The Probability of Connectedness

Edward A. Bender
Department of Mathematics, University of California, San Diego

November 2, 1998

[summary by Bruno Salvy]

A graph is a set of connected components. Graphs of various kinds are obtained by imposing constraints on these components. If c_n is the number of different components of size n and a_n the number of graphs of size n, then c_n/a_n is the probability that a graph selected uniformly at random among all graphs of size n is connected. The aim of this work is to study to what extent structural properties of the sequence $\{c_n\}$ make it possible to determine the asymptotic probability of connectedness (as the size n tends to infinity).

The asymptotic properties of c_n/a_n are closely related to properties of the generating functions of these sequences. Two cases are to be considered. In the *labelled* case, the generating functions under study are

$$A(x) = \sum_{n\geq 0} a_n \frac{x^n}{n!}, \qquad C(x) = \sum_{n\geq 0} c_n \frac{x^n}{n!}$$

and they are connected by

$$A(x) = \exp(C(x)). \tag{1}$$

In the *unlabelled* case, the generating functions are

$$A(x) = \sum_{n\geq 0} a_n x^n, \qquad C(x) = \sum_{n\geq 0} c_n x^n$$

and they are connected by

$$A(x) = \exp\left(C(x) + C(x^2)/2 + C(x^3)/3 + \cdots\right). \tag{2}$$

(See for instance [3] for a proof.) From the asymptotic point of view, these two identities are sufficiently close to make most of the proofs go through from one case to the other, with technical complications in the unlabelled case.

Example. General labelled rooted trees with n vertices are counted by n^{n-1}. The exponential generating function is the tree function $T(z)$ defined by

$$T(z) = z\exp(T(z)).$$

The corresponding forests have generating function $\exp(T(z)) = T(z)/z$. The dominant singularity of $T(z)$ is $\exp(-1)$ where the singularity is of square root type. Singularity analysis then shows that the asymptotic probability of connectedness is $\exp(-1)$.

	possible values for (ρ_ℓ, ρ_u)
$R = 0$	$[0,1] \times \{1\}$
C divergent at R	$\{0\} \times [0,1]$
C convergent at R	$[0,1) \times (0,1]$ with $\rho_\ell \le \rho_u$

TABLE 1. Conjectured possibilities for (ρ_ℓ, ρ_u)

Let R be the radius of convergence of the series $C(x)$. In the unlabelled case, since the c_n are integers, $R \le R_{\max} = 1$, while $R_{\max} = \infty$ in the labelled case. It is useful to distinguish three situations: $R = 0$, C converges at R, or C diverges at R (which can be infinite). Defining

$$\rho_\ell = \liminf_{n\to\infty} c_n/a_n, \qquad \rho_u = \limsup_{n\to\infty} c_n/a_n,$$

the aim of this work is to study when $\rho_\ell = \rho_u$ and to show as much as possible of Table 1.

A first result in this area is the following.

Theorem 1 ([5]). *A necessary and sufficient condition for asymptotic connectedness ($\rho_\ell = \rho_u = 1$) is that $R = 0$ and*

$$\sum_{i=1}^{n-1} h_i h_{n-i} = o(h_n) \tag{3}$$

where h_n is any of a_n or c_n.

Note that (3) is satisfied with $h_n = a_n$ if and only if it is satisfied with $h_n = c_n$.

Example. General undirected graphs with n vertices are enumerated by $a_n = 2^{n(n-1)/2}$ which accounts for all choices of edges. The theorem shows that $c_n \sim a_n$.

The remainder of this summary is devoted to proving parts of Table 1.

1. It is Always Possible that $\rho_\ell = 0$ and $\rho_u = 1$

This is shown by constructing an *ad hoc* sequence c_n which is 1 for most n and very large at rare points. Then a_n tends to infinity so that $\rho_\ell = 0$ and $\rho_u = 1$ because for those large c_n, $a_n \sim c_n$.

This idea might extend to obtain $0 = \rho_\ell < \rho_u < 1$ by taking more frequent large c_n in order to break the last equivalence.

2. Divergent Case: $\rho_\ell = 0$

This is a result of [4], which is proved as follows. If there exists $\delta > 0$ such that $c_n > \delta a_n$ for all n sufficiently large, then we get for $0 \le z < R$

$$C(z) > \delta e^{C(z)(+\cdots)} + P(z),$$

where P is a polynomial and the dots indicate more positive terms that are present in the unlabelled case. In both cases, this inequality implies that C is convergent at R.

3. **Convergent Case:** $\rho_u > 0$ **and** $\rho_\ell < 1$

The second inequality is a consequence of Wright's theorem.

The first one can be proved as follows in the labelled case. Differentiating (1) and extracting coefficients yields

$$\frac{a_n}{n!} = \frac{1}{n}\sum_{k=0}^{n} k\frac{c_k}{k!}\frac{a_{n-k}}{(n-k)!}.$$

If $c_n = o(a_n)$, then cutting the sum at $n^{1/2}$ and using $c_k \le a_k$ in the first part shows that

$$[x^n]A(x) = o([x^n]A(x)^2),$$

which implies divergence of A at R.

4. **When** $\rho_\ell = \rho_u$

The result is that every time (ρ, ρ) is present in Table 1, then there are sequences a_n and c_n of nonnegative integers making this happen. The first two lines of the table are dealt with by exhibiting appropriate examples: general labelled graphs for the first one; partitions of sets for the other one in view of the asymptotics of Bell numbers.

In the convergent case, an important tool is the following theorem.

Theorem 2. *In the convergent case, if* $\lim c_{n-1}/c_n$ *exists (then it is* R*) and* $\sum_{k=\omega}^{n-\omega} c_k c_{n-k} = o(c_n)$ *for any* $\omega(n) \to \infty$*, then* $\rho_\ell = \rho_u = 1/A(R)$.

We first show how this theorem is used to prove that every $\rho \in (0,1)$ is reached in the labelled case. The principle is to construct a sequence of generating functions $C^{[i]}(x)$ such that the coefficients $j![x^j]C^{[i]}(x)$ are nonnegative integers for $0 \le j < i$ and the value of $\lim a_n^{[i]}/c_n^{[i]}$ is ρ. Start with

$$C^{[0]}(x) = \alpha \sum \frac{(x/R)^n}{n^2}.$$

Then by the theorem, $\rho = \exp(-\alpha\pi^2/6)$, which fixes α. To construct $C^{[k+1]}$ from $C^{[k]}$, the coefficient of $k!x^k$ is replaced by its integer part, and the coefficient of x^{k+1} is increased to keep ρ unchanged. The increase is at most $R^{-1}/k!$ which is sufficiently small compared to its original value so that the conditions of the theorem still hold. Therefore the limit $C^{[\infty]}$ also satisfies the theorem. A similar argument gives the unlabelled case.

Proof of the theorem. In the labelled case, the hypothesis is used in an induction on d to obtain the following asymptotic estimates and bounds on $c_n^{(d)} = [x^n]C(x)^d$:

- $c_n^{(d)} < K^{d-1}c_n$ for some K and sufficiently large n;
- $c_n^{(d)} \sim dC(R)^{d-1}c_n$ uniformly for $d \le D(n)$, where $D(n) \to \infty$.

The conclusion follows from there by extracting the coefficient of x^n in $A(x) = \sum C(x)^d/d!$.

A proof in the unlabelled case is given in [2]. □

5. **Conclusion**

Many properties related to connectedness can be deduced from very little information on the counting sequence of the connected components. Much more than indicated here is known if extra smoothness conditions on the sequence are satisfied. Also, results regarding the limiting distribution are known. We refer to [2] for details. Still, a large part of Table 1 remains unproved, mostly regarding the existence of structures with the announced (ρ_ℓ, ρ_u).

Bibliography

[1] Bender (E. A.), Cameron (P. J.), Odlyzko (A. M.), and Richmond (L. B.). – Connectedness, classes and cycle index. *Combinatorics, Probability and Computing*, vol. 8, 1999, pp. 31–43.

[2] Bender (Edward A.), Cameron (Peter J.), and Richmond (L. Bruce). – Asymptotics for the probability of connectedness and the distribution of number of components, 1999. Preprint.

[3] Bergeron (F.), Labelle (G.), and Leroux (P.). – *Combinatorial species and tree-like structures.* – Cambridge University Press, Cambridge, 1998, xx+457p. Translated from the 1994 French original by Margaret Readdy, With a foreword by Gian-Carlo Rota.

[4] Cameron (Peter J.). – On the probability of connectedness. *Discrete Mathematics*, vol. 167/168, 1997, pp. 175–187.

[5] Wright (E. M.). – A relationship between two sequences. *Proceedings of the London Mathematical Society*, vol. 17, 1967, pp. 296–304.