ON LARGE MATCHINGS AND CYCLES IN SPARSE RANDOM GRAPHS

A.M. FRIEZE*

Department of Computer Science and Statistics, Queen Mary College (University of London),
London, E1 4NS, United Kingdom

Received 28 February 1984
Revised 5 September 1985

Let \( k \) be a fixed positive integer. A graph \( H \) has property \( M_k \) if it contains \( \lfloor \frac{1}{2}k \rfloor \) edge disjoint hamilton cycles plus a further edge disjoint matching which leaves at most one vertex isolated, if \( k \) is odd. Let \( p = c/n \), where \( c \) is a large enough constant. We show that \( G_{n,p} \) a.s. contains a vertex induced subgraph \( H_k \) with property \( M_k \) and such that \( |V(H_k)| = 1 - (1 + \varepsilon(c))c^{k-1}e^{-c} \) \((k - 1)!)n\), where \( \varepsilon(c) \to 0 \) as \( c \to \infty \). In particular this shows that for large \( c \), \( G_{n,p} \) a.s. contains a matching of size \( \frac{1}{2}(1 - (1 + \varepsilon(c))c^{-c})n \) \((k = 1)\) and a cycle of size \( (1 - (1 + \varepsilon(c))c^{-c})n \) \((k = 2)\).

1. Introduction

In this paper we study the size of the largest matching and cycle in random graphs with edge probability \( c/n \), where \( c \) is a large constant. We continue the analysis of Bollobás [2], Bollobás, Fenner and Frieze [3] and confirm the conjecture in the final paragraph of the latter paper.

We shall let \( G_{n,p} \) denote a random graph with vertex set \( V_n = \{1, 2, \ldots, n\} \) in which edges are chosen independently with probability \( p \). We say that \( G_{n,p} \) has a property \( Q \) almost surely (a.s.) if \( \lim_{n \to \infty} \Pr(G_{n,p} \in Q) = 1 \).

For \( c > 0 \) define \( \alpha(c), \beta(c) \) by

\[
\alpha(c) = \sup(\alpha \geq 0): G_{n,c/n} \text{ a.s. contains a matching of size at least } \frac{1}{2}cn.
\]

(1.1)

and

\[
\beta(c) = \sup(\beta \geq 0): G_{n,c/n} \text{ a.s. contains a cycle of size at least } \beta n.
\]

(1.2)

Our main result is an improved estimate of \( \beta(c) \).

In what follows \( p = c/n \) and \( \varepsilon_1(c), \varepsilon_2(c) \) are unspecified functions satisfying \( \lim_{c \to \infty} \varepsilon_i(c) = 0, i = 1, 2 \).

* Research carried out while the author was a visiting professor at Carnegie-Mellon University, Pittsburgh, U.S.A.
To prove (2.3) we observe that
\[
\text{Exp}(\{|v \in V_n: d_G(v) > 4 \log n\}|) = n \sum_{k > 4\log n} \binom{n - 1}{k} p^k (1 - p)^{n-k-1} \\
\leq n \sum_{k > 4\log n} \left(\frac{ce}{k}\right)^k = o(1)
\]
as \(ce \leq 3 \log n\).

Since the expectation of the number of cycles of length 3 or 4 is \(o(c^4)\) their contribution is easily absorbed into what follows.

Next let \(P_k = \{\text{paths of length } k \text{ in } G \text{ with small endpoints}\}\). Now clearly
\[
|W_k| \leq 2 |P_k| \quad \text{for } k = 1, 2, 3, 4. \tag{2.7}
\]
Furthermore
\[
\text{Exp}(|P_k|) = \binom{n}{2} p\lambda^2, \tag{2.8}
\]
where \(\lambda = BS(c/10 - 1, n - 2) \leq e^{-0.669c}\). Now
\[
\text{Exp}(|P_1|^2) = \text{Exp}(|P_1|) + \binom{n}{2} \binom{n-2}{2} p^2 \lambda_1 + 2(n-2)\binom{n}{2} p^2 \lambda_2,
\]
where
\[
\lambda_1 = \text{Pr}(\text{SMALL} \supseteq \{1, 2, 3, 4\} \setminus E(G) \supseteq \{1, 2\}) \cup \{3, 4\}) \\
\leq \text{Pr}(|N_G(1) \cap \{5, 6, \ldots, n\}| \leq c/10 - 1)^4 \\
\leq (\lambda(1 - p)^{-3})^4
\]
and
\[
\lambda_2 = \text{Pr}(\text{SMALL} \supseteq \{1, 2, 3\} \setminus E(G) \supseteq \{1, 2\}) \cup \{2, 3\}) \\
\leq (\lambda(1 - p)^{-1})^3.
\]
This gives
\[
\text{Var}(|P_1|) \leq ce^{-4c/3} n \quad \text{for } n \text{ large.} \tag{2.9}
\]
Similar calculations give
\[
|P_k| = \frac{1}{2}(1 + o(1))n^{k+1}p^k\lambda^2 \quad \text{for } k = 2, 3, 4. \tag{2.10}
\]

(2.4) now follows from (2.7), (2.8), (2.9) and (2.10).

To prove (2.5) we take \(c \geq 20(l + 1)\log(l + 1)\) and first consider \(S\) for which \(1 \leq s = |S| \leq n/(200e^3(l + 1)^3)\). Let \(T = S \cup N_G(S)\) and \(t = |T|\). If (2.5) does not hold for \(S\) then \(|T| \leq m_1 = [n/(200e^3(l + 1)^3)]\) and \(T\) contains at least \(m_2 = \frac{1}{2} \sum_{k > 4\log n} \binom{n-1}{k} p^k (1 - p)^{n-k-1} \leq e^{-3 \log n} \leq e^{-0.669c}\).
edges of $G$. The probability that such a $T$ exists is no more than

$$\sum_{i=1}^{m_1} \binom{n}{t} \binom{t}{2} p^{m_2} \leq \sum_{i=1}^{m_1} \left(\frac{ne^t}{t} \frac{r^2}{2m_2}\right)^{m_2} \leq \sum_{i=1}^{m_1} \left(\frac{ne^t}{t} \frac{10e(l+1)r}{n}\right)^r \leq \sum_{i=1}^{m_1} \left(\frac{100e^3(l+1)^2r^3}{n}\right)^i = o(1)$$

For $|S| \geq m_3 = \lceil n/(300e^3(l+1)^3) \rceil$ we can ignore the fact that the vertices of $S$ are large. Let $m_4 = \lceil n/2l \rceil$. The probability that such an $S$ exists violating (2.5) is no more than

$$\sum_{s=m_3}^{m_4} \binom{n}{s} \binom{n}{ls} (1-p)^{s(n-s)} \leq \sum_{s=m_3}^{m_4} \left(\frac{ne}{s} \frac{1}{ls}\right)^{ls} e^{-3cs^2} \leq \sum_{s=m_3}^{m_4} (300e^4(l+1)^3 e^{-8(l+1)\log(l+1)}(l+1)^{l+1} = o(1)$$

which proves (2.5).

the probability that (2.6) does not hold is not more than

$$\sum_{s=m_4}^{l_1} \binom{n}{s} BS\left(cs/3l, s(n-s)\right) \leq 2 \sum_{s=m_4}^{l_1} \left(\frac{ne}{s}\right)^s \left(\frac{3ls(n-s)e}{cs}\right)^{cs/3l} \left(\frac{c}{n}\right)^{cs/3l} e^{-cs/3} \leq 2 \sum_{s=m_4}^{l_1} \left(2le(3le)^{cs/3l} e^{-cs/3}\right)^s = o(1). \quad \square$$

The proofs of our theorems rely on the removal of a certain set of vertices. We must show that this set is not too large. The following lemma deals with part of this set.

**Lemma 2.2.** Let $X_0 = \text{small}$ and let the sequence of sets $X_1, X_2, \ldots, X_s$ be defined by

$$X_i = \left\{ u \in V_n : \left| N_G(u) \cap \bigcup_{j=0}^{i-1} X_j \right| \geq 2 \right\}$$

and let $s$ be the smallest $i \geq 1$ such that $X_i+1 = X_i$. Let $X = \bigcup_{i=1}^{s} X_i$, then

$$|X| \leq 2e^4 c^4 e^{-4c^3} n \text{ a.s.} \quad (2.11)$$

**Proof.** For $x \in X \cup X_0$ let $i(x) = \min\{i : x \in X_i\}$ and let $D(x) = (V(x), A(x))$ denote a digraph inductively constructed as follows: for $x \in X_0$, $D(x) = (\{x\}, \emptyset)$ and for $x \in X_0$ let $y_1, y_2$ be 2 distinct neighbours of $x$ satisfying $i(x) > i(y_1), i(y_2)$. 


Then
\[ D(x) = (V(y_1) \cup V(y_2) \cup \{x\}, A(y_1) \cup A(y_2) \cup \{(x, y_1), (x, y_2)\}) \]

Each \( D(x) \) is acyclic, (weakly) connected and satisfies

each \( v \in V(x) \) has outdegree 0 or 2 and \( x \) is the unique vertex of indegree 0.

Let
\[ k = \text{the number of vertices of outdegree 2} = |K(x)|, \]
where \( K(x) = S(x) - X_0 \),

and let
\[ l = \text{the number of vertices of outdegree 0} = |L(x)|, \]
where \( L(x) = S(x) \cap X_0 \).

It follows then that
\[ |A(x)| = 2k \quad (2.13a) \]
and we will show
\[ l \leq k + 1 \text{ and if } l = k + 1, \text{ then } D(x) \text{ is a binary tree rooted at } x. \quad (2.13b) \]

This is most easily proved by induction on \( k \). A digraph satisfying (2.12) has at least one vertex \( y \) whose outneighbours \( z_1, z_2 \) both have outdegree zero. Removing arcs \( (y, z_1) \) and \( (y, z_2) \) and any vertex which becomes isolated we obtain a smaller digraph satisfying (2.12).

We obtain from the above that we can associate with each \( x \in X \), a set \( V(x) \) of vertices and a partition of \( V(x) \) into \( K(x), L(x) \) satisfying

\[ x \neq x' \text{ implies } V(x) \neq V(x'); \]
\[ \text{if } k = |K(x)|, \text{ then } l = |L(x)|, \text{ then } 2 \leq l \leq k + 1; \quad (2.14a) \]
\[ L(x) \subseteq \text{small}; \quad (2.14b) \]
\[ G(x) = G[V(x)] \text{ is connected and has at least } 2k \text{ edges}; \quad (2.14c) \]
\[ \text{if } l = k + 1 \text{ and } G(x) \text{ has } 2k \text{ edges, then } G(x) \text{ is a tree with leaves } L(x). \quad (2.14d) \]

We estimate \([X_0 - X_0]\) by counting sets of vertices satisfying (2.14). For a given \( k, l, m \) let \( \lambda_{k,l,m} \) be the expected number of sets \( K, L \) with \(|K| = k, |L| = l \) satisfying (2.14) above, where \( G[K \cup L] \) has \( m \) edges. Then

\[
\lambda_{k,l,m} \leq \binom{n}{k} \binom{n}{l} \binom{k+l}{2} \binom{m}{l} BS(c/10, n-k-l) \]
\[
\leq \left( \frac{ne^k}{k} \right) \left( \frac{ne}{l} \right) \left( \frac{(k+l)^e}{2m} \right) \left( \frac{c}{n} \right)^m e^{-2e(l/3)}(1 - \frac{c}{n})^{-(l(k+l))} \]
\[
= \mu_{k,l,m}. \]
Now if \( c \leq 2 \log n, \ k, \ l \leq n^{1/3} \), then \( \mu_{k,l,m+1}/\mu_{k,l,m} \leq n^{-1/4} \) for \( n \) large. Thus
\[
\sum_{m=2k}^{k+1} \lambda_{k,l,m} \leq (1 + o(1))\mu_{k,l,2k}.
\] (2.15)

With the same bounds on \( c, \ k, \ l \) and with \( n \) large and \( l \leq k + 1 \) we have
\[
\mu_{k,l,2k} \leq 21n^{l-k}(e^{4c^2}k)^{\frac{k-1}{l}}e^{-2c/3}
\] (2.16)

which implies
\[
\sum_{l=2}^{k+1} \mu_{k,l,2k} \leq 21(e^{4c^2}k/n)^{k} \sum_{l=2}^{k+1} (n/le^{2c/3})^l \leq n(e^{4c^2})^{k}e^{-2ck/3} \leq ne^{-ck/2} \quad \text{as} \quad c \geq 300.
\]

It follows that \( s \leq \log n \) a.s., and we can assume \( k \leq \log n \). Now, using (2.16),
\[
\sum_{k=2}^{\log n} \sum_{l=2}^{k} \mu_{k,l,2k} \leq 21 \sum_{k=2}^{\log n} (e^{4c^2})^{k}e^{-2ck/3} \leq 22(e^{4c^2})^{k}e^{-4c/3}
\]
and so
\[
\text{the number of sets, } K, \ L \text{ with } 2 \leq l \leq k \text{ is a.s. less than } n^{1/2}e^{-4c/3}. \quad (2.17)
\]

We only need to consider the case \( l = k + 1 \) from now on. But as \( \mu_{k,k+1,m+1}/\mu_{k,k+1,m} \leq 3ck/n \) we have
\[
\sum_{m \geq 2k} \mu_{k,k+1,m} \leq (1 + o(1))\mu_{k,k+1,2k}.
\] (2.18)

So we are finally reduced to estimating
\[
\tau_k = \text{the number of vertex induced binary trees with } k \text{ leaves (k-b-trees) in which each leaf is small.}
\]

Let \( \theta_k \) be the number of (vertex labelled) \( k \)-b-trees contained in a complete graph with \( 2k - 1 \) vertices. (Clearly \( \theta_k \leq (2k - 1)^{2k-3} \). Then
\[
\text{Exp}(\tau_k) = \binom{n}{2k-1}\theta_k p^{2k-2}(1-p)^{(k+1)^2-2k+2} BS(c/10 - 1, n - 2k + 1)^k 
\]
\[
\leq n(e^{2c^2}e^{-2c/3})^k \quad \text{for } n \text{ large.} \quad (2.19)
\]

To estimate \( \text{Var}(\tau_k) \), let \( \{T_1, T_2, \ldots, T_B\}, \ B = (2k-1)\theta_k \), be the set of \( k - b \)-trees contained in a complete graph with \( n \) vertices. Let \( A_i \) be the event that \( T_i \) is a vertex induced subgraph of \( G_{n,p} \) in which all leaves are small.

Next let \( Y_p = \{(i,j) \mid |V(T_i) \cup V(T_j)| = p\} \) for \( p = 2k - 1, \ldots, 4k - 2 \) and let
\[
Z_{p,q} = \{(i,j) \in Y_p \mid |E(T_i) \cup E(T_j)| = q\}. \quad \text{Then}
\]
\[
\text{Exp}(\tau_k^2) = \text{Exp}(\tau_k) + \Delta_1 + \Delta_2,
\] (2.20)
where

\[ \Delta_1 = \sum_{(i,j) \in Y_{4k-2}} \Pr(A_i \cap A_j) \]

and

\[ \Delta_2 = \sum_{p=2k-1}^{4k-3} \sum_{(i,j) \in Y_p} \Pr(A_i \cap A_j). \]

Now

\[ \Delta_1 \leq \left( \frac{n}{2k-1} \right)^2 \left( \theta_k p^{2k-2} (1-p)^{(2k-1) - 2k+2} \right)^2 \sigma, \]

where

\[ \sigma = BS(c/10 - 1, n - 2k + 1)^k BS(c/10 - 1, n - 4k + 2)^k \]

is an estimate of the probability that all leaves of 2 particular disjoint trees are small. It follows that

\[ \Delta_1 \leq \text{Exp}(\tau_k)^2 (1-p)^{-2k^2}. \]

(2.21)

Now for \( p \leq 4k - 3 \) we have

\[ \sum_{(i,j) \in Y_p} \Pr(A_i \cap A_j) = \sum_{q=p-1}^{4k-4} \sum_{(i,j) \in Z_{p,q}} \Pr(A_i \cap A_j) \]

\[ \leq \sum_{q=p-1}^{4k-4} \binom{n}{p} \binom{q}{2k-1} \left( \frac{c}{n} \right)^q e^{-2ck/3} (1-p)^{-8k^2} \]

\[ \leq ne^{-ck/2} \text{ for } n \text{ large.} \]

(2.22)

(2.19), (2.20), (2.21), (2.22) plus the Chebycheff inequality implies that \( \tau_k \) is a.s. within a factor \((1 + o(1))\) of the right-hand side of (2.19). This together with (2.17) and (2.18) proves the result. \( \square \)

For a positive integer \( k \), the \( k \)-core \( V_k(G) \) is defined to be the largest set \( S \subseteq V_n \) such that \( \delta(G[S]) \geq k \). This is well defined, for if \( \delta(G[S_i]) \geq k \) for \( i = 1, 2 \), then \( \delta(G[S_1 \cup S_2]) \geq k \). We let \( G_k \) denote the subgraph of \( G \) induced by \( V_k(G) \).

The \( k \)-core can be constructed using the following algorithm.

\begin{verbatim}
begin
    H := G;
    while \( \delta(H) < k \) do
        begin
            Y := \{ v \in V(H) : d_H(v) < k \};
            H := H[V(H) - Y]
        end
end
\end{verbatim}

On termination \( H = G_k \). This is because one can easily show inductively that
each iteration removes vertices that are not in \( V_k(G) \) and as \( \delta(H) \geq k \) we have \( V(H) \subseteq V_k(G) \).

Clearly any matching of \( G \) is contained in \( G_1 \) (= \( G \) minus isolated vertices) and any cycle of \( G \) is contained in \( G_2 \).

Now for \( k \geq 1 \) let \( A_k = A_k(G_{n,p}) = V_k(G_{n,p}) - (W \cup X \cup Y_k) \), where \( W, X \) are as defined in Lemmas 2.1, 2.2 respectively and

\[
Y_k = \{ y \in V_p : d_{G_{n,p}}(y) = k \quad \text{and} \quad N_{G_{n,p}}(y) \cap X \neq \emptyset \}.
\]

Let \( H_k = H_k(G_{n,p}) = G_{n,p}[A_k] \), then we have

**Lemma 2.3.** For \( k \geq 1 \) let \( M \) be any matching of \( G_{n,p}[A_k] \) which is not incident with any small vertex. Let \( \hat{H}_k = H_k - M \), then for large \( c \)

\[
\emptyset \neq S \subseteq A_k, |S| \leq n/(2k + 8) \quad \text{implies} \quad |N_{\hat{H}_k}(S)| \geq k |S| \text{ a.s.} \tag{2.23}
\]

**Proof.** Let \( G = G_{n,p}, H = \hat{H}_k \) and for a given \( S \) let \( S_1 = S \cap \text{SMALL} \) and \( S_2 = S - S_1 \). Now

\[
|N_H(S)| \geq |N_H(S_1)| - |S_2| + |N_H(S_2)| - \min(|S_1|, |S_2|). \tag{2.24}
\]

This follows from \( S \cap (W \cup X) = \emptyset \).

Also, we claim

\[
|N_H(S_1)| \geq k |S_1|. \tag{2.25}
\]

Note first that \( v \in S_1 \) implies \( d_G(v) \geq k \) and no pair of vertices of \( S_1 \) are adjacent, since \( S_1 \cap W_1 = \emptyset \). Note that no pair of vertices of \( S_1 \) have a common neighbour as \( S_1 \cap W_2 = \emptyset \). Also \( N_G(S_1) \cap (W \cup Y_k) = \emptyset \) as \( S_1 \cap W_1 = \emptyset \). Furthermore \( v \in S_1 \) implies \( |N_G(v) \cap X| \leq 1 \) as \( S_1 \cap X = \emptyset \). Thus to prove (2.25) we need only show that if \( v \in S_1 \) and \( d_G(v) = k \), then \( N_G(v) \cap X = \emptyset \). But this follows from \( S_1 \cap Y_k = \emptyset \).

We claim next that if (2.5) holds with \( l = k + 4 \), then

\[
|N_H(S_2)| \geq (k + 2) |S_2|. \tag{2.26}
\]

For then \( |N_G(S_2)| \geq (k + 4) |S_2| \) and for each \( v \in S_2 \), \( |(N_G(v)) \cap |N_H(v)| \geq 2. \) This is because \( v \) is incident with at most one edge of \( M \) and is adjacent to at most one vertex of \( W \cup X \cup Y_k \). It is a simple matter to verify (2.23) from (2.24), (2.25) and (2.26). \( \square \)

**Lemma 2.4.**

\[
|A_k| \geq n \left( 1 - (1 + \varepsilon(c)) \frac{e^{k-1}}{(k-1)!} e^{-c} \right) \text{ a.s.,} \tag{2.27}
\]

where \( \varepsilon(c) \to 0 \) as \( c \to \infty \).
Proof. 

\[ |A_k| \geq |V_k(G)| - |W| - |X| - |Y_k - W \cup X|. \]

We show first that

\[ |Y_k - W \cup X| \leq |X|. \] (2.28)

For \( y \in Y_k - X \) there is, by definition, a unique \( x(y) \in X \) such that \( y \) is adjacent to \( x(y) \) in \( G \). Now for distinct \( y_1, y_2 \in Y_1 - W \) we have \( x(y_1) \neq x(y_2) \) else \( y_1 \in W_2 \) and (2.28) follows.

Now let \( Z_0 \) be the set of vertices of degree \( \leq k - 1 \) in \( G \) and let \( Z_1, Z_2, \ldots \) be the sequence of sets removed in each iteration of the \( k \)-core finding algorithm. Now, it is well known that

\[ |Z_0| = (1 - o(1))n \left( 1 - \sum_{i=0}^{k-1} \frac{c^{i+1}e^{-c}}{i!} \right) \text{ a.s.} \]

We show that

\[ Z_i \subseteq X \cup W_i \cup Y_k \quad (i = 1, 2, \ldots) \]

Thus assume inductively that \( Z_1, Z_2, \ldots, Z_{i-1} \subseteq X \cup W_i \cup Y_k \) for some \( i \geq 1 \) (true vacuously for \( i = 1 \)) and let \( T = \bigcup_{i=0}^{i-1} Z_i \). Then \( y \in Z_i \) implies \( d_G(y) \geq k \) but \( |N_G(y) - T| \leq k - 1 \).

Case 1. \( |N_G(y) \cap T| \geq 2 \)

By assumption \( T \subseteq X \cup \text{SMALL} \) and so \( y \in X \).

Case 2. \( |N_G(y) \cap T| = 1 \)

Then \( d_G(y) = k \) implies \( y \in X \cup W_i \cup Y_k \).

Hence \( |V_k(G)| \geq |Z_0| - |X \cup W_i \cup Y_k| \) and the lemma follows.

Lemma 2.5. Let \( c \) be large and \( G \) satisfy the conditions in Lemmas 2.1, 2.2 and 2.3. Let \( X \) be a \( t \)-factor of \( H_k \) where, \( t < k \). Then \( H = (A_k, E(A_k) - X) \) is connected.

Proof. If \( H \) is not connected, then there exists a nonempty \( S \subseteq A_k \) such that \( N_H(S) = \emptyset \). We show that this is not possible for \( c \) large enough. (2.23) implies that \( |S| \geq n/(2k + 8) \). (2.27) implies that, for \( c \) large, fewer than \( 2c^{k-1}e^{-c}n \) vertices are deleted from \( G \) in producing \( H \). Then (2.2) implies that at most \( 8c^k e^{-c}n \) edges are lost in the construction. But then (2.6) with \( l = k + 4 \) implies that not all edges with one vertex in \( S \) have been deleted.

Suppose a graph \( G \) contains \( h \) edge-disjoint hamilton cycles. Let the graph obtained from \( G \) by deleting the edges in these cycles be referred to as an \( h \)-subgraph of \( G \).
Define $\phi(G) = (h, p)$ by

$$
\begin{align*}
    h &= \text{maximum number of disjoint hamilton cycles in } G; \\
    p &= \begin{cases} \\
        0 & \text{if } k \leq 2h \\
        \text{maximum cardinality of a matching} & \text{if } k = 2h + 1 \\
        \text{in any } h\text{-subgraph of } G & \text{if } k \geq 2h + 2 \\
        \text{maximum length of a path} & \text{in any } h\text{-subgraph of } G
    \end{cases}
\end{align*}
$$

If $\phi(G) = (h, p)$ we define a $\phi$-subgraph $H$ of $G$ to be any $h$-subgraph of $G$ containing either a matching of size $p$ or a path of length $p$ as the case may be. Let the edges in $E(G) - E(H)$ be referred to as a $\phi$-set.

**Lemma 2.6.** Let $H$ be a graph which cannot be disconnected by the removal of a $t$-factor, $t < k$. Suppose that $H$ does not have property $M_k$. Then there exists $U = \{u_1, u_2, \ldots, u_t\} \subseteq V(H)$ and for each $u_i \in U$, a set $U_i \subseteq V(H)$ such that

(i) $u_i \in U$, $w \in U_i$ implies $(u_i, w) \notin E(H)$ and $\phi(\hat{H}) > \gamma(H)$ (in the lexicographic ordering), where $\hat{H}$ is obtained from $H$ by adding the edge $(u_i, w)$.

(ii) $|N_H(U_i)| < k |U_i|$, $i = 1, 2, \ldots, t$.

**Proof.** Let $(h, p) = \phi(H)$ and $H'$ be a $\phi$-subgraph of $H$. We deduce that $H'$ is connected.

**Case 1.** $h < \lceil \frac{1}{2}k \rceil$

Let $U = \{u_1, u_2, \ldots, u_t\}$ be the set of vertices which are endpoints of longest paths of $H'$. Posa [12] has shown that for each $u_i \in U$ there exists a set $U_i \subseteq U$ such that

(a) for each $w \in U_i$ there is a longest path in $H'$ with endpoints $u_i$, $w$;

(b) $|N_{H'}(U_i)| < 2 |U_i|$.

Since $H'$ is connected and non-hamiltonian no edge joins the endpoints of any longest path. Adding such an edge must increase $\phi$ (in the lexicographic sense).

**Case 2.** $h = \lceil \frac{1}{2}k \rceil$, $k$ odd

Let $\mathcal{M}$ be the set of maximum cardinality matchings of $H$. Let $U = \{u_1, u_2, \ldots, u_t\}$ be the set of vertices left isolated by some $M \in \mathcal{M}$.

Let $u_i \in U$ and let some $M_i \in \mathcal{M}$ leave $u_i$ isolated. Let $S_i \neq \emptyset$ be the set of vertices, different from $u_i$, left isolated by $M_i$. Let $U'_i$ be the set of vertices reachable from $S_i$ by an even length alternating path w.r.t. $M_i$. Let $U_i = S_i \cup U'_i \subseteq U$. It is clear that (1) holds.

If $u \in N_H(U_i)$, then $u \notin S_i$ and so there exists $y_i$ such that $\{u, y_i\} \in M_i$. We show that $y_i \in U_i$ which will prove that $|N_{H'}(U_i)| < |U_i|$ and the lemma. Now there exists $y_2 \in U_i$ such that $\{u, y_2\} \in E(H)$. Let $P$ be an even length alternating path from some $s \in S_i$ terminating at $y_2$. If $P$ contains $\{u, y_1\}$ we can truncate it to terminate with $\{u, y_1\}$, otherwise we can extend it using edges $\{y_2, x\}$ and $\{x, y_1\}$.
We are now ready for the

**Proof of Theorem 1.3.** We use a coloring argument that was introduced in Fenner and Frieze [6]. Suppose that after generating $G = G_{n,p}$ all its edges are colored blue, and then each edge of $G$ is re-colored green with probability $p' = (\log n)/cn$ and left blue with probability $1 - p'$. These recolorings are done independently of each other.

Let $E^b, E^g$ denote the blue and green edges respectively and let $G^b = (V_n, E^b)$, $H_k = H_k(G)$ and $H^b_k = H_k(G^b)$.

**Remark 2.7.** It is important to note that for a fixed value of $E^b$, $E^g$ is a random subset of $E^b$, where each $e \in E^b$ is independently included in $E^g$ with probability $p_1 = pp'/(1 - p(1 - p'))$ and excluded with probability $1 - p_1$.

Consider next the following 2 events:

$i = G = G_{n,p}$ satisfies the conditions of Lemmas 2.1, 2.2, 2.3 and

$$\phi(H_k) < ([\frac{1}{2}k], (\frac{1}{2}a)(k - 2[\frac{1}{2}k])), \quad \text{where } a = |A_k(G)|.$$  

(2.29)

To prove (2.29) we shall prove that for $c$ large

$$\Pr(\mathcal{G} \mid i \not\subset A_k(G^b), |S| \leq n/(2k + 8) \implies |N_{H_k^b}(S)| \geq k |S|);$$

(b) there does not exist $e = \{v, w\} \in E^g$, $e \subseteq A_k(G^b)$ such that $\phi(H_k^b + e) > \phi(H_k^b)$.

In consequence of what has already been proved, we need only prove

$$\lim_{n \to \infty} \Pr(\mathcal{G}) = 0.$$  

(2.30a)

which together imply (2.29).

**Proof of (2.30a).** Let $G_0 \in \mathcal{G}$ be fixed and let $F_0$ be any fixed $\phi$-set of $H_k$. We prove

$$\Pr(\mathcal{G} \mid G_{n,p} = G_0) \geq (1 - 2c^{kn})^{(1 - p')^{kn}}.$$  

(2.31)

We can readily verify this once we have shown that

$$\mathcal{G} \supseteq \mathcal{G}_1 \cap \mathcal{G}_2 \cap \mathcal{G}_3 \cap \mathcal{G},$$

where

$$\mathcal{G}_1 = E^g$$

is a matching of $G_0$;

$$\mathcal{G}_2 = \text{no green edge meets any vertex of degree less than } c/10 + 2 \text{ in } G_0 \text{ or any vertex in } W \cup X \cup Y_k;$$

$$\mathcal{G}_3 = F_0 \cap E^g = \emptyset.$$
For $\mathscr{E}_1 \cap \mathscr{E}_2$ implies
\[ A_k(G_0^b) = A_k(G_0) \]
and then $\mathscr{E}_1$ implies (see Lemma 2.3) that (2.23) holds, which verifies $\mathscr{E}(a)$. $\mathscr{E}_3$ implies $\mathscr{E}(b)$.

Now it follows from (2.3) that
\[ \Pr(\mathscr{E}_1) \leq 16(\log n)^4/c^2 n. \]

From Lemmas 2.1, 2.2 and (2.27) we find that the total number of edges of $G_0$ that are excluded by the conditions in $\mathscr{E}_2$, $\mathscr{E}_3$ is no more than
\[ n((c/10 + 1)e^{-2c^3} + 4c^ke^{-c}) + \frac{1}{2}kn \leq kn \]
Thus
\[ \Pr(\mathscr{E}_1 \cup \mathscr{E}_2 \cup \mathscr{E}_3) \leq 1 - (1 - p')^kn + 16(\log n)^4/c^2 n, \]
which proves (2.31). □

**Proof of (2.30b).** Now
\[ \Pr(\mathscr{E}) = \sum_I \Pr(\mathscr{E} | G^b = I) \Pr(G^b = I), \]
where $I$ is an arbitrary graph with vertices $V_n$.

Now if $H_k(I)$ fails to satisfy $\mathscr{E}(a)$, then $\Pr(\mathscr{E} | G^b = I) = 0$. So let us assume that $\mathscr{E}(a)$ holds.

Now if $U, U_1, \ldots, U_t$ are as defined in Lemma 2.6 with $H = H_k$, then each set is of size at least $n/(2k + 8)$ and for $\mathscr{E}(b)$ to hold no green edge can join $u_i \in U$ to $w \in U_i$. But then in view of Remark 2.7 and $\mathscr{E}(a)$ we have
\[ \Pr(\mathscr{E}(b) | G^b = I) \leq (1 - p_i)^{n^2/(2(2k + 8)^2)}, \]
which implies (2.30b). □

Finally, let us consider what happens when $c \to \infty$. The above proof shows that $H_k$ a.s. has property $M_k$. For $k = 1$ and $c = \log n + x$, $x$ constant, one can easily show that $A_1$ a.s. comprises all non-isolated vertices of $G$. Thus we obtain Erdős and Renyi's result [5] as a corollary. Similarly, when $k = 2$ and $c = \log n + \log \log n + x$, $A_2$ a.s. comprises all vertices of degree at least 2 and so we obtain Komlós and Szemerédi's result [9] as well. (Tomasz Luczak pointed out an error in an earlier statement of these last two results). □

Corollary 1.2 follows directly from Theorem 1.2 and the Percolation Theorem of McDiarmid [11].

**References**