THE PROBABILITY OF CONNECTEDNESS OF AN UNLABELLED GRAPH CAN BE LESS FOR MORE EDGES

E. M. WRIGHT

Abstract. We write $\beta = \beta(n, q)$ for the probability that a graph on $n$ unlabelled nodes with $q$ edges is connected; that is $\beta$ is the ratio of the number of connected graphs to the total number of graphs. We write $N = n(n-1)/2$. For fixed $n$ we might expect that $\beta$ would increase with $q$, at least nonstrictly. On the contrary, we show that, for any given integer $s$, we have $\beta(n, q+1) < \beta(n, q)$ for $N - n - s \leq q \leq N - n$ and $n > n_0(s)$. We can show that $\beta(n, q+1) < \beta(n, q)$ for a much longer range, but this requires much more elaborate arguments.

An $(n, q)$ graph has $n$ nodes and $q$ edges, where each edge is a different unordered pair of nodes. We write $N = n(n-1)/2$, so that $0 \leq q \leq N$. We write $F_{nq}$ (resp. $T_{nq}$) for the number of $(n, q)$ graphs in which the nodes are labelled (resp. unlabelled) and $f_{nq}$ (resp. $t_{nq}$) for the number of these graphs which are connected. Then $\alpha_{nq} = f_{nq}/F_{nq}$ (resp. $\beta_{nq} = t_{nq}/T_{nq}$) is the probability that a labelled (resp. unlabelled) $(n, q)$ graph is connected. It is natural to expect that, for fixed $n$, these probabilities will increase (in the nonstrict sense) as $q$ increases. For $\alpha_{nq}$, this is true and can be proved trivially. For $\beta_{nq}$ it is false; the simplest counterexample is that $\beta_{69} = 20/21$ and $\beta_{610} = 14/15$ (see, for example, the diagrams of all $(6, 9)$ and $(6, 10)$ graphs in [6]).

The slightly surprising phenomenon that $\beta_{nq}$ can decrease as $q$ increases seems worth further study. Here we prove the following theorem.

**Theorem.** For any positive integer $v$, there is an $n_0 = n_0(v)$, such that

$$(1) \quad \beta_{n, N-n-s} > \beta_{n, N-n-s+1} \quad (0 \leq s \leq v)$$

when $n > n_0$ and, indeed, such that

$$(2) \quad 1 - \beta_{n, N-n-s+1} > C n^{1/2}(1 - \beta_{n, N-n-s}).$$

Presented to the Society, September 15, 1971 under the title *Decreasing probability of connectedness of a graph as number of edges increases*; received by the editors October 9, 1971.

AMS 1969 subject classifications. Primary 0565.

Key words and phrases. Unlabelled graphs, probability of connectedness.

$^1$ The research reported herein has been sponsored by the United States Government.

\[ c \quad \text{American Mathematical Society 1972} \]
Thus, for these values of \( q \), \( \beta_{n_0} \) retreats rapidly from 1 as \( q \) increases. Every statement in what follows is subject to the condition "for large enough \( n \)." The positive number \( C \), not always the same at each occurrence, is independent of \( n \) and \( q \).

For the moment we consider only unlabelled graphs. Any disconnected \((n, q)\) graph consists of a \((k, q_k)\) graph together with a \((n-k, q-q_k)\) graph for some \( q_k \) and for some \( k \) (not necessarily unique) such that \( 1 \leq k \leq n/2 \). Compared with the complete \((n, N)\) graph, the disconnected \((n, q)\) graph lacks at least \( k(n-k) \) edges, so that \( q \leq N-k(n-k) \). If \( q = N-n-s \), where \(-1 \leq s \leq n-5\), we have \( k = 1 \). Hence a disconnected \((n, N-n-s)\) graph has just 2 components, one an isolated node and the other an \((n-1, N-n-s)\) graph. The latter graph, compared with the complete \((n-1, (n-1)(n-2)/2)\) graph, lacks \( (n-1)(n-2)/2 - (N-n-s) = s+1 \) edges. Hence it is the complement of an \((n-1, s+1)\) graph. The relation is 

\[
T_{n,N-n-s} = T_{n-1,s+1} \quad ( -1 \leq s \leq n-5 ).
\]

Clearly \( T_{n-1,0} = T_{n-1,1} = 1 \) and so \( \beta_{n,N-n+1} < \beta_{n,N-n} \) provided \( T_{n,N-n+1} < T_{n,N-n} \), i.e. provided \( T_{n-1,n} < T_{n,n} \). The smallest \( n \) for which this is true is 6, where \( T_{65} = 15, T_{66} = 21 \) (see [6]). Hence \( \beta_{6,10} < \beta_{6,9} \), the counterexample given above.

If \( n \geq 2q \), the \((n, q)\) graphs can be put into \((1, 1)\) correspondence with the \((2q, q)\) graphs by removing \( n-2q \) of the isolated nodes from each of the former and conversely. Hence \( T_{n,q} = T_{2q,q} \) and, from (3), we have

\[
T_{n,N-n-s} = t_{n,N-n-s} = T_{2(s+1),s+1} \quad ( -1 \leq s \leq v ).
\]

To prove (2), we have then to show that

\[
T_{2s+2,s+1}/T_{2s,s+1} > C n^{1/2} T_{2s+2,s+1}/T_{2s,s+1}
\]

Since \( T_{2s+2,s+1}/T_{2s,s} = C \), it is enough to show that \( T_{n,N-n-s} > C n^{1/2} T_{n,N-n-s+1} \), that is \( T_{n,n+s} > C n^{1/2} T_{n,n+s-1} \).

We shall prove a little more, viz. the following lemma.

**LEMMA.** If \( Cn < q < Cn \), then \( T_{n,q+1} > C n^{1/2} T_{n,q} \).

Henceforth we suppose the hypothesis of the lemma to be satisfied. We now use Polya’s famous counting theorem [1], [5], [8]. \( S_n \) is the symmetric group (of degree \( n \) and order \( n! \)) of all the permutations \( \pi \) of the \( n \) labelled nodes in a labelled \((n, q)\) graph, and \( F_\pi = F_\pi(n, q) \) is the number of labelled \((n, q)\) graphs invariant under \( \pi \). Then Polya’s theorem tells us that

\[
n! T_{n,q} = \sum_{\pi \in S_n} F_\pi.
\]
We are henceforth concerned with the enumeration of labelled graphs. The permutation \( \pi \) has a unique expression as a product of disjoint cycles, which contains \( p_j \) node-cycles of length \( j \), where \( 1 \leq j \leq n \). The node permutation \( \pi \) induces a permutation on the \( N \) possible edges in the graph which contains \( P_j \) edge-cycles of length \( j \). We have

\[
\sum_{j=1}^{n} j p_j = n, \quad \sum_{j=1}^{n} j P_j = N.
\]

Oberschelp [7] gives formulae for \( P_j \) in terms of the \( p_j \). All we require here is to note that

\[
P_1 = \frac{1}{2} p_1 (p_1 - 1) + p_2.
\]

Any graph invariant under \( \pi \) must contain all the edges of a particular edge-cycle or none. Hence

\[
F_\pi(n, q) = \sum' \prod_{j=1}^{n} B(P_j, s_j),
\]

where \( B(h, k) = h!/(k!(h-k)!) \), \( B(h, 0) = 1 \) and \( \sum' \) denotes summation over all \( s_1, \cdots, s_n \) such that \( \sum j s_j = q \). We write \( M = n^{3/4} \) and separate the sum on the right-hand side of (4) into two parts, so that

\[
n! T_{nq} = T_1(q) + T_2(q),
\]

where \( T_1(q) \) contains all the \( F_\pi \) for which \( p_1 > M \) and \( T_2(q) \) those for which \( p_1 \leq M \).

First we take \( p_1 > M \), so that, \( p_1 > Cn^{3/2} \). In (7), we replace \( q \) by \( q+1 \), discard those terms on the right in which \( s_1 = 0 \) and replace \( s_1 \) by \( s_1 + 1 \) in the others. We have then

\[
F_\pi(n, q + 1) \geq \sum' \lambda(P_1, s_1) \prod_{j=1}^{n} B(P_j, s_j) \geq Cn^{1/2} F_\pi(n, q),
\]

where

\[
\lambda(P_1, s_1) = (P_1 - s_1)/(s_1 + 1) \geq (Cn^{3/2} - q)/(q + 1) \geq Cn^{1/2}
\]

under the hypothesis of the lemma. Hence

\[
T_1(q + 1) > Cn^{1/2} T_1(q).
\]

We write \([R(X)]_q\) for the coefficient of \( X^q \) in the polynomial \( R(X) \). From (7),

\[
F_\pi(n, q) = \left[ \prod (1 + X^j)^{P_j} \right]_q \leq Y^{-q} \prod (1 + Y^j)^{P_j},
\]

where \( Y \) is any positive number. Again, if \( j > 2 \), we have \( (1 + Y^j)^2 \leq (1 + Y^2)^j \). Hence

\[
\prod (1 + Y^j)^{P_j} \leq (1 + Y)^{P_1} (1 + Y^2)^{(N - P_1)/2}
\]
by (5). We now put $Y = ((q/(N-q)))^{1/2}$ and have $F_\varepsilon(n, q) \leq V^{1/2} W$, where

$$V^2 = (1 + Y)^2/(1 + Y^2) \leq 1 + Cn^{-1/2},$$

and

$$W = Y^{-3}(1 + Y^2)^{N/2} = (N/q)^{q/2}(N/(N-q))^{(N-q)/2}.$$

If $p_1 \leq M$ we have $P_1 \leq Cn^{3/2}$ and $P_1 \log V < Cn$. Again

$$\log W \leq (q/2)\log(N/q) + {(N - q)/2}\log(1 + q/(N-q))$$

$$\leq (q/2)\log q + Cn.$$

Hence $\log F_\varepsilon(n, q) \leq (q/2)\log q + Cn$. There are less than $n!$ terms in $T_2(q)$ and so

$$T_2(q) \leq n! \quad q^{q/2} e^{Cn}.$$  \hspace{1cm} (10)

We now write $p = \lfloor(2q)/\log(2q)\rfloor$ so that $M < p \leq n$. The number of $\pi$ for which $p_1 = p$ is $B(n, p)D(n-p)$, where $D(m)$ is the number of permutations of the $m$ numbers 1, 2, ..., $m$ in which no number remains unmoved, i.e. there are no unit cycles. Then $D(m)$ is Euler's rencontre number and is the nearest integer to $m!/e$ (see, for example, [3], [4], [9], [12]). Hence the number of $\pi$ for which $p_1 = p$ is greater than $Cn!/p!$. Again, by (6), for such a $\pi$, we have $P_1 \geq \frac{1}{2} p(p-1)$ and so

$$F_\varepsilon \geq B(P_1, q) \geq B(p(p-1)/2, q) = p! \Lambda_p$$

(say). Now $\log(p!) = O(p \log p) = O(q)$ and so

$$\log \Lambda_p = 2q \log p - q \log q + O(q) = q \log q - 2q \log \log q + O(n).$$

Hence

$$T_1(q) \geq C(n!)\Lambda_p \geq C(n!)(q/(\log q)^2)^q e^{-Cn}.$$  \hspace{1cm} (11)

It follows from (8), (10) and (11) that

$$T_2(q) = o(T_1(q)), \quad n! T_{nq} = T_1(q)(1 + o(1)).$$

This is true when $q+1$ replaces $q$ and so our lemma follows from (9).

We can improve the factor $n^{1/2}$ in (2) to $n/(\log n)^2$. More significantly, we can show that (1) holds over a range of $q$ of length $C_1n$ for any fixed positive $C_1$ and large enough $n$. It is possible that we can replace $C_1n$ by a constant multiple of $n \log n$. But these results require the development of a much more elaborate theory, and, in particular, a study of the asymptotic behaviour of $T_{nq}$ and of $T_{n,q-1}/T_{nq}$ for large $n$ and large $q$ such that $q < (n \log n)/2$. 


REFERENCES


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