/2}$.

We now show some examples that parts (a) and (b) apply. Suppose

$$c_n \sim \frac{K}{n^s R^n} = n^{1-s}\frac{K}{nR^n}.$$

For $s<1$ we have $1-s>0$, a positive exponent of n in $f(n) = Kn^{1-s}$. Therefore we get $Kn^{1-s} = o(K^2(\alpha n)^{2(1-s)})$ and hence part (a) of the theorem applies. On the other hand if $s>1$ the exponent of n is negative and we get $K^2(\alpha n)^{2(1-s)} = o(Kn^{1-s})$ and part (b) applies.

In the following we give some examples of graphs, which we describe by their components. Trees give $s = 5/2$ in the unrooted case and $s = 3/2$ in the rooted case. This holds for both labelled and unlabelled graphs [4,7].

• *Rooted and unrooted trees*: Let $A(x)$ and $C(x)$ be the exponential generating functions for rooted trees. We know that $A(x) = \exp(C(x))$ and $C(x) = xA(x)$. It follows that $C_n = n^{n-1}$ and hence $c_n = n^{n-1}/n!$. On the other hand we have that

$$R = \lim_{n\to\infty} \frac{nC_{n-1}}{C_n} = \lim_{n\to\infty} \frac{n(n-1)^{n-2}}{n^{n-1}} = \lim_{n\to\infty} \left(\frac{n-1}{n}\right)^{n-2} = \lim_{n\to\infty} \left(1 + \frac{1}{n-1}\right)^{n-1} = \frac{1}{e}.$$

Using Stirling formula, that is $n! \sim \sqrt{2\pi}(n/e)^n \sqrt{n}$, we get

$$c_n = \frac{n^{n-1}}{n!} \sim \frac{e^n}{\sqrt{2\pi}n\sqrt{n}} = \frac{1}{\sqrt{2\pi}n^{3/2}(1/e)^n} = \frac{1}{\sqrt{2\pi}n^{3/2}R^n},$$

which gives rise to $s = 3/2$. This result can be found in [4].

Let $C'(x)$ be the ordinary generating function of unrooted trees. Otter [8] showed that $c_n{}' = C_n{}' \sim Kn^{-5/2}(R')^{-n}$, where R' is the radius of convergence of $C'(x)$. He also obtained $s = 3/2$ for unlabelled rooted trees. Asymptotics of unlabelled unrooted forests was studied by Palmer and Schwenk [9].

• *Achiral trees*: A plane graph is one that can be drawn in the plane with no pair of edges crossing. An achiral graph is a plane graph with plane symmetry. In other words it is its own mirror image. The concept of chirality was motivated by organic chemists. Achiral plane trees have been studied by Wormald in [12]. Generating functions of plane, rooted or achiral trees and any combinations of those are given by Harary and Robinson [6]. The generating function of achiral plane trees $A(x)$, as it is proven in [6], is

$$A(x) = \frac{x}{2}\left(-1 + (1+2x)(1-4x^2)^{\frac{1}{2}}\right).$$

The radius of convergence of this function is $1/2$ and extracting coefficients results in

$$A_{2n} = \binom{2n-2}{n-1} \quad \text{and} \quad A_{2n+1} = \frac{1}{2}\binom{2n}{n}.$$

Using Stirling's approximation we get

$$A_{2n+1} \sim \frac{2^{2n-1}}{\sqrt{\pi n}},$$
$$A_{2n} \sim \frac{2^{2n-2}}{\sqrt{\pi n}}$$

which gives us $s = 1/2$.

• *k-neighbour tree*: A k-neighbour tree is a labelled tree consisting of a vertex of degree $k-1$ with $k-1$ neighbours of degree one with no other vertices or edges. These trees with $k > 2$ give $s = 1 - k$ since $C_n = n!\binom{n-2}{k-1}/k!$. To see this, write a permutation of $\{1, \ldots, n\}$ in one line form, choose the first element for the center, choose $k-1$ places to divide the remainder into neighbours of the center, and ignore the order of the neighbours.

3. Proofs

We begin with a simple lemma.

Lemma 3.1. *If $a<b$ are integers,*

$$\left| \int_a^b f(x)\,dx - \sum_{x=a+1}^{b} f(x) \right| \leqslant V_f(a,b),$$

where $V_f(a,b)$ is the total variation of f on $[a,b]$.

Proof. It suffices to consider the interval $[k, k+1]$ and sum. The difference between the largest and smallest value of f on this interval is at most $V_f(k, k+1)$. Since $f(k+1)$ and $\int_k^{k+1} f(x)\,dx$ both lie between the max and min, we are done. $\square$

The rest of this section is devoted to a proof of the theorem. It will be useful to bound $\sum_{i=m}^{n/2} \frac{n}{i(n-i)}$ for $\varepsilon n \leqslant m \leqslant n/2$. The function $\frac{1}{x(n-x)}$ has negative derivative for $0 \leqslant x < n/2$ and hence at $i = n/2$ we have the minimum value and at $i = m$ the maximum value of the terms in the sum $\sum_{i=m}^{n/2} \frac{n}{i(n-i)}$. Therefore the terms lie between $4/n$ (at $i = n/2$) and $\frac{2}{n\varepsilon}$ (at $i = m$). Since the number of terms is $n/2 - m + O(1)$, we have

$$\sum_{i=m}^{n/2} \frac{n}{i(n-i)} \leqslant \left(\frac{2}{n\varepsilon}\right)(n/2 - m + O(1)) \leqslant \frac{1}{\varepsilon} - \frac{2}{\varepsilon}\left(\frac{m}{n}\right) + O\left(\frac{1}{n\varepsilon}\right).$$

This shows that the sum is bounded and is at least $(4/n)(n/2 - m + O(1))$ which is bounded away from 0 if m/n is bounded away from $1/2$.

Since there cannot be two components of size exceeding $n/2$, $a_{n,m} = c_n$ when $m > n/2$.

To prove part (a), we let $\alpha = \beta$ and $\alpha = \gamma$ to obtain

$$f(n) = o(f^2(\beta n)) \quad \text{and} \quad f(n) = o(f^2(\gamma n))$$

uniformly for $\varepsilon \leqslant \beta \leqslant \gamma \leqslant 1 - \varepsilon$. Multiplying the results, and taking the square root we obtain

$$f(n) = o(f(\beta n) f(\gamma n))$$

uniformly for $\varepsilon \leqslant \beta \leqslant \gamma \leqslant 1-\varepsilon$. Considering only two components, one of size i with $m \leqslant i < n/2$, we have

$$\begin{aligned}\frac{a_{n,m}}{c_n} &\geqslant \sum_{m\leqslant i<n/2} \frac{c_i c_{n-i}}{c_n} \sim \sum_{m\leqslant i<n/2} \frac{f(i)\,f(n-i)n}{i(n-i)\,f(n)} \\ &= \sum_{m\leqslant i<n/2} \frac{n}{i(n-i)o(1)},\end{aligned}$$

which goes to infinity with n since $\sum \frac{n}{i(n-i)}$ is bounded away from zero due to $m/n \leqslant 1/2-\varepsilon$.

For (b) and (c), we use induction on k where $k \leqslant n/m < k+1$ after dealing with the case $m=n/2$. In this case $a_{n,m} \leqslant c_n + c_m^2$ and $c_m^2 = o(c_n)$. Thus $a_{n,m} \sim c_n$.

Let i be the size of the smallest component. If we insist that there be only one component of size i, we obtain the lower bound

$$\frac{a_{n,m}}{c_n} \geqslant 1 + \sum_{i=m}^{n/2} \frac{c_i a_{n-i,i+1}}{c_n}. \tag{3.1}$$

On the other hand, if we mark a component of size i and allow other components of size i, we obtain the upper bound

$$\frac{a_{n,m}}{c_n} \leqslant 1 + \sum_{i=m}^{n/2} \frac{c_i a_{n-i,i}}{c_n}. \tag{3.2}$$

Since $1 \leqslant \frac{n-i}{i+1} < \frac{n-i}{i} \leqslant \frac{n}{m} - 1 < k$, we can induct.

To prove (b) we use the induction hypothesis and the condition of f to rewrite (3.2) as

$$\frac{a_{n,m}}{c_n} \leqslant 1 + \sum_{i=m}^{n/2} \frac{n\,o(1)}{i(n-i)},$$

which equals $1+o(1)$ since $\sum \frac{n}{i(n-i)}$ is bounded. Similarly, (3.1) yields that $a_{n,m}/c_n| \geqslant 1 + o(1)$.

We now turn our attention to (c). By the induction hypothesis and assumptions about c_n,

$$\frac{c_i a_{n-i,i+1}}{c_n} = \frac{c_i c_{n-i}}{c_n} \frac{a_{n-i,i+1}}{c_{n-i}} \sim \frac{n}{i(n-i)}\,\Omega_K\left(\frac{n-i}{i+1}\right). \tag{3.3}$$

Note that, for any $C>1$, $\Omega_K(x) \geqslant 1$ and is bounded and uniformly continuous on $[1,C]$. Hence $V_{\Omega_K}(1,C)$ exists. It is known that

$$V_{gh}(a,b) \leqslant M(h)V_g(a,b) + M(g)V_h(a,b), \tag{3.4}$$

where $M(f) = \sup\{|f(x)| \mid x \in [a,b]\}$. (For example, see Section 8.4 of [1].) Apply the lemma with $f = gh$, $g = \frac{n}{i(n-i)}$ and $h = \Omega_k\left(\frac{n-i}{i+1}\right)$, using (3.1), (3.3) and (3.4) to obtain

$$\frac{a_{n,m}}{c_n} \geqslant 1 + \int_m^{n/2} \frac{Kn}{x(n-x)} \Omega_K\left(\frac{n-x}{x+1}\right) dx + o(1) \sum_{i=m}^{n/2} \frac{n}{i(n-i)} + O\left(\frac{1}{n}\right),$$

where

- the $o(1)$ comes from the induction hypothesis and the uniformity of the approximation and
- the $O(1/n)$ comes from (3.4) and the fact that the magnitude and total variation of g are $O(1/n)$ while those of h are bounded.

If we replace (3.1) by (3.2) and use (3.3) and (3.4), then we get an upper bound for $a_{n,m}/c_n$ which is the same as the lower bound we got. This gives an asymptotic formula for $a_{n,m}/c_n$.

We observe that

$$\Omega_K\,(x+\delta) - \Omega_K(x) = K \int_x^{x+\delta} \frac{\Omega_K(t-1)}{t-1}\, dt.$$

It follows that $\Omega_K(x)$ is continuous for each $x \geqslant 1$ and differentiable except at $x = 2$ (the left- and right-hand derivatives are 0 and K, respectively). However for $x > 1$ we have

$$|\Omega_K(x+\delta) - \Omega_K(x)| \leqslant M\delta,$$

where M is a constant and depends only on ε. Thus,

$$\Omega_K\left(\frac{n-x}{x+1}\right) = \Omega_K\left(\frac{n-x}{x}\right) + O\left(\frac{1}{n}\right).$$

Using the asymptotic formula for $a_{n,m}/c_n$ from above, we now have

$$\frac{a_{n,m}}{c_n} = 1 + K \int_m^{n/2} \frac{n}{x(n-x)}\, \Omega_K\left(\frac{n-x}{x}\right) dx + o(1).$$

Substituting $t = n/x$, we obtain

$$\begin{aligned}\frac{a_{n,m}}{c_n} &\sim 1 + K \int_{n/m}^{2} \frac{t^2}{n(t-1)}\, \Omega_K(t-1)\,(-n/t^2\, dt)\\ &= 1 + K \int_2^{n/m} \frac{1}{(t-1)}\, \Omega_K(t-1)\, dt = \Omega_K\left(\frac{n}{m}\right).\end{aligned}$$

This completes the proof. □

References

[1] T.M. Apostol, Mathematical Analysis, Addison-Wesley, Reading, MA, 1957.
[2] J.P. Bell, E.A. Bender, P.J. Cameron, L.B. Richmond, Asymptotics for the probability of connectedness and the distribution of number of components, Electron. J. Combin. 7 (2000) R33, 22.
[3] A. Buchstab, Rec. Math. [Mat. Sbornik] 44 (2) (1937) 1239–1246.
[4] K.J. Compton, Some methods for computing component distribution probabilities in relational structures, Discrete Math. 66 (1987) 59–77.
[5] P. Flajolet, M. Soria, Gaussian limiting distribution for the number of components in combinatorial structures, J. Combin. Theory A 53 (1990) 165–182.
[6] F. Harary, R.W. Robinson, The number of achiral trees, J. Reine Angew. Math. 278 (1975) 322–335.
[7] A. Knopfmacher, J. Knopfmacher, Arithmetical semigroups related to trees and polyhedra, J. Combin. Theory A 86 (1999) 85–102.
[8] R. Otter, The number of trees, Ann. Math. 49 (1948) 583–599.
[9] E.M. Palmer, A.J. Shwenk, On the number of trees in a random forest, J. Combin. Theory B 27 (1979) 109–121.
[10] D. Panario, L.B. Richmond, Smallest components in decomposable structures: exp-log class, Algorithmica 29 (2001) 205–226.
[11] G. Tenenbaum, Introduction to Analytic and Probabilistic Number Theory, Cambridge University Press, Cambridge, 1995.
[12] N. Wormald, Achiral plane trees, J. Graph Theory 2 (1978) 189–208.