Annals of Discrete Mathematics 13 (1982) 33-50

DISTINGUISHING VERTICES OF RANDOM GRAPHS

BÉLA BOLLOBÁS

University of Cambridge
England

The *distance sequence* of a vertex x of a graph is $(d_i(x))_1^n$, where $d_i(x)$ is the number of vertices at distance i from x. The paper investigates under what condition it is true that almost every graph of a probability space is such that its vertices are uniquely determined by an initial segment of the distance sequence. In particular, it is shown that for $r \geq 3$ and $\varepsilon > 0$ almost every labelled r-regular graph is such that every vertex x is uniquely determined by $(d_i(x))_1^u$, where $u = \lfloor (\frac{1}{2}+\varepsilon)\frac{\log n}{\log(r-1)} \rfloor$. Furthermore, the paper contains an entirely combinatorial proof of a theorem of Wright [10] about the number of unlabelled graphs of a given size.

§0. INTRODUCTION

In this note we shall study two models of random graphs: $G(n,M)$ and $G(n,r\text{-reg})$. The first consists of all graphs of size M with vertex set $V = \{1,2,\ldots,n\}$ and the second is the set of all r-regular graphs with vertex set V. In both models all graphs are given the same probability. When dealing with the model $G(n,r\text{-reg})$ we shall assume that rn is *even*, so there are r-regular graphs of order n if $n \geq r+1$. For the basic facts concerning random graphs we refer the reader to [2, Ch.VII]. In particular, we say that *almost every* (a.e.) graph in a model has a certain property if the probability of having the property tends to 1 as the number of vertices tends to ∞.

Let G be $G(n,M)$ or $G(n,r\text{-reg})$. We are interested in the existence of indistinguishable vertices of a graph $G \in G$ for two different definitions of being distinguishable. First we consider two vertices of G *distinguishable* if no automorphism of G maps one of the vertices into the other one. Thus $G \in G$ consists of distinguishable vertices iff the automorphism group of G is trivial. To give the second definition of being distinguishable, given a vertex $x \in G$ write $d_i(x)$ for the number of vertices at distance i from x. Call $(d_i(x))_1^n$ the *distance sequence of* x and define two vertices to be *distinguishable* if they have different

sequences. Clearly two vertices distinguishable in this sense are also distinguishable in the first sense.

A fundamental theorem of Wright [10] implies that if $M = M(n)$ is such that almost no graph G_M in $G(n,M)$ has two isolated vertices or two vertices of degree $n-1$, then almost every random graph G_M has trivial automorphism group. In fact the theorem states much more namely that the number of labelled graphs of order n and size M divided by the number of unlabelled graphs of order n and size M is asymptotic to $n!$. In §1 we give an entirely combinatorial proof of this result.

The rest of the paper is concerned with the problem of distinguishing vertices with the aid of their distance sequences. In §2 we treat the random graphs $G_M \in G(n,M)$ and in §3 we study random regular graphs of a fixed degree r. In both cases our aim is to use small portions of the distance sequences to identify the vertices.

§1. THE NUMBER OF UNLABELLED GRAPHS

Denote by $U_M = U_{n,M}$ the number of unlabelled graphs of order n and size $M = M(n)$ and write $L_M = L_{n,M}$ for the number of labelled graphs. We shall put $N = \binom{n}{2}$ so that $L_M = \binom{N}{M}$. Our aim is to show that under suitable conditions on M we have

$$U_M \sim L_M/n! = \binom{N}{M}/n! \,. \tag{1}$$

This is clearly considerably stronger than the assertion that the automorphism group of a.e. $G_M \in G(n,M)$ is trivial.

If the automorphism group of a graph of order n is trivial then the graph contains at most one isolated vertex and at most one vertex of degree $n-1$. It is easily seen (see e.g. Erdős and Rényi [6] or [7]) that this happens if and only if

$$\frac{2M}{n} - \log n \to \infty \quad \text{and} \quad \frac{2(N-M)}{n} - \log n \to \infty \,.$$

It is surprising that this simple necessary condition on M is sufficient to imply (1). This important result is due to Wright [10]; Earlier Pólya (see [8]) and Oberschelp [9] had proved (1) under considerably stronger conditions on M. Our aim is to give a combinatorial proof of this theorem.

Consider the symmetric group S_n acting on $V = \{1,2,\ldots,n\}$. For $\omega \in S_n$ let $G(\omega)$ be the set of graphs in G_M invariant under ω and put $I(\omega) = |G(\omega)|$. Then

$$\sum_{\omega \in S_n} I(\omega) = \sum_{G \in G_M} a(G) \,,$$

where $a(G)$ is the order of the automorphism group of $G \in G_M$: $a(G) = |\mathrm{Aut}\, G|$. Since G_M has exactly $n!/a(G)$ graphs isomorphic

to a given graph $G \in G_M$,

$$U_M = \sum_{G \in G_M} (n!/a(G))^{-1} = \frac{1}{n!} \sum_{G \in G_M} a(G) = \frac{1}{n!} \sum_{\omega \in S_n} I(\omega) \ .$$

For the identity permutation $1 \in S_n$ we have $I(1) = L_M$, so (1) holds iff

$$\sum_{\substack{\omega \in S_n \\ \omega \neq 1}} I(\omega) = o(L_M) \ . \tag{2}$$

The number $I(\omega)$ depends only on the size of the orbits of ω acting on $V^{(2)}$, the set of N pairs of vertices, because a graph G_M belongs to $G(\omega)$ if and only if the edge set of G_M is the union of some entire orbits of ω acting on $V^{(2)}$.

Let A be a fixed set with a elements. We shall consider set systems $\mathcal{A} = \{A_1, A_2, .., A_s\}$ partitioning A : $A = \bigcup_{i=1}^{s} A_i$. Denote by $F(b;\mathcal{A}) = F(b;A_1,\ldots,A_s)$ the collection of those b-element subsets of A which are the unions of some of the sets A_i . Thus $B \in F(b;\mathcal{A})$ iff $|B| = b$ and for every i we have $A_i \subset B$ or $B \cap A_i = \emptyset$. Suppose there are m_1 sets A_i of size 1, m_2 sets of size 2,..., m_t sets of size t and no set with more than t elements. Note that these parameters satisfy

$$a = \sum_{j=1}^{t} j m_j \ . \tag{3}$$

The definition of $F(b;\mathcal{A})$ implies that if $f(b;\mathcal{A}) = |F(b;\mathcal{A})|$ then

$$f(b;\mathcal{A}) = f(b;m_1,m_2,\ldots,m_t) = \sum_{(i_j)} \binom{m_1}{b-k} \prod_{j=2}^{t} \binom{m_j}{i_j} \ , \tag{4}$$

where the summation is over all $(i_j)_2^t$ satisfying $0 \le i_j \le m_j$ and $k = k((i_j)_2^t) = \sum_{j=2}^{t} j i_j \le b$, $b-k \le m_1$.

If $\mathcal{A}'$ is obtained from $\mathcal{A}$ by decomposing an A_i into some sets (that is if the partition $\mathcal{A}'$ is a *refinement* of $\mathcal{A}$) then clearly $F(b;\mathcal{A}) \subset F(b;\mathcal{A}')$, so

$$f(b;\mathcal{A}) \le f(b;\mathcal{A}') \ . \tag{5}$$

In particular, if $m_1+2m_2 = a$ and $0 \le k \le m_2$ then

$$f(b;m_1,m_2) \le f(b;m_1 + 2k, m_2-k) \ . \tag{6}$$

Furthermore, if $\mathcal{A}$ is the union of two disjoint systems, say $\mathcal{A}'$ and $\mathcal{A}''$, then

$$f(b;\mathcal{A}) = \sum_{i=0}^{b} f(b-i;\mathcal{A}')f(i;\mathcal{A}'') \ .$$

Consequently if $m_j = m_j' + m_j''$ with $m_j' \geq 0$ and $m_j'' \geq 0$ then

$$f(b;m_1,\ldots,m_t) = \sum_{i=0}^{b} f(b-i;m_1',\ldots,m_t')f(i;m_1'',\ldots,m_t'') \ . \qquad (7)$$

The last property of the function $f(b;m_1,\ldots,m_t)$ we need is not quite immediate so we state it as a lemma.

Lemma 1. Suppose $m_1 \leq a-2$. If $a-m_1$ is odd,

$$f(b;m_1,\ldots,m_t) \leq f(b;m_1+1,\tfrac{1}{2}(a-m_1-1))$$

and if $a-m_1$ is even then

$$f(b;m_1,m_2,\ldots,m_t) \leq f(b;m_1+2,\tfrac{1}{2}(a-m_1-2)) \ .$$

Proof. Since a set A_i with more than 3 elements can be partitioned into sets of size 2 and 3, by (5) we may suppose that A contains no sets of size greater than 3 , that is $m_4 = m_5 = \ldots = m_t = 0$. Furthermore, if $m_3 = 0$ we are home since then $a-m_1 = 2m_2$ and by (6)

$$f(b;m_1,m_2,\ldots,m_t) = f(b;m_1,m_2) \leq f(b;m_1+2,m_2-1) \ .$$

Now suppose that $m_3 > 0$. We distinguish two cases according to the parity of m_3 , which is also the parity of $a-m_1$. First suppose that m_3 is odd, say $m_3 = 2k+1 \geq 1$. We know from (7) that

$$f(b;m_1,m_2,2k+1) = \sum_{i=0}^{b} f(b-i;m_1,m_2)f(i;0,0,2k+1)$$

and

$$f(b;m_1+1,m_2+3k+1) = \sum_{i=0}^{b} f(b-i;m_1,m_2)f(i;1,3k+1) \ .$$

Hence it suffices to show that

$$f(i;0,0,2k+1) \leq f(i;1,3k+1) \ . \qquad (8)$$

In order to show (8) we use (3) to write out the two sides explicitly. If the left-hand side of (8) is not zero then i is a multiple of 3. Furthermore, if $i = 6j$ then (8) becomes

$$\binom{2k+1}{2j} \leq \binom{3k+1}{3j}$$

and if $i = 6j+3$ then (8) is

$$\binom{2k+1}{2j+1} \leq \binom{3k+1}{3j+1} \ .$$

Both inequalities are obvious.

Secondly, if m_3 is even, say $m_3 = 2k+2$ then

$$f(b;m_1,m_2,2k+2) \leq f(b;m_1+1,m_2+1,2k+1 \leq f(b;m_1+2,m_2+3k+2) \ ,$$

completing the proof. □

Let us return to our central theme, relation (2). We shall prove it by estimating $I(\omega)$ in terms of the number of vertices fixed by ω . Denote by $S_n^{(m)}$ the set of permutations moving m vertices,

that is fixing n-m vertices. Then $S_n^{(0)} = \{1\}$ and $S_n^{(1)} = \emptyset$. Furthermore, if $\omega \in S_n^{(m)}$ then the m vertices moved by ω can be selected in $\binom{n}{m}$ ways and there are at most m! permutations moving the same set of m vertices. Hence

$$|S_n^{(m)}| \le \binom{n}{m} m! = (n)_m \ . \tag{9}$$

Lemma 2. Let $2 \le m \le n$, $n \ge 3$, and denote by $N_1 = N_1(m)$ the integer satisfying

$$\binom{n-m}{2} + \frac{m+1}{2} \le N_1 \le \binom{n-m}{2} + \frac{m+4}{2}$$

for which $N_2 = (N-N_1)/2$ is also an integer.

Then for $\omega \in S_n^{(m)}$ we have

$$I(\omega) \le f(M;N_1,N_2) \ .$$

Proof. Suppose $\omega \in S_n^{(m)}$ as a permutation acting on $V^{(2)}$ has M_i orbits of size i, i = 1,2,...,n . Since a graph G_M belongs to $G(\omega)$ iff its edge set is the union of some entire orbits,

$$I(\omega) = |G(\omega)| = f(M;M_1,M_2,\ldots,M_n) \ .$$

Hence if M_j' is chosen to be M_1+1 or M_1+2 so that $M_2' = (N-M_1')/2$ is an integer, then by Lemma 1 we have

$$I(\omega) \le f(M;M_1',M_2') \ .$$

At most how large is M_1 ? If a permutation fixes two pairs of vertices, say $\{x,y\}$ and $\{y,z\}$, then it also fixes y , the common vertex, and so x and z as well. Consequently a pair fixed by $\omega \in S_n^{(m)}$ either consists of two vertices fixed by ω or it is disjoint from all other pairs fixed by ω . Hence

$$M_1 \le \binom{n-m}{2} + m/2$$

and so

$$M_1' \le N_1 \ .$$

Therefore by inequality (6) we have

$$I(\omega) \le f(M;M_1',M_2') \le f(M;N_1,N_2) \ . \qquad \square$$

It so happens that (2) can be proved fairly easily under a somewhat stronger assumption on M than necessary. Here we only state this result; for a proof we refer the reader to Wright [10]. As a consequence of this preliminary version of the main result in the proof of the main theorem we may assume that $M = O(n \log n)$.

Theorem 3. If

$$2n \log n \le M \le N - 2n \log n$$

then

$$U_M \sim L_M/n! \ . \qquad \square$$

Now we are ready to prove the fundamental theorem of Wright [10].

Theorem 4. Suppose $\psi(n) \to \infty$ and

$$\tfrac{1}{2}n(\log n + \psi(n)) \le M \le N - \tfrac{1}{2}n(\log n + \psi(n)) .$$

Then

$$U_M \sim L_M/n! .$$

Proof. The map sending a graph into its complement preserves the automorphism group and sets up a 1-1 correspondence between G_M and G_{N-M}. Hence in proving this result we may assume that $M \le N/2$. Furthermore, because of Theorem 3 we may and shall assume that

$$\tfrac{1}{2}n \log n + \psi(n)n \le M \le 2n \log n .$$

We shall also assume that $n \ge n_o$ where n_o depends on the function $\psi(n)$ and is chosen so that all our inequalities hold. It will be easy to check that there is such an n_o.

Let $N_1 = N_1(m)$ and $N_2 = N_2(m) = \frac{1}{2}(N-N_1)$ be as in Lemma 2 and set

$$L(m,i) = \binom{N_1}{M-2i}\binom{N_2}{i} ,$$

$$L(m) = \sum L(m,i) ,$$

where the summation is over all i satisfying $0 \le i \le N_2$ and $M-N_1 \le 2i \le M$. If $\omega \in S_n^{(m)}$ then by Lemma 2 we have

$$I(\omega) \le L(m) .$$

Hence by (2) and (9) the theorem follows if we show that

$$\sum_{m=2}^{n} (n)_m \, L(m)/L_M = o(1) . \tag{10}$$

For the sake of convenience let us recall the inequality N_1 satisfies:

$$m(n - \frac{m}{2} - 2) - 2 \le N-N_1 \le m(n - \frac{m}{2} - 2) - \tfrac{1}{2} . \tag{11}$$

Using standard estimates of binomial coefficients we find that

$$L(m,0)/L_M = \binom{N_1}{M}/\binom{N}{M} \le (N_1/N)^M \le \exp\{-2M \frac{m}{n}(1 - \frac{m+5}{2n})\} . \tag{12}$$

Set

$$\ell(m,i) = \frac{L(m,i)}{L(m,i-1)} = \frac{N-N_1-2i+2}{2i} \, \frac{(M-2i+2)(M-2i+1)}{(N_1-M-2i-1)(N_1-M+2i)}$$

and note that for a fixed value of m the ratio $\ell(m,i)$ is a decreasing function of i. In fact, if $1 \le i \le j$ then

$$\ell(m,j) \le \frac{i}{j} \, \ell(m,i)$$

and so

$$L(m,j) \le \frac{(\ell(m,1))^j}{j} \, L(m,0) . \tag{13}$$

In order to prove (10) we shall decompose the range of m into three intervals and show that the sum over each of these intervals is $o(1)$.

(a) Suppose that $2 \le m \le 10 \frac{n}{(\log n)^2}$. Then

$$\ell(m,1) = \frac{N-N_1}{2} \frac{M(M-1)}{(N_1-M+1)(N_1-M+2)}$$

$$\le \frac{mn}{2} \frac{(2n \log n)^2}{(N-mn-M)^2} \le 100 .$$

Since

$$\ell(m,1) \le \frac{1}{i} \ell(m,1) ,$$

this shows that

$$L(m) \le c\, L(m,0)$$

for some absolute constant c . Therefore by (12)

$$L(m)/L_M \le c \exp\{-2M \frac{m}{n} (1 - \frac{m}{2n})\}$$

$$\le c \exp\{-m(\log n + \psi(n))(1 - \frac{m}{2n})\}$$

$$\le c \exp\{-m \log n - \psi(n)m - 2\frac{m^2}{n} \log n\}$$

$$\le \exp\{-m \log n - \tfrac{1}{2}\psi(n)m\} .$$

Summing over the m in our range we find that

$$\sum_m (n)_m L(m)/L_M \le \sum_m \exp\{-\tfrac{1}{2}\psi(n)m\} = o(1) .$$

(b) Now let us turn to the largest interval we shall consider:

$$10 \frac{n}{(\log n)^2} \le m \le n - n^{\frac{3}{4}} (\log n)^2 .$$

It is easily checked that in this range $\ell(m,1) \ge 4$.

Put

$$j = \lfloor 2\ell(m,1) \rfloor .$$

Then

$$\ell(m,j+1) \le \frac{\ell(m,1)}{j+1} \le \frac{\ell(m,1)}{2\ell(m,1)} = \tfrac{1}{2}$$

so, estimating rather crudely,

$$L(m) \le 2 \sum_{i=1}^{j} \frac{(\ell(m,1))^i}{i} L(m,0) \le 3(\ell(m,1))^{2\ell(m,1)} L(m,0) . \qquad (14)$$

Since

$$N_1/N \le (\binom{n-m}{2} + m/2 + 2)/N \le (\frac{n-m}{n})^2 (1+n^{-\frac{1}{2}}) ,$$

we have

$$\log\{L(m,0)/L_M\} \le 2M \log \frac{n-m}{n} + M n^{-\frac{1}{2}} .$$

Furthermore,

$$N_1 - M > \frac{2}{5}(n-m)^2$$

so

$$\ell(m,1) \le \frac{n^2}{4} \frac{(2n \log n)^2}{(\frac{2}{5} n^{3/2} (\log n)^4)^2} < 7 n/(\log n)^6 .$$

Hence by (14) we have

$$\sum (n)_m L(m)/L_M = o(1)$$

for the sum over the range at hand, if, say,

$$(m+2)\log n + 14 \frac{n}{(\log n)^6} \log n + 2M \log \frac{n-m}{n} + Mn^{-\frac{1}{2}} < 0 . \qquad (15)$$

The derivative of the left-hand side with respect to m is

$$\log n - 2M/(n-m) < 0$$

so it suffices to check (15) for the minimal value of m, namely $m = \lceil 10n/(\log n)^2 \rceil$. Since

$$2M \ge n(\log n + \psi(n)) \quad \text{and} \quad \log \frac{n-m}{n} \le -\frac{m}{n} ,$$

the left-hand side of (15) is at most

$$2 \log n + 14n/(\log n)^5 - \psi(n)m + 2n^{\frac{1}{2}} \log n < 15n/(\log n)^5 - m < 0,$$

as required.

(c) Finally, suppose that

$$m \ge n - n^{\frac{3}{4}} (\log n)^2 .$$

In this range the crudest estimates will do. Since $N_2 \le N/2$,

$$L(M,i)/L_M = \binom{N_2}{i}\binom{N_1}{M-2i}/\binom{N}{M} \le \frac{(N/2)^i}{i!} \frac{N_1^{M-2i}}{(M-2i)!} \frac{M!}{(N-M)^M} = h(m,i) .$$

If $2(i+1) \le M$, we have

$$\frac{h(m,i+1)}{h(m,i)} = \frac{N}{2(i+1)} \frac{N_1^2}{(M-2i)(M-2i-1)} \ge \frac{n^4}{8M^3} > 2 ,$$

as $N_1 \ge n/2$ holds for every m. Furthermore, at the maximal value of i, $\lfloor M/2 \rfloor$, we have

$$h(m,\lfloor M/2 \rfloor) \le n^4 M^{M/2}/(N-M)^{M/2} \le n^{-3n} .$$

Consequently in our range we have

$$\sum (n)_m L(m)/L_M \le \sum (n)_m 2h(m,\lfloor M/2 \rfloor) \le n^{-n} .$$

This completes the proof of our theorem. □

§2. DISTANCE SEQUENCES OF GRAPHS IN $G(n,M)$

Recall that the *distance sequence* of a vertex x is $(d_i(x))$, where $d_i(x) = \{z \in V : d(z,x) = i\}$. Our aim is to show that in a certain range of M almost every random graph G_M is such that its vertices are uniquely determined by their distance sequences. The distance sequence of an isolated vertex is $0,0,\ldots$, so a graph with two isolated vertices does not have the property above. The next result shows that if M is not too large but it is large enough to ensure that almost no G_M contains two isolated vertices then almost every graph is such that its vertices are uniquely determined by the initial segments of their distance sequences.

Theorem 5. Suppose $\psi(n) \to \infty$ and

$$\tfrac{1}{2}n(\log n + \psi(n)) \le M \le n^{13/12} .$$

Then a.e. graph in $G(n,M)$ is such that for every pair of vertices (x,y) there is an $\ell \le \ell_o = 3\lceil(\log n/\log(M/n))^{\frac{1}{2}}\rceil$ with $d_\ell(x) \neq d_\ell(y)$.

Proof. For the sake of convenience we shall prove the analogous result for the model $G(n,P(\text{edge}) = p)$, where $p = 2M/n^2$. It will be clear that this does not change the essence of the proof only the calculations become a little more transparent.
Furthermore, the difficulties depend on the size of M: for M rather close to the lower bound we need a slightly different argument.

(a) Assume first that $pn/(\log n)^3 \to \infty$ and $p \le 2n^{-12/13}$. Pick two vertices in V, say x and y. It suffices to show that in $G(n,P(\text{edge}) = p)$ the probability of $d_i(x) = d_i(y)$ for all $i \le \ell_o = 3\lceil(\log n/\log(M/n))^{\frac{1}{2}}\rceil$ is $o(n^{-2})$. Set

$$\Gamma_\ell(x,y) = \{z : d(z,x) = d(z,y) = \ell\} ,$$

$$\Delta_\ell(x) = \{z : d(z,x) = \ell,\ d(z,y) > \ell\} ,$$

$$\Delta_\ell(y) = \{z : d(z,y) = \ell,\ d(z,x) > \ell\} ,$$

$$\delta_\ell(x) = |\Delta_\ell(x)| \quad \text{and} \quad \delta_\ell(y) = |\Delta_\ell(y)| .$$

Easy calculations show that in our range

$$(2pn)^{\ell_o} \le n .$$

By Lemmas 3 and 5 of [4] we find that for some function $\varepsilon = \varepsilon(n) \to 0$ with probability $1 - o(n^{-3})$ we have for all $\ell \le \ell_o$:

$$|\delta_\ell(x) - (pn)^\ell| \le \varepsilon(pn)^\ell \quad \text{and} \quad d_\ell(x) - \delta_\ell(x) \le \varepsilon(pn)^\ell . \qquad (16)$$

Similar relations hold for $\delta_\ell(y)$ and $d_\ell(y)$.

Now let $0 \le \ell \le \ell_0 - 1$ and suppose we are given the edges of a random graph $G_p \in G(n, P(\text{edge}) = p)$ spanned by the union of the set of vertices at distance at most $\ell+1$ from x and the set of vertices at distance at most ℓ from y. Denote by S the set of vertices not belonging to the union above and set $s = |S|$. [If (16) holds then clearly $s \ge n/(pn)$.]

What is the probability that $d_{\ell+1}(x) = d_{\ell+1}(y)$, conditional on the set of edges given so far? As the event $d_{\ell+1}(x) = d_{\ell+1}(y)$ conditional on the edges so far *determines* the number of vertices in S joined to some vertex in $\Delta_\ell(y)$, this probability is at most

$$\max_k b(k;\, s,\, 1-(1-p)^{\delta_\ell(y)}) ,$$

where $b(k;\, s,\gamma)$ denotes the kth term of the binomial distribution:

$$b(k;\, s,\gamma) = \binom{s}{k} \gamma^k (1-\gamma)^{s-k} .$$

Consequently the probability of $d_{\ell+1}(x) = d_{\ell+1}(y)$ conditional on (16) is at most

$$(pn)^{-(\ell+1)/2} .$$

Therefore the probability of $d_i(x) = d_i(y)$ for $i = 1,2,\dots,\ell_0$, is at most

$$o(n^{-3}) + \prod_{\ell=1}^{\ell_0} (pn)^{-\ell/2} = o(n^{-2}) ,$$

since

$$\prod_{\ell=1}^{\ell_0} (pn)^{\ell/2} \ge pn(pn)^{\ell_0^2/4} .$$

This completes the proof of the assertion in our range.

(b) Now assume that

$$\log n + \psi(n) \le pn \le (\log n)^4$$

for some function $\psi(n) \to \infty$. It is easily seen that under these assumptions a.e. random graph $G_p \in G(n;\, P(\text{edge}) = p)$ has the following properties:

(i) G_p has no isolated vertex,

(ii) every vertex has degree at most $3pn$,

(iii) for any two adjacent vertices there are at least $\frac{1}{2} \log n/\log\log n$ vertices adjacent to at least one of them,

(iv) for every path of length two there are at least $\frac{1}{2} \log n$ vertices adjacent to at least one of them,

(v) for every integer $\ell \le 20$, every set of ℓ vertices spans a subgraph of size at most ℓ.

Denote by A the event that a random graph G_p has the properties above. If A holds then we can proceed more or less as in part (a). Conditional on A, with probability at least $1 - n^{-3}$ we have

$$\sum_1^{\ell_o} (d_\ell(x) + d_\ell(y)) = o(n)$$

and

$$\min\{\delta_\ell(x), \delta_\ell(y)\} \geq (pn)^{\ell-3}$$

for all $\ell \leq \ell_o$. As in (a), these imply that, conditional on A, the probability of $d_i(x) = d_i(y)$ for all $i \leq \ell_o$ is at most

$$\prod_{i=1}^{\ell_o-3} (pn)^{i/2} = o(pn)^{5\ell_o^2/21} = o(n^{-2}) .$$

Hence the probability that for *some* pair of vertices (x,y) we have $d_i(x) = d_i(y)$ for all $i \leq \ell_o$ is at most

$$1 - P(A) + n^2\, o(n^{-2}) = o(1) . \qquad \square$$

Remarks

1. Theorem 5 is essentially best possible. One can show that a.e. random graph $G_M \in G(n,M)$ contains two vertices x and y such that $d_i(x) = d_i(y)$ whenever $i \leq (\log n/\log(M/n))^{\frac{1}{2}}$. Furthermore, it is easily seen that if $M = \lfloor n^{11/8} \rfloor$, say, then a.e. G_M has diameter 3 and a.e. G_M contains vertices x and y with $d_i(x) = d_i(y)$ for all i. [In fact $d_i(x) = d_i(y) = 0$ for $i \geq 4$.] Thus Theorem 5 does not hold for large values of M, though the bound $n^{13/12}$ could easily be improved.

2. The proof shows that the vertices can be distinguished by other parts of the distance sequence as well. For example if M is in our range, ℓ_1 and ℓ_2 are natural numbers,

$$(4M/n)^{\ell_1+\ell_2} < n$$

and

$$(M/n)^{\ell_1\ell_2} > n^5$$

then almost no random graph G_M contains two vertices x and y with $d_i(x) = d_i(y)$ for all $i : \ell_1 \leq i \leq \ell_2$.

§3. DISTANCE SEQUENCES OF RANDOM REGULAR GRAPHS

The study of random regular graphs was started only recently. The main reason for this is that the asymptotic number of labelled r-regular graphs of order n was found only in 1978 by Bender and

Canfield [1] . Even more recently in [3] a simpler proof was given for the same formula; furthermore in [3] a model was given for the set of regular graphs with a fixed vertex set, which can be used to investigate random labelled regular graphs. In particular, Bollobás and de la Vega [5] studied the diameter of random regular graphs. Our approach here is fairly close to that of [3].

We shall give a brief description of the model mentioned above, in a form suitable for the study of distance sequences. Let $r \geq 3$ be fixed and denote by $G(n,r\text{-reg})$ the probability space of all r-regular graphs with a fixed set of n labelled vertices. We shall assume that rn is even and any two graphs have the same probability. As customary, we shall say that *almost every* (a.e.) r-regular graph has a certain property if the probability of having the property tends to 1 as $n \to \infty$. Let $W_1,W_2,\dots,W_n$ be disjoint r-element sets. A *configuration is a partition of* $W = \bigcup_1^n W_i$ *into unordered pairs*, called *edges*. If (a,b) is an edge of a configuration F then we say that a is *joined to* b . Denote by Ω the set of configurations and turn it into a probability space by giving all configurations the same probability. Given a configuration $F \in \Omega$, define $\phi(F)$ to be the multigraph having vertex $V = \{W_1,W_2,\dots,W_n\}$ in which W_i is joined to W_j if some element of W_i is joined to some element of W_j in F . Clearly every r-regular graph with vertex set V is of the form $\phi(F)$ for some $F \in \Omega$ and $|\phi^{-1}(G)| = (r!)^n$ for every r-regular graph with vertex set V . In fact, as shown in [3], for about $e^{-(r^2-1)/4}$ of all configurations F is $\phi(F)$ an r-regular graph. As an immediate consequence of this we see that a.e. r-regular graph has a certain property if and only if a.e. configuration has the corresponding property.

The property we are interested in is that of having two vertices with the same distance sequence. Let F be a configuration. In order to define the distance sequences of F we define the *distance* $d(W_i,W_j) = d_F(W_i,W_j)$ between two classes W_i,W_j as the minimal k for which there are classes $W_{i_0} = W_i, W_{i_1},\dots,W_{i_k} = W_j$ such that for every ℓ , $0 \leq \ell < k$, an edge of F joins an element of W_{i_ℓ} to an element of $W_{i_{\ell+1}}$. Equivalently, $d_F(W_i,W_j)$ is the distance between W_i and W_j in the graph $\phi(F)$. For $W_i \in V$ set $d_\ell(W_i) = \{W_j : d(W_i,W_j) = \ell\}$ and call $(d_\ell(W_i))_{\ell=1}^n$ the *distance sequence of* W_i . Thus W_i has the same distance sequence

in a configuration F as in the graph $\phi(F)$.

Theorem 6. Let $r \geq 3$ and $\varepsilon > 0$ be fixed. Set $\ell_o = \lfloor (\frac{1}{2}+\varepsilon)\frac{\log n}{\log(r-1)} \rfloor$. Then a.e. r-regular labelled graph of order n is such that every vertex x is uniquely determined by $(d_i(x))_1^{\ell_o}$.

Proof. Analogously to the proof of Theorem 5, our aim is to show that the probability of two fixed classes W_i and W_j of a random configuration satisfying

$$d_\ell(W_i) = d_\ell(W_j) , \qquad 1 \leq \ell \leq \ell_o ,$$

is $o(n^{-2})$. Let $1 \leq i < j \leq n$ be fixed integers. We shall select the edges of a random configuration one by one, taking those nearest to W_i and W_j first. This approach is closely modelled on that in Bollobás and de la Vega [5].

Suppose we have selected all edges at least one endvertex of which belongs belongs to a class at distance less than k from the set $\{W_i,W_j\}$. Take all classes at distance k from $\{W_i,W_j\}$ and one by one select pairs for those elements in these classes that have not been paired so far. Having done this consecutively for $k = 0,1,\ldots,n-2$ and $n-1$, we have constructed the union of the components of W_i and W_j in a random configuration. In fact it is rather trivial that a.e. configuration is connected so after this sequence of operations we almost surely end up with the entire random configuration.

As in [5], we call an edge *indispensable* if it is the first edge that ensures that another class W_ℓ is at a certain distance from $\{W_i,W_j\}$. Equivalently, an edge is dispensable if each class containing an endvertex of the edge is either one of W_i and W_j or else it contains a vertex incident with an edge we have already chosen. Note that an edge joining two vertices of the same class is dispensable. What is the probability that the kth edge we select is dispensable? As the $k-1$ edges selected so far are incident with vertices in at most $k+1$ classes, this probability is at most

$$\frac{(k+1)(r-1)}{(n-1)r} .$$

Therefore the probability that more than 2 of the first $k_o = \lfloor n^{1/6} \rfloor$ edges are dispensable is at most

$$\binom{k_o}{3}\left(\frac{k_o}{n-k_o}\right)^3 = o(n^{-2}) , \tag{17}$$

the probability that more than $\ell_1 = \lfloor n^{1/8} \rfloor$ of the first $k_1 = \lfloor n^{6/13} \rfloor$ edges are dispensable is at most

$$\binom{k_1}{\ell_1+1}\left(\frac{k_2}{n-k_2}\right)^{\ell_1+1} = o(n^{-2}) \tag{18}$$

and the probability that more than $\ell_2 = \lfloor n^{5/13} \rfloor$ of the first $k_2 = \lfloor n^{2/3} \rfloor$ edges are dispensable is at most

$$\binom{k_2}{\ell_2+1}\left(\frac{k_2}{n-k_2}\right)^{\ell_2+1} = o(n^{-2}) \ . \tag{19}$$

Let A be the event that at most 2 of the first k_o edges are dispensable, at most ℓ_1 of the first k_1 and at most ℓ_2 of the first k_2. By (17)-(19) we find that the probability of A is $1 - o(n^{-2})$. Therefore it suffices to show that the probability of $d_\ell(W_i) = d_\ell(W_j)$ for $1 \le \ell \le \ell_o$ *conditional on* A is $o(n^{-2})$. In order to estimate this conditional probability we describe another, slightly different way of selecting the edges of a configuration. As before, suppose we have selected all edges incident with vertices in classes at distance less than k from $\{W_i, W_j\}$. Partition the elements that are unpaired at this stage into edges and put an edge into our random configuration under construction if it satisfies one of the following three conditions. (i) One of the vertices incident with the edge belongs to a class at distance k from W_i, (ii) both vertices incident with the edge belong to classes at distance k from W_j, (iii) if we add the edges satisfying (i) then one endvertex of our edge belongs to a class at distance k from W_j and the other belongs to a class at distance $k+1$ from W_i. We call this the $(2k)$*th operation*.

Suppose that after the completion of the $(2k)$th operation there are t_k vertices in the classes at distance k from W_j that have not been paired so far and there are s_k classes consisting entirely of unpaired elements. The $(2k+1)$*st operation* consists of pairing the t_k vertices above with t_k elements of the s_k classes and adding all these pairs to our random configuration.

Having completed the $(2k+1)$st operation, we have determined all edges incident with vertices in classes at distance at most k from $\{W_i, W_j\}$ so we can proceed to the $(2k+2)$nd operation. Once again, if we perform consecutively the 0th, 1st, 2nd, ..., $(2n)$th operations, then we construct all edges in the components containing W_i and W_j.

We may assume that the ε appearing in the statement of the theorem satisfies $0 < \varepsilon < 1/8$. Note that then

$$2 + r(r-1)^{\ell_o} \le 2 + rn^{\frac{1}{2}+\varepsilon} < n^{2/3} .$$

Now let us assume that A holds. It is easily seen that then for $k \le \ell_o$ we have

$$t_k \ge (r-1)^{k-3} \quad \text{and} \quad s_k \ge n/2 . \tag{20}$$

(In fact both bounds are rather crude.) Note that $d_{k+1}(W_i)$ is determined before we begin the $(2k+1)$st operation so if $d_{k+1}(W_i) = d_{k+1}(W_j)$ has to hold then the pairs of the t_k vertices must belong to a given number of the s_k classes. Hence the probability that $d_{k+1}(W_i) = d_{k+1}(W_j)$ conditional on t_k and s_k satisfying (20) is at most the maximum of the probability that a t_k element subset of the union of s_k classes of r elements each meets exactly ℓ classes, the maximum being taken over all values of ℓ, t_k and r_k satisfying (20). The following lemma gives an upper bound for this maximum.

Lemma 7. Let $R = \bigcup_1^s R_i$ be a partitioning of an rs-element set into r-element sets, where $r \ge 3$ is fixed. Suppose $t \le cs^{5/8}$ for some constant c. Denote by X_T the number of R_i's intersected by a random t-subset T of R. Then

$$\max_\ell P(X_T = \ell) \le c_o s^{1/2}/t$$

for some constant c_o.

Proof. We may assume without loss of generality that $t \ge s^{\frac{1}{2}}$.

The probability that T has at least 3 elements in some R_i is at most

$$s\binom{r}{s}\binom{rs-3}{t-3}/\binom{rs}{t} \le \frac{1}{6}t^3/s^2 \le \frac{c^4}{6} s^{\frac{1}{2}}/t .$$

Since

$$P(X_T = \ell) \le P(\max_i |T \cap R_i| \ge 3) + P(X_T = \ell \text{ and } \max_i |T \cap R_i| \le 2) ,$$

the lemma follows if we show that

$$P_\ell = P(X_T = \ell \text{ and } \max_i |T \cap R_i| \le 2) \le c_1 s^{\frac{1}{2}}/t$$

for some constant c_1.

Clearly $P_\ell = 0$ if $\ell < t/2$ and otherwise

$$P_\ell = \binom{s}{\ell}\binom{\ell}{t-\ell} r^{2\ell-t}\binom{r}{2}^{t-\ell}/\binom{rs}{t} .$$

For $1 \le u \le (t-1)/2$ set

$$Q_u = \frac{(t-2u)(t-2u-1)}{(s-t+u+1)(u+1)} \frac{r-1}{2r} .$$

It is easily seen that if u is an integer then we have

$$P_{t-u-1}/P_{t-u} = Q_u .$$

Define u_o, $1 < u_o < t/2$, by

$$Q_{u_o} = 1 .$$

Then

$$u_o \sim \frac{t^2(r-1)}{2rs} = o(t^{\frac{1}{2}})$$

and P_{t-u} increases with u for $u \le u_o-1$ and decreases with u for $u \ge u_o-1$. Furthermore, easy calculations show that for $u_o - u_o^{\frac{1}{2}} \le u \le u_o-1$

$$Q_u \le 1 + c_2 u_o^{-\frac{1}{2}}$$

and for $u_o - 1 \le u \le u_o + u^{\frac{1}{2}}$

$$Q_u \ge 1 - c_2 u_o^{-\frac{1}{2}}$$

where c_2 is some constant. Hence

$$\min\{P_{t-u} : u_o - u_o^{\frac{1}{2}} \le u \le u_o + u_o^{\frac{1}{2}}\} \ge c_3 \max_{\ell} P_\ell \qquad (21)$$

for some positive constant c_3. Finally, by the definition of P_ℓ we have

$$\sum_{\ell} P_\ell \le 1 ,$$

so (21) implies that

$$\max_{\ell} P_\ell \le \frac{1}{c_3 u_o^{\frac{1}{2}}} \le c_u s^{\frac{1}{2}}/t$$

for some positive constant c_u. □

The proof of Theorem 6 is easily completed with the aid of Lemma 7. The probability that $d_\ell(W_i) = d_\ell(W_j)$ for $\ell \le \ell_o$ is at most

$$1 - P(A) + \prod_{h}^{\ell_o} \{c_o n^{\frac{1}{2}}/(r-1)^{\ell-3}\} ,$$

where $h = \lfloor \frac{1}{2} \frac{\log n}{\log(r-1)} \rfloor + 3$. Since

$$(r-1)^{\ell_o} \ge n^{(1+\varepsilon)/2} ,$$

the sum above is $o(n^{-2})$, completing the proof of our theorem. □

Remark. A slight variant of the proof above gives the following extension of the theorem:

Let $r \geq 3$ and $\frac{1}{2} < \delta < 1$ be fixed and set $h_0 = \lfloor \delta \frac{\log n}{\log(r-1)} \rfloor$, $h_1 = h_0 + \lceil \frac{6}{2\delta - 1} \rceil + 3$. Then a.e. r-regular labelled graph of order n is such that every vertex x is uniquely determined by $\{d_i(x) : h_0 \leq i \leq h_1\}$.

REFERENCE

[1] E.A. Bender and E.R. Canfield, The asymptotic number of labelled graphs with given degree sequences, J. Combinatorial Theory (A) 24 (1978) 296-307.

[2] B. Bollobás, Graph Theory - An Introductory Course, GTM vol.63, Springer-Verlag, New York - Heidelberg - Berlin, 1979.

[3] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, Europ. J. Combinatorics 1 (1980) 311-316.

[4] B. Bollobás, The diameter of random graphs, Trans. Amer. Math. Soc. to appear.

[5] B. Bollobás and W.F. de la Vega, The diameter of random regular graphs, to appear.

[6] P. Erdös and A. Rényi, On the evolution of random graphs, Publ. Math. Inst. Hungar. Acad. Sci. 5 (1960) 17-61.

[7] P. Erdös and A. Rényi, On the strength of connectedness of a random graph, Acta Math. Acad. Sci. Hungar. 12 (1961) 261-267.

[8] G.W. Ford and G.E. Uhlenbeck, Combinatorial problems in the theory of graphs, Proc. Nat. Acad. Sci. U.S.A. 43 (1957) 163-167.

[9] W. Oberschelp, Kombinatorische Anzahlbestimmungen in Relationen, Math. Ann. 174 (1967) 53-78.

[10] E.M. Wright, Graphs on unlabelled vertices with a given number of edges, Acta Math. 126 (1971) 1-9.