THE PROBABILITY OF CONNECTEDNESS OF A LARGE UNLABELLED GRAPH

BY E. M. WRIGHT

Communicated by Gian-Carlo Rota, December 18, 1972

An \((n, q)\) graph is one with \(n\) nodes and \(q\) edges, in which any two different nodes are or are not joined by a single edge. We write \(T = T(n, q)\) for the number of different \((n, q)\) graphs with unlabelled nodes and \(t\) for the number of these graphs which are connected, so that \(\beta = t/T\) is the probability that an unlabelled \((n, q)\) graph is connected. We write \(F, f\) and \(\alpha\) for the corresponding numbers for \((n, q)\) graphs whose nodes are labelled. We write also \(N = n(n - 1)/2\), \(B(h, k) = h!/[k!(h - k)!]\) and \(\gamma = (2q - n \log n)/n\). Clearly \(q \leq N\). In what follows, \(A\) (not always the same at each occurrence) is a fixed positive number at our choice and all statements are true only for \(n > n_0, q > q_0\), where \(n_0\) and \(q_0\) depend on the \(A\).

Erdös and Renyi [1] put \(q = [n(log n + a)/2]\), where \(a\) is independent of \(n\) and \(q\), and showed that, for these \(q\), we have

\[
\alpha \to \exp(e^{-a})
\]
as \(n \to \infty\). For given \(n\), it can be shown trivially that \(\alpha\) increases steadily (in the nonstrict sense) as \(q\) increases. Hence, from (1), it can be at once deduced that, as \(n \to \infty\), we have \(\alpha \sim \exp(e^{-\gamma})\) and, in particular, that

\[
\alpha \to 1 \ (\gamma \to + \infty), \quad \alpha \to 0 \ (\gamma \to - \infty).
\]

Elsewhere [4] I have shown that, if \(\gamma \to + \infty\), then \(f\) has an asymptotic expansion of which the first two terms are

\[
f = B(N, q) - nB(N - n + 1, q) - \cdots .
\]

Now \(F = B(N, q)\) and

\[
\frac{nB(N - n + 1, q)}{B(N, q)} = n \prod_{s=0}^{q-1} \frac{N - n + 1 - s}{N - s} \leq n(N - n + 1)^q N^{-q}
\]

and the logarithm of this is less than \(\log n - \{q(n - 1)/N\} = - \gamma\). Hence my result leads to \(\alpha = 1 - O(e^{-\gamma})\), a statement which is only nontrivial...
when $\gamma \to +\infty$. Thus, for this range of $q$, I obtain a much more detailed result than Erdös and Renyi. On the other hand, my method (depending on Gilbert's [2] generating functions identity) appears incapable of extension to obtain (1), as indeed Erdös and Renyi remark.

My first theorem here gives a result for $\beta$ corresponding to, but differing from, Erdös and Renyi's result for $\alpha$. The proof depends heavily on the results of [5] and [7].

**Theorem 1.** As $n \to \infty$, we have

$$\beta \sim 1 - e^{-\gamma} \quad (A < \gamma < A),$$

$$\beta \to 0 \quad (\lim_{} \gamma \leq 0),$$

$$\beta \to 1 \quad (\gamma \to +\infty).$$

These results are in striking contrast to Erdös and Renyi's. They imply that, when $-A < \gamma < A$, a substantially higher proportion of the labelled graphs are connected than of the unlabelled, at least in the limit as $n \to \infty$.

But there is another, and much more interesting difference in the proof required when $\beta \to 0$ or $\beta \to 1$. Erdös and Renyi [1] did not need to consider the corresponding cases for $\alpha$ since, for fixed $n$, the number $\alpha$ increases (nonstrictly) with $q$. No such result is known for $\beta$ and indeed, as I showed in [6], no such result is true.

The behavior of $\beta$ for fixed $n$ as $q$ increases presents an interesting problem. Obviously $\beta = 0$ for $q \leq n - 2$ and $\beta = 1$ for $N - n + 2 \leq q \leq N$. What appears to be true otherwise (by calculations based on the table [3]) is that, for fixed $n \geq 6$ and some $q_1 = q_1(n)$, we have

$$\beta(n, q) < \beta(n, q + 1) \quad (n - 2 \leq q < q_1),$$

$$\beta(n, q) > \beta(n, q + 1) \quad (q_1 \leq q \leq N - n).$$

All that I can prove, however, is the following theorem.

**Theorem 2.** For $n > n_0$ and some $q_1 = q_1(n)$, we have

$$(2) \quad \beta(n, q) < \beta(n, q + 1) \quad (m(A + \log n)/2 < q < q_1),$$

$$(3) \quad \beta(n, q) > \beta(n, q + 1) \quad (q_1 \leq q \leq N - n).$$

We can calculate the integer $q_1$ with a possible error of 1.

It is surprising that we can define so precisely the range of validity of the unexpected result (3). On the other hand, I cannot prove (2) for $\gamma \leq 0$, i.e. for $2q \leq n \log n$, although the tables [3] and common sense (that dubious guide) combine to indicate that it must be true. In fact, the proof of (2) for $N/2 \leq q < q_1$ is easier than that for $q \leq N/2$ and, in particular, my present proof of (2) for $A < \gamma < A$ is not at all simple.
REFERENCES


Principal's Office, University of Aberdeen, Aberdeen, United Kingdom