COMBINATORICA 7 (3) (1987) 265—274

LARGE HOLES IN SPARSE RANDOM GRAPHS

A. M. FRIEZE and B. JACKSON

Received 18 October 1985

We show that there is a function $\alpha(r)$ such that for each constant $r \geqq 3$, almost every r-regular graph on n vertices has a hole (vertex induced cycle) of size at least $\alpha(r)n$ as $n \to \infty$. We also show that there is a function $\beta(c)$ such that for $c > 0$ large enough, $G_{n,p}$, $p = c/n$ almost surely has a hole of size at least $\beta(c)n$ as $n \to \infty$.

Introduction

This paper is concerned with the existence of large *holes* i.e. induced cycles, in random graphs. We study in particular random regular graphs and sparse random graphs with constant average degree.

In a recent paper [5] we showed that such graphs have large induced trees with high probability. For the model of random graphs with constant average degree this confirmed a conjecture of Erdős and Palka [3]. This conjecture was confirmed independently and contemporaneously by Fernandez de la Vega [4]. Thus this paper strengthens the above results by specifying that the induced tree can be taken to be a path.

For a graph G let $\sigma(G)$ denote the size of its largest hole. Our first result concerns random regular graphs. Let $R(r, n)$ denote the set of regular graphs of degree r with vertex set $V_n = \{1, 2, ..., n\}$. We turn $R(r, n)$ into a probability space by giving each graph in $R(r, n)$ the same probability.

Theorem 1. *Let $r \geqq 3$ be fixed and let $R(r, n)$ be chosen randomly from $R(r, n)$. Then where $q = q(r) = (2r-3)(2r-4)$ we have*

$$\lim_{n\to\infty} \Pr\left(\sigma(R(r, n)) \geqq nr(r-2)^{-1}(1-q\log_e(1+q^{-1})-\varepsilon)\right) = 1$$

for any fixed $\varepsilon > 0$.

The second result concerns the model $G_{n,p}$, $p = c/n$ where c is a constant.

AMS subject classification (1980):

Here each of the $\binom{n}{2}$ possible edges is independently included with probability p and excluded with probability $1-p$.

We prove

Theorem 2. *For large enough* c, *and* $q=q(4c)$,

$$\lim_{n\to\infty} \Pr\left(\sigma(G_{n,c/n}) \geqq (n/4c)(1-c^6e^{-c})(1-q\log_e(1+q^{-1})-\varepsilon)\right) = 1$$

for any fixed $\varepsilon>0$.

The restriction, for large enough c' in the above theorem comes from our method of proof. We conjecture that a similar result holds for any $c>1$.

Regular graphs

We prove Theorem 1 by analysing a simple algorithm FINDPATH that searches for a long induced path. We show that FINDPATH almost surely finds a path of the stated length. (By almost surely, a.s., we simply mean with probability $\to 1$ as $n\to\infty$). It is then a simple matter to show there is a.s. a large hole.

FINDPATH keeps a path $P_k=(v_1, v_2, \ldots, v_k)$ such that P_{k-1} is induced. It checks as to whether or not P_k itself is induced and if so it tries to grow its path from v_k. It searches the currently available edges $A(v_k)$ which are incident with v_k and then proceeds as follows:

(1) $A(v_k)=\emptyset$ or there is an edge in $A(v_k)$ of the form (v_k, v_i), $i\leqq k-2)$. In this case FINDPATH backtracks i.e. removes vertices $v_k, v_{k-1}, \ldots$ from P until a v_i is found for which $A(v_i)\neq\emptyset$, otherwise

(2) FINDPATH chooses an edge $(v_k, w_1)\in A(v_k)$, extends P_k by making $v_{k+1}=w_1$ and removes edge (v_k, w_1) from $A(v_k)$.

More formally we define

Algorithm FINDPATH
Input graph $G=(V_n, E)$;
begin
  $k:=1$; $v_1:=1$; **for** $v=1$ **to** n **do** $A(v):=\{e\in E\colon v\in e\}$
  repeat
    $W:=\{v\neq v_k\colon e\in A(v_k)$ and $v\in e\}=\{w_1, w_2, \ldots, w_p\}$;
    if $W=\emptyset$ or $W\cap\{v_1, v_2, \ldots, v_{k-2}\}\neq\emptyset$ *then*
    begin
      for $e\in A(v_k)$ **do** remove (e); backtrack
    end else
    begin
      $v_{k+1}:=w_1$; remove $((v_k, w_1))$;
      $k:=k+1$
    end
  until forever
end;

```
procedure backtrack;
begin
    repeat
      k:=k-1; if k=1 then STOP;
    until A(v_k)≠∅
end;

procedure remove (e: edge);
begin
    let e=(x, y); A(x):=A(x)-{e}; A(y):=A(y)-{e}
end.
```

It is clear that $(v_1, v_2, \dots, v_{k-1})$ is always an induced path, any v_k that stops the path from being induced is immediately discarded.

We now turn to the performance of FINDPATH on random regular graphs. In order to obtain a result for $G_{n,c/n}$, c constant, we shall prove a slightly stronger result.

Thus let $\Delta \geqq 3$ be an integer constant and let $\underline{d}=d_1, d_2, \dots, d_n$ satisfy $3 \leqq d_i \leqq \Delta$ for $i=1, 2, \dots, n$. Let $\mathscr{G}(\underline{d})$ denote the set of simple graphs with vertex set v_n satisfying $d(i)=d_i$ (degree of vertex i) for $i=1, 2, \dots, n$. Thus graphs in $\mathscr{G}(\underline{d})$ have no loops or multiple edges. Assume that $\underline{d}$ is graphic i.e. $\mathscr{G}(\underline{d}) \neq \emptyset$.

We turn $\mathscr{G}(\underline{d})$ into a probability space by giving all members of $\mathscr{G}(\underline{d})$ the same probability $1/|\mathscr{G}(\underline{d})|$. Theorem 1 follows immediately from Lemma 1.

Lemma 1. *Let G be chosen randomly from $\mathscr{G}(\underline{d})$, Then*

$$\lim_{n\to\infty} \Pr\left(\sigma(G) \geqq (2m/\Delta)(\Delta-2)^{-1}(1-q\log_e(1+q^{-1})-\varepsilon)\right) = 1$$

where $q=q(\Delta)$ and $\varepsilon>0$ is arbitrary.

Proof. In order to study $\mathscr{G}(\underline{d})$ we consider the model defined in Bollobás [1]. Let $D_1, D_2, \dots, D_n$ be disjoint sets with $|D_i|=d_i$ and set

$$D = \bigcup_{i=1}^{n} D_i$$

and $2m=|D|$.

A *configuration* C is a partition of D into m pairs, the edges of C. Let $\mathbf{\Phi}$ be the set of all $N(2m)=(2m)!2^{-m}/m!$ configurations. Turn $\mathbf{\Phi}$ into a probability space by giving all members of $\mathbf{\Phi}$ the same probability. For $C \in \mathbf{\Phi}$ let $\Phi(C)$ be the *multigraph* with vertex set V_n in which i is joined to j whenever C has an edge with one end vertex in D_i and the other in D_j. Clearly $\mathscr{G}(\underline{d}) \subseteq \Phi(\mathbf{\Phi})$ and

$$|\Phi^{-1}(G)| = \prod_{i=1}^{n} d_i!$$

for every $G \in \mathscr{G}(\underline{d})$.

Let Q be a property of the graphs in $G(\underline{d})$ and let Q^* be a property of the configurations in $\mathbf{\Phi}$. Suppose these properties are such that for $G \in \mathscr{G}(\underline{d})$ and $C \in \Phi^{-1}(G)$ the configuration C has Q^* if and only if G has Q. All we shall need from [2] is that if almost every C has Q^* then almost every G has Q.

We shall thus be able to prove '90%' of the lemma if we can show that FINDPATH applied to a multigraph $\Phi(C)$, C chosen randomly from $\mathbf{\Phi}$, almost surely finds an induced path of the required size. It will then be straightforward to show that a large hole a.s. exists.

In the following analysis we do not assume that the C chosen randomly from $\mathbf{\Phi}$ is given to us completely. Instead we assume that C is constructed by an oracle and we have to ask the oracle to get information about C. Done this way we can easily handle the conditioning introduced by the knowledge that the path $(v_1, v_2, \ldots, v_{k-1})$ is induced.

We shall now give the version of FINDPATH applied to configurations in this context. At any stage $A(i) \subseteqq D_i$ is the set of points of D_i whose pairings in C are not completely known to us, $i=1, 2, \ldots, n$. Initially $A(i)=D_i$ for $i=1, 2, \ldots, n$. $F \subseteqq D$ is the set of free points, i.e. points associated with vertices that have never been on our path. For such vertices, $A(v) \subseteqq F$ although $A(v)$ need not equal D_v as vertices removed from further consideration may have had v as a neighbour.

$$L = \bigcup_{i=1}^{k-1} A(v_i)$$

is the set of unpaired points associated with vertices on the induced path.

In the following, t is an iteration count.

Algorithm FINDPATH (applied to configurations)
begin
  $k:=1$; $v_1:=1$; $F:=\{1, 2, \ldots, 2m\}-A(1)$; $L:=\emptyset$; $t:=1$
  repeat
   $x:=A(v_k)=\{x_1, x_2, \ldots, x_p\}$
  Case Analysis
   Ask oracle if any x_i is paired with anything in $L \cup X$.
   (a) Answer YES: Ask oracle what is paired with x_i, $i=1, 2, \ldots, p$.
    Answer y_i, $i=1, 2, \ldots, p$.
    for $i=1$ **to** p **do** remove (x_i, y_i);
     backtrack.
   (b) Answer NO: Ask oracle what is paired with x_1
    Answer y_1.
    Let $y_1 \in A(v)$; remove (x_1, y_1);
    $k:=k+1$; $v_k:=v$; $L:=L \cup (X-\{x_1\})$.
  end of case analysis
  $t=t+1$
  until forever
end;

procedure backtrack
begin
  repeat
   $k:=k-1$; **if** $k=1$ **then** STOP
  until $A(v_k) \neq \emptyset$
  $L:=L-A(v_k)$
end;

procedure remove (x, y);
begin
let $x \in A(v)$, $y \in A(w)$;
$A(v) := A(v) - \{x\}$; $A(w) := A(w) - \{y\}$;
$F := F - \{x, y\}$; $L := L - \{x, y\}$
end.

We note first that throughout the algorithm $|L| \leqq (k-1)(\Delta-2)+1$ as $|A(v_1)| \leqq \Delta-1$ and $|A(v_i)| \leqq \Delta-2$, $i=2, \ldots, k-1$ by construction. We use $|L|/(\Delta-2)$ as an estimate for $k-1$ in our analysis. The reason for this is that it is difficult to analyse the reductions in k caused by backtracking whereas $|L|$ decreases by at most $\Delta-2$.

An execution of FINDPATH is used to define a random sequence $s_1, s_2, \ldots, s_i, \ldots$ where $s_t=1$ means that case (b) with $p \geqq 2$ occurred during the t'th iteration of the algorithm and $s_t=0$ otherwise.

Let l_t denote $|L|$ at the beginning of iteration t, then

(1a) $$s_t = 1 \rightarrow l_{t+1} \geqq l_t+1$$

(1b) $$s_t = 0 \rightarrow l_{t+1} \geqq l_t-(2\Delta-3)$$

If case (b) occurs and $p=1$ then $l_{t+1}=l_t$. If case (a) occurs then $p \leqq \Delta-1$ at this stage and so remove (x_i, y_i), $i=1, 2, \ldots, p$ reduces $|L|$ by at most $\Delta-1$. Backtracking reduces $|L|$ by at most $\Delta-2$ more. So, if $S_t=s_1+\ldots+s_t$, then

(2) $$l_t \geqq t-(2\Delta-3)(t-S_t).$$

We now prove a lower bound for $\Pr(s_t=1)$, for $t \geqq 2$. Note that $\Pr(s_1=1) \geqq 1-\binom{\Delta}{2}/(2m-1)$. Assume then that at the start of iteration $t-1$ we have $l_{t-1}=l$, $|F|=f$, $f^*=\left|\{x\colon x \in A(v) \subseteqq F \text{ where } |A(v)| \geqq 3\}\right|$ and $|X|=p$ as usual. We consider the case where backtrack was not called at the previous iteration. It can be seen that if backtrack had been called then the probability that $s_t=1$ increases slightly (β below decreases while α is unchanged).

Since each $x \in L$ must be paired with an element of $F \cup X$ the number of ways of continuing our pairings is

$$\beta = (f+p)(f+p-1)\ldots(f+p-l+1)\,N(f+p-l)$$
$$= (f+p)!\Big/\left(\left(\frac{1}{2}f+\frac{1}{2}p-\frac{1}{2}l\right)!\,2^{\frac{1}{2}f+\frac{1}{2}p-\frac{1}{2}l}\right).$$

The number of completions that lead to $s_t=1$ is at least

$$\alpha = f^*(f-\Delta)\ldots(f-\Delta-l+1)(f-l-1)\ldots(f-l-p+1)\,N(f-l-p)$$
$$= f^*(f-\Delta)!(f-l-1)(f-l-2)\ldots(f-l-\Delta+1)\Big/\left(\frac{1}{2}f-\frac{1}{2}l-\frac{1}{2}p\right)!\,2^{\frac{1}{2}f-\frac{1}{2}l-\frac{1}{2}p}$$

To see this we associate with each $x \in A(v) \subseteqq F$ a set $B(x) \subseteqq F$ such that $A(v) \subseteqq B(x)$ and $|B(x)|=\Delta$. Our expression for α is the number of pairings in which

$x_1 \in X$ is paired with $y_1 \in A(v) \subseteq F$, $|A(v)| \geqq 3$; each $x \in L$ is paired with something in $F - B(y_1)$, $x_2, x_3, \ldots, x_p$ are paired with elements of F and the remaining elements are paired arbitrarily. Each such pairing yields $s_t = 1$.

Thus

$$\Pr(s_t = 1 \,|\, |X| = p) \geqq \alpha/\beta \geqq \frac{f^*}{f}\left(1 - \frac{l+p}{f}\right)^p \left(1 - \frac{l}{f-\Delta}\right)^{\Delta-1} \tag{3}$$

We now use the inequalities

$$l \leqq (t-2)(\Delta-2)+1 \tag{4}$$

$$f \geqq 2m - (t-1)\Delta \tag{5}$$

$$f^* \geqq f - 2\big((t-2)(\Delta-2)+1\big). \tag{6}$$

(4) and (5) are straightforward. To veryfi (6) let $f_i = |\{v\colon D(v) \subseteq F \text{ and } |A(v)| = i\}|$, $i = 1, 2$. Then $f_1 + f_2 \leqq (t-2)(\Delta-2)+1$ and $f^* = f - (f_1 + 2f_2)$ and (4c) follows.

Using (4)—(6) in (3) and $p \leqq \Delta - 1$ (for $t \neq 1$) we deduce that for $t < 2m/\Delta$

$$\Pr(s_t = 1) \geqq 1 - \frac{2\Delta(\Delta-2)(t-2) + \Delta^2 + 1}{2m - t\Delta} \tag{7}$$

given that at least t iterations are executed and regardless of the history of the algorithm.

Thus let $z_1, z_2, \ldots, z_t$, $t \leqq m/\Delta$ be independent Bernouilli random variables with

$$\Pr(z_t = 0) = \frac{2\Delta(\Delta-2)(t-2) + \Delta^2 + 1}{2m - t\Delta}$$

Let $Z_t = z_1 + z_2 + \ldots + z_t$. It follows that

$$\Pr(S_t \geqq a) \geqq \Pr(Z_t \geqq a) \quad \text{for any} \quad a \geqq 0. \tag{8}$$

Our aim now is to show that FINDPATH a.s. executes at least something close to $t^* = 2m/(q+1)\Delta$ iterations where $q = q(\Delta)$, and that a.s.,

$$l_{t^*} \geqq (1 - o(1))\frac{2m}{\Delta}\left(1 - q \log\left(1 + \frac{1}{q}\right)\right). \tag{9}$$

Let $\tau = \lceil n^{1/3} \rceil$ and note from (7) that

$$\Pr(S_\tau < \tau) = O(n^{-1/3}). \tag{10}$$

For $t > \tau$ we use the following result which is a simple corollary of Theorem 1 of Hoeffding [6]:

Let X_i, $0 \leqq X_i \leqq 1$, $i = 1, 2, \ldots, N$ *be independent random variables and let* $\mu = E(X_1 + X_2 + \ldots + X_N)/N$. *Then*

$$\Pr(|X_1 + X_2 + \ldots + X_N - N\mu| \geqq \varepsilon\mu) \leqq 2e^{-\varepsilon^2 N\mu/3}. \tag{11}$$

Now for $t \geqq \tau$

$$\mu_t = E(z_1+z_2+\ldots+z_t) \tag{12}$$

$$= \left(1+\left(\frac{1}{n}\right)\right)\sum_{i=1}^{t}\left(1-\frac{2\Delta(\Delta-2)i}{2m-\Delta i}\right)$$

$$= (1-o(1))\int_0^t\left(1-\frac{2\Delta(\Delta-2)x}{2m-\Delta x}\right)dx$$

$$= (1-o(1))\left(t+(2\Delta-4)\left(t-\frac{2m}{\Delta}\log\frac{2m}{2m-\Delta t}\right)\right)$$

$$\geqq (1-o(1))\left(t+(2\Delta-4)\left(t-\frac{2mt}{2m-\Delta t}\right)\right) \quad (\text{using } x \geqq \log(1+x))$$

$$\geqq (1-o(1))t\frac{2\Delta-4}{2\Delta-3} \qquad \text{for} \quad t \leqq t^*$$

$$\geqq t/2 \qquad \text{for large } n.$$

Using (8) and (11) we deduce that for $\varepsilon > 0$ and small

$$\Pr(\exists t, \tau \leqq t \leqq t^*: S_t \leqq (1-\varepsilon)\mu_t) \leqq \Pr(\exists t, \tau \leqq t \leqq t^*: Z_t \leqq (1-\varepsilon)\mu_t)$$

$$\leqq \sum_{t=\tau}^{t^*} e^{-\varepsilon^2 t/6} = O(e^{-\varepsilon^2\tau/6}).$$

Hence, using (2)

$$\Pr(\exists t, \tau \leqq t \leqq t^*: l_t < t-(t-(1-\varepsilon)\mu_t)(2\Delta-3)) = O(e^{-\varepsilon^2\tau/6}). \tag{13}$$

Now for n large

$$t-(2\Delta-3)(t-(1-\varepsilon)\mu_t) \geqq t-(2\Delta-3)\left[t-(1-2\varepsilon)\left(t+(2\Delta-4)\left(t-\frac{2m}{\Delta}\log\frac{2m}{2m-\Delta t}\right)\right)\right]$$

$$\geqq t-2\varepsilon(2\Delta-3)t+\left(t-\frac{2m}{\Delta}\log\frac{2m}{2m-\Delta t}\right)q \tag{14}$$

$$\geqq t\left(1-2\varepsilon(2\Delta-3)-\frac{q\Delta t}{2m-\Delta t}\right) \quad \text{using} \quad \log(1+x) \leqq x \tag{15}$$

$$\geqq t\left(1-2\varepsilon(2\Delta-3)-\frac{q}{q+\varepsilon'}\right) \qquad \text{for} \quad t \leqq t' = \frac{2m}{(q+1+\varepsilon')\Delta}$$

$$\geqq \varepsilon t \qquad \text{if} \quad \varepsilon' = (q+1)(4\Delta-5)\varepsilon \tag{16}$$

We deduce from (10) and (16) that FINDPATH a.s. executes t' iterations. However, since ε can be made arbitrarily small and the RHS of (14) is a continuous function of t, which is strictly positive at $t^{*\prime}$ we can extend the inequality (16) to t^*

and deduce that t^* iterations a.s. take place. Substituting $t=t^*$ into (14) and using (13), yields (9).

In what follows $\varepsilon>0$ is arbitrarily small. For $u>\tau$ let $E_1(u)$ denote the event:

FINDPATH continues for at least u iterations, $S_\tau=\tau$ and

$$l_t \geqq t-(t-(1-\varepsilon)\mu_t)(2\Delta-3) \quad \text{for} \quad \tau \leqq t \leqq u.$$

Let $E_1=E_1(t^*)$. We have shown that $\Pr(E_1)=1-o(1)$. Let $t_1=\lfloor(1-\varepsilon)t^*\rfloor$ and $L_1(t)=\{j<\varepsilon n\colon\ A(v_j)\neq\emptyset$ at the start of iteration $t\}$.

If E_1 occurs then using (9) we have $l_t\geqq(1-q\log(1+q^{-1})-a\varepsilon)(2m/\Delta)>\varepsilon n$ for $t_1\leqq t\leqq t^*$ where $a>0$ is a constant.

Let $E_2(t)$ denote the event: some $x\in A(v_k)$ is paired with some $y\in A(v_i)$ for some $i\in L_1(t)$ and no element of $A(v_k)-\{x\}$ is paired with an element of L. Let

$$E_2 = \bigcup_{t=t_1}^{t^*} E_2(t).$$

If $E_1\cap E_2$ occurs then $\sigma(\Phi(C))\geqq(1-q\log(1+q^{-1})-b\varepsilon)(2m/\Delta)(\Delta-2)^{-1}$ where $b>0$ is a constant, and so we can prove the lemma by showing

$$\Pr(E_1\cap E_2) = 1-o(1).$$

Next let E_3 denote the event: $|L_1(t)|\geqq\varepsilon e^{-1/q}n/2$, for $t_1\leqq t\leqq t^*$. Then arguing similarly to the way we argued for (5) we can show

$$\Pr\Big(E_2(t)\,|\,E_1(t)\cap E_3\cap \bigcap_{k=t_1}^{t-1} \bar{E}_2(k)\Big) \geqq \frac{|L_1(t)|}{2m}\left(1-\frac{\Delta(\Delta-2)t^*}{2m}\right)\left(1-O\left(\frac{1}{n}\right)\right)$$

$$\geqq \alpha(\varepsilon,\Delta) \quad \text{for } n \text{ large}.$$

Where $\alpha(\varepsilon,\Delta)$ is strictly positive and independent of n. Hence, where $\theta=1-\alpha(\varepsilon,\Delta)$ we have

$$\Pr\Big(E_1(t)\cap E_3\cap \bigcap_{k=t_1}^{t} \bar{E}_2(k)\Big) \leqq \theta\Pr\Big(E_1(t)\cap E_3\cap \bigcap_{k=t_1}^{t-1} \bar{E}_2(k)\Big)$$

$$\leqq \theta\Pr\Big(E_1(t-1)\cap E_3\cap \bigcap_{k=t_1}^{t-1} \bar{E}_2(k)\Big)$$

and hence

$$\Pr(E_1\cap E_3\cap \bar{E}_2) \leqq \theta^{\varepsilon t^*}.$$

As $\Pr(E_1)=1-o(1)$, we can complete the proof of the lemma by showing that

$$\Pr(E_3) = 1-o(1). \tag{17}$$

Let now $t_2=\lceil(1+\varepsilon/(\Delta-2))\varepsilon n\rceil$. We find after some simplification using (15), that, for n large

$$t_2-(t_2-(1-\varepsilon)\mu_t)(2\Delta-3) \geqq (1-a\varepsilon)t_2,$$

where $a>0$ is constant.

We thus deduce that if $E_1(t_2)$ occurs then

$$|L(t_2)| \geqq (1-a\varepsilon)\,t_2.$$

Now at time t_2 there are at most $(\Delta-2)(t_2-\varepsilon n)\leqq\varepsilon^2 n$ points in $L(t_2)-L_1(t_2)$ and hence if $E_1(t_2)$ occurs

(18) $$|L_1(t_2)| \geqq \big(1-(a+1)\,\varepsilon\big)\,n.$$

Now fix a t, $t_1\leqq t\leqq t^*$. For $x\in L_1(t_2)$ we now, estimate a lower bound for the probability that $x\in L_1(t)$ which is independent of which other members of $L_1(t_2)$ are in $L_1(t)$. In fact

$$\Pr\big(x\in L_1(t)\big) \geqq \prod_{u=t_1+1}^{t}\left(1-\frac{\Delta-1}{2m-\Delta u}\right)$$

$((\Delta-1)/(2m-\Delta u))$ is an upper bound to the probability that x is paired with an element of $A(v_k)$ by the oracle in iteration u).

(19) $$\geqq \left(1-\frac{\Delta-1}{2m-\left(\dfrac{2m}{q+1}\right)}\right)^{\frac{2m}{(q+1)\Delta}}$$

$$\geqq e^{-1/q} \qquad \text{for large } n.$$

Now (18) and (19) imply (17), and the lemma, since then $|L_1(t)|$ dominates probabilistically a binomial random variable with parameters $\varepsilon\big(1-(a+1)\varepsilon\big)n$ and $e^{-1/q}$.

Sparse graphs

Bollobás, Fenner and Frieze [3] proved the following result: for sufficiently large c, $G_{n,c/n}$ a.s. contains a vertex induced subgraph H such that, after relabelling the vertices of H as $\{1, 2, \ldots, h\}$.

(a) $h\geqq n(1-c^6e^{-c})$.

(b) $6\leqq\delta(H)<\Delta(H)\leqq 4c$.

(c) Conditional on the degree of vertex i in H being d_i, $i=1, 2, \ldots, h$, H is equally likely to be any graph with such a degree sequence.

Theorem 2 follows from Lemma 1 and the above.

Note added in proof. Tomász Luczak has proved the conjecture given after the statement of Theorem 2.

References

[1] B. Bollobás, A probablistic proof of an asymptotic formula for the number of labelled regular graphs, *European Journal of Combinatorics* **1** (1980), 311—316.

[2] B. Bollobás, T. I. Fenner and A. M. Frieze, Long cycles in sparse random graphs, *Graph Theory and Combinatorics, A volume in honour of Paul Erdős*, (ed. B. Bollobás), Academic Press, London (1984), 59—64.

[3] P. Erdős and Z. Palka, Trees in random graphs, *Discrete Mathematics* **46** (1983), 145—150.

[4] W. Fernandez de la Vega, Induced trees in sparse random graphs, *to appear.*
[5] A. M. Frieze and B. Jackson, Large induced trees in sparse random graphs, *Journal of Combinatorial* B **42** (1987), 181—195.
[6] W. Hoeffding, Probability inequalities for sums of bounded random variables, *Journal of the American Statistical Association* **58** (1963), 13—30.

A. M. Frieze

Queen Mary College
London University
England

and

Carnegie Mellon University
Pittsburg, USA

B. Jackson

Goldsmiths College
London University
England