ON LARGE MATCHINGS AND CYCLES IN SPARSE RANDOM GRAPHS

CARNegie-MELLON UNIV Pittsburgh PA MANAGEMENT

SCiENCES RESEARCH GROUP A M FriEZe JAN 84 MSRR-504

UNCLASSIFIED N80014-75-C-0621

F/G 12/1
ON LARGE MATCHINGS AND
CYCLES IN SPARSE RANDOM GRAPHS
by
A. M. Frieze
January 1984
Contract N00014-75-C-0621

Carnegie-Mellon University
PITTSBURGH, PENNSYLVANIA 15213

GRADUATE SCHOOL OF INDUSTRIAL ADMINISTRATION
WILLIAM LARIMER MELLON, FOUNDER

DISTRIBUTION STATEMENT A
Approved for public release, Distribution Unlimited

84 09 14 046
ON LARGE MATCHINGS AND
CYCLES IN SPARSE RANDOM GRAPHS

by

A. M. Frieze*

January 1984

Contract N00014-75-C-0621

* Graduate School of Industrial Administration, Carnegie-Mellon
University, Pittsburgh, PA 15213, U.S.A.
(On leave from Queen Mary College, London)

This report was prepared as part of the activities of the Management
Sciences Research Group, Carnegie-Mellon University. Reproduction in
whole or in part is permitted for any purpose of the U.S. Government.

Management Sciences Research Group
Graduate School of Industrial Administration
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213
Abstract

Let \( p = \frac{c}{n} \) where \( c \) is a large constant. We show that the random graph \( G_n, p \) a.s. contains a matching of size \( n(1 - (1+\varepsilon(c))e^{-c})/2 \) and a cycle of size \( n(1-(1+\varepsilon(c))ce^{-c}) \) where \( \varepsilon(c) \) is some function satisfying \( \lim_{c \to \infty} \varepsilon(c) = 0 \).
1. 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

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

and

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

Our main result is an improved estimate of \(\beta(c)\). However the same methods can be used to estimate \(\alpha(c)\) and we shall do this first as the analysis is marginally simpler.

In what follows \(p = c/n\) and \(\epsilon_1(c), \epsilon_2(c)\) are unspecified functions satisfying \(\lim_{c \to \infty} \epsilon_i(c) = 0, \quad i=1,2\).

Theorem 1.1

\[
(1.3) \quad \alpha(c) = 1 - (1 + \epsilon_1(c))e^{-c}
\]

and this remains valid if \(c \to \infty\).
As far as we know the only other paper dealing with this question is by Karp and Sipser [7] who prove some strong results about a simple heuristic for finding a large cardinality matching.

There has been more work done on estimating $a(c)$. Ajtai, Komlós and Szemerédi [1] and Fernandez de la Vega [6] showed that $a(c) \geq 1 - c_0/c$. Bollobás made a significant step forward by showing that $G_{n,p}$ a.s. contains a large Hamiltonian subgraph and that $a(c) \geq 1 - c^{24}e^{-c/2}$. By refining this analysis, Bollobás, Fenner and Frieze [3] showed that $a(c) \geq 1 - c^6e^{-c}$. The main result of this paper is

**Theorem 1.2**

(1.4) \[ a(c) = 1 - (1 + \varepsilon_2(c))ce^{-c} \]

and this remains valid if $c = 0$.

**Corollary 1.3**

A random digraph with edge density $c/n$ a.s. contains a directed cycle of size $n(1 - (1 + \varepsilon_2(c))ce^{-c})$.

**Notation**

The following notation is used throughout. Let $G$ be a graph. $V(G)$, $E(G)$ denote the sets of vertices and edges of $G$.

For $S \subseteq V(G)$ we let $G[S] = (S,E(S))$ where $E(S) = \{ e \in E(G) : e \subseteq S \}$.

$N_G(S) = \{ w \in S : \text{there exists } v \in S \text{ such that } \{v,w\} \in E(G) \}$. 
For $v \in V(G)$ we write $N_G(v)$ for $N_G(\{v\})$ and $d_G(v)$ for the degree of $v$.

$\mu(G)$ is the maximum cardinality of a matching of $G$.

$$BS(x,m) = \sum_{k=0}^{\lfloor x \rfloor} \binom{m}{k} p^k (1-p)^{m-k}$$

As the case $c > \log n$ is well known we shall assume for convenience that $c \ll 3\log n$. 
2. **Lemma 2.1**

Let $G = G_{n,p}$ and let vertex $v$ be **small** if $d_G(v) \leq c/10$ and **large** otherwise. Let SMALL, LARGE be the sets of small and large vertices respectively.

Let $W = W_1 \cup W_2$ where for $k=1,2$

$W_k = \{ v : v \text{ is small and there exists a small } w \text{ such that } v \text{ and } w \text{ are joined by a path of length } k \}$

Then for $c \geq 300$ $G$ a.s. satisfies the following:

1. $(2.1) \quad |\{ v \in V_n : d_G(v) \leq c/10 + 1 \}| \leq ne^{-2c/3}$;
2. $(2.2) \quad$ there does not exist $S \subseteq V_n$ with $|S| \geq ne^{-c}$ and
   $$|\{ e \in E(G) : e \cap S \neq \emptyset \}| \geq 4c |S|;$$
3. $(2.3) \quad d_G(v) \leq 4\log n$ for $v \in V_n$;
4. $(2.4) \quad |W| \leq c^2e^{-4c/3}n;$
5. $(2.5) \quad \emptyset \neq S \subseteq V_n, |S| \leq n/14$ and $S \subseteq \text{LARGE}$ implies $|N_{G}(S)| \geq 6 |S| ;$
6. $(2.6) \quad S \subseteq V_n, n/14 \leq |S| \leq n/2$ implies
   $$|\{ (v,w) \in E(G) : v \in S, w \in S \} | \geq c |S|/10;$$

**Proof**

To prove (2.1) note that for $n$ large

$\text{Exp}( |\{ v \in V_n : d_G(v) \leq c/10 + 1 \} ) = nB5(c/10 + 1, n-1) \leq ne^{-0.669c}.$
Now the variance of this set size can be shown to be $\leq ne^{-2c/3}$.

Thus one can use either the Chebycheff or Markov inequality depending on whether or not c remains bounded as n tends to infinity.

Next note that the probability there exists a set $S$ violating (2.2) is no more than

$$\sum_{s \geq ne^{-c}} \binom{n}{s} \left(\frac{sn}{4cs}\right)^{4cs} \leq \sum_{s \geq ne^{-c}} \left(\frac{ne}{s}\right)^s \left(\frac{snep}{4cs}\right)^{4cs} \leq \sum_{s \geq ne^{-c}} \left(\frac{e^{s+1/c}}{256}\right)^{cs} = o(1).$$

To prove (2.3) we observe that

$$\exp(|\{v \in V_n : d_G(v) > 4\log n\}|) = n \sum_{k>4\log n}(n-1)p^k(1-p)^{n-k-1}$$

$$\leq n \sum_{k>4\log n}(ce^k)^k = o(1)$$

as $ce \leq 3\log n$.

Next let $P_k = \{\text{paths of length } k \text{ in } G \text{ with small endpoints} \}$. Now clearly

$$|W_k| \leq 2 |P_k| \quad \text{for } k=1,2.$$  

Furthermore

$$\exp(|P_1|) = \binom{n}{2}p^2\lambda_1^2$$

where $\lambda = \text{BS}(c/10 - 1, n-2) \leq e^{-0.669c}$

Now

$$\exp(|P_1|^2) = \exp(|P_1|) + \binom{n}{2}(n-2)p^2\lambda_1 + 2(n-2)\binom{n}{2}p^2\lambda_2$$

where
\begin{align*}
\lambda_1 &= \Pr(\text{SMALL}\supseteq \{1,2,3,4\}, \text{E}(G) \supseteq \{\{1,2\}, \{3,4\}\}) \\
&\leq \Pr(|\mathcal{N}_G(1) \cap \{5,6,\ldots,n\}| \leq c/10 - 1)^4 \\
&\leq (\lambda(1-p)^{-2})^4
\end{align*}
and

\begin{align*}
\lambda_2 &= \Pr(\text{SMALL}\supseteq \{1,2,3\}, \text{E}(G) \supseteq \{\{1,2\}, \{2,3\}\}) \\
&\leq (\lambda(1-p)^{-1})^3.
\end{align*}

This gives

\begin{align*}
\text{(2.9) } \text{Var}(|P_1|) &\leq ce^{-4c/3n} \quad \text{for } n \text{ large.}
\end{align*}

Similar calculations give

\begin{align*}
\text{(2.10a) } \text{Exp}(|E_2|) &= (1+o(1))n^3p^2\lambda^2/2 \\
\text{and}
\text{(2.10b) } \text{Var}(|E_2|) &\leq n^3p^2\lambda^2 \quad \text{for } n \text{ large}
\end{align*}

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

To prove (2.5) we first consider \( S \) for which \( 1 \leq s = |S| \leq n/35000e^4 \). Let \( T = S \cup \mathcal{N}_G(S) \) and \( t = |T| \). If (2.5) does not hold for \( S \) then \( |T| \leq m_1 = \lceil n/50000e^4 \rceil \) and \( T \) contains at least \( m_2 = \lceil ct/140 \rceil \) edges of \( G \). The probability that such a \( T \) exists is no more than

\begin{align*}
\sum_{t=1}^{m_1} \binom{m_1}{t} \left( \binom{t}{m_2} \right) p^{m_2} &= \sum_{t=1}^{m_1} \binom{ne}{t} \left( \frac{t}{2m_2} \right)^{m_2} \\
&\leq \sum_{t=1}^{m_1} \binom{ne}{t} \left( \frac{70et}{n} \right)^{ct/140} \leq \sum_{t=1}^{m_1} \left( \frac{4900e^4 t}{n} \right)^{ct/280} = o(1)
\end{align*}

using \( c \geq 300 \).

For \( |S| \geq m_3 = \lceil n/360000e^4 \rceil \) we can ignore the fact that the vertices of \( S \) are large. The probability that such an \( S \) exists violating (2.5) is no more than
\[ \frac{|n/14|}{s = m_3} \binom{n}{s} \binom{n-s}{6s} (1-p)^s(n-7s) \]

\[ \leq \left( \frac{|n/14|}{s = m_3} \right) \binom{ne}{6s} e^{-cs/2} \]

\[ \leq \sum_{s = m_3}^{\left| n/14 \right|} \left( 6^8 \cdot 10^{21} \cdot e^{35} \cdot e^{-c/2} \right)^s = o(1) \]

which proves (2.5).

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

\[ \sum_{s = \left| n/14 \right|}^{n/2} \binom{n}{s} BS(cs/10, s(n-s)) \]

\[ \leq 2 \sum_{s = \left| n/14 \right|}^{n/2} \left( \frac{ne}{s} \right)^s \left( \frac{10s(n-s)e}{cs} \right)^{cs/10} \left( \frac{c}{n} \right)^{cs/3} e^{-cs/3} \]

\[ \leq 2 \sum_{s = \left| n/14 \right|}^{n/2} \left( 14e(10e)c/10e - c/3 \right)^s = o(1). \]

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

\[ X_1 = \{ v \in V_n : |N_G(v) \cap \bigcup_{t=0}^{i-1} X_t| \geq 2 \} \]

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

(2.11) \[ |X| \leq 2e^{4c4e^{-4c/3n}} \text{ a.s.} \]
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

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

Let

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

and let

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

It follows then that

(2.13a) \( |A(x)| = 2k \)

and we will show

(2.13b) \( \varepsilon \leq k + 1 \) and if \( \varepsilon = k + 1 \) then \( D(x) \) is a binary tree rooted at \( x \).

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

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

$$
\lambda_{k,\varepsilon,m} \leq \binom{n}{k} \binom{n}{\varepsilon} \left(\frac{k+\varepsilon}{m}\right)^{k+\varepsilon} p^{m/10}(n-k-\varepsilon)^{\varepsilon} \leq \binom{n}{k} \binom{n}{\varepsilon} \left(\frac{k+\varepsilon}{2m}\right)^{k+\varepsilon} e^{m/2} e^{-2e^{\varepsilon}/3} (1 - \frac{c}{n})^{-\varepsilon(k+\varepsilon)}
$$

$$
\lambda_{k,\varepsilon,m} = u_{k,\varepsilon,m}
$$

Now if $c \leq 2\log n$, $k, \varepsilon \leq n^{1/3}$ then $u_{k,\varepsilon,m+1}/u_{k,\varepsilon,m} \leq n^{-1/4}$ for $n$ large.

Thus

$$
(2.15) \sum_{m=2k}^{(k+\varepsilon)} \lambda_{k,\varepsilon,m} \leq (1+o(1)) u_{k,\varepsilon,2k}.
$$

With the same bounds on $c, k, \varepsilon$ and with $n$ large and $\varepsilon \leq k+1$ we have

$$
(2.16) u_{k,\varepsilon,2k} \leq 21^{\varepsilon-k}(e^4c^2k)^{k-\varepsilon}e^{-2e^{\varepsilon}/3} \text{ which implies }
$$

$$
\sum_{\varepsilon=2}^{k+1} u_{k,\varepsilon,2k} \leq 21(e^4c^2k/n) \sum_{\varepsilon=2}^{k+1} (n/\varepsilon e^{2c/3})^{\varepsilon}
$$
\[ s \leq n(e^{4c^2}) e^{-2ck/3} \]

\[ \leq ne^{-ck/2} \quad \text{as } c \geq 300. \]

It follows that \( s \leq \log n \text{ a.s.} \), and we can assume \( k \leq \log n \). Now, using (2.16),

\[
\sum_{k=2}^{\log n} \sum_{\ell=2}^{k} \mu_{k,\ell,2k} \leq 21 \sum_{k=2}^{\log n} (e^{4c^2})^{k} e^{-2ck/3} \\
\leq 22(e^{4c^2})^{4} e^{-4c/3}
\]

and so

(2.17) the number of sets \( K, L \) with \( 2 \leq \ell \leq k \) is a.s. less than \( n^{1/2}e^{-4c/3} \).

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

(2.18) \[ \sum_{m \geq 2k} \mu_{k,k+1,m} \leq (1+o(1))\mu_{k,k+1,2k} \]

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

(2.19) \[ \text{Exp}(\tau_k) = \left( \frac{n}{2k-1} \right)^{2k-2} \binom{2k-1}{2} \left( 1-p \right)^{2k-2} \text{BS}(c/10-1,n-2k+1)^{k} \leq n(e^{2c^2}e^{-2c/3})^{k} \quad \text{for } n \text{ large.} \]
To estimate \( \text{Var}(\tau_k) \), let \( \{T_1, T_2, \ldots, T_B\} \), \( B=\binom{n}{2k-1}^k \), 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 \( G_p \) in which all leaves are small.

Next let \( Y_p = \{(i,j): |V(T_i) \cup V(T_j)| = p\} \) for \( p=2k-1, \ldots, 4k-2 \) and let \( Z_{p,q} = \{(i,j) \in Y_p : |E(T_i) \cup E(T_j)| = q\} \). Then

\[
(2.20) \quad \text{Exp}(\tau_k^2) = \text{Exp}(\tau_k) + \Delta_1 + \Delta_2
\]

where

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

and

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

Now

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

where

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

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

It follows that

\[
(2.21) \quad \Delta_1 \leq \text{Exp}(\tau_k)^2 (1-p)^{-2k^2}
\]
Now for \( p \leq 4k-3 \) we have

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

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

(2.22) \leq ne^{-ck/2} \quad \text{for } n \text{ large.}

(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 R.H.S. of (2.19). This together with (2.17) and (2.18) proves the result.

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

\( H:=G; \)

while \( \delta(H) < k \) do

begin

\( Y:= \{ v \in V(H) : d_H(v) < k \}; \)

\( H:= H[V(H) - Y] \)

end

end
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 $s(H) \geq k$ 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 $k=1,2$ let $A_k = A_k(G_n,p) = V_k(G_n,p)-(WUXUYK)$ where $W,X$ are as defined in Lemmas 2.1, 2.2 respectively and

$$Y_k = \{ y \in V_n : d_{G_n,p}(y) = k \text{ and } N_{G_n,p}(y) \cap X \neq \emptyset \}.$$  

Let $H_k = H_k(G_n,p) = G_n,p[A_k]$, then we have

Lemma 2.3

For $k=1,2$ let $M$ be any matching of $G_n,p[A_k]$ which is not incident with any small vertex. Let $H_k = H_k - M$, then (2.5) implies:

$$(2.23) \quad \emptyset \not\subseteq A_k, |S| \leq n/14 \text{ implies } |N_{H}(S)| \geq k|S|.$$

Proof

Let $G = G_n,p$, $H = H_k$ and for a given $S$ let $S_1 = S \cap \text{SMALL}$ and $S_2 = S - S_1$. Now

$$(2.24) \quad |N_H(S)| \geq |N_H(S_1)| - |S_2| + |N_H(S_2)| - \min(|S_1|, |S_2|).$$

We can write $\min(|S_1|, |S_2|)$ in place of $|S_1|$ as no vertex of $S_2$ is adjacent to more than one vertex of $S_1$, as $S_2 \cap X = \emptyset$.

Also, we claim

$$(2.25) \quad |N_H(S_1)| \geq k|S_1|.$$  

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 $N_G(S_1) \cap (WUYK) = \emptyset$ 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 then

$\text{(2.26)} \quad |N_H(S_2)| \geq 4|S_2|$

For then $|N_G(S_2)| \geq 6|S_2|$ and for each $v \in S_2$, $|N_G(v)| \leq |N_H(v)| + 2$. This is because $v$ is incident with at most one edge of $M$ and $v$ is adjacent to at most one vertex of $W \times Y_k$. It is a simple matter to verify (2.23) from (2.24), (2.25) and (2.26) by considering $|S_1| \geq |S_2|$ and $|S_1| < |S_2|$ as separate cases.
3. **Matchings**

Let $H_1$ be the subgraph of $G$ defined in Lemma 2.3. We are going to prove that $H_1$ a.s. has a perfect or near perfect matching. We first establish that $H_1$ is large.

**Lemma 3.1**

\[
|V(H_1)| = n(1 - (1+\epsilon_1(c))e^{-c}) \quad \text{a.s.} \]

where $\epsilon_1(c) \to 0$ as $c \to \infty$.

**Proof**

\[
|V(H_1)| \geq |V_1(G)| - |W| - |X| - |Y_1-W|. 
\]

It is well known that

\[
|V_1(G)| = (1+o(1))n(1-e^{-c}) \quad \text{a.s.} 
\]

where the $o(1)$ term in (3.2) could for example be taken to be $\pm n^{-1/4}e^{-c/2}$, using the Chebycheff inequality.

Lemmas 2.1 and 2.2 give a.s. upper bounds on $|W|$, $|X|$ and (3.1) will follow from

\[
|Y_1-W| \leq |X| 
\]

For $y \in Y_1$ 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 (3.4) follows.

We establish next the following condition that goes with a graph not having a (near) perfect matching.
Lemma 3.2

Suppose \( \mu(H) < \lfloor |V(H)|/2 \rfloor \). 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 \( M \in \mathcal{M} \). For \( i=1, 2, \ldots, t \) there exists a set \( U_i \subseteq U \) satisfying

\[
\text{(3.4a)} \quad |N_H(U_i)| < |U_i|;
\]

\[
\text{(3.4b)} \quad w \in U_i \text{ implies } e=\{u, w\} \notin E(H) \text{ and } \mu(H) < \mu(H+e).
\]

Proof

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 \( S_i \) be an even length alternating path w.r.t. \( M_i \). Let \( U_i = S_i \cup U_i' \subseteq U \). Then (3.4b) holds otherwise \( M_i \) has an augmenting path.

If \( u \in N_H(U_i) \) then \( u \notin S_i \) and so there exists \( y_1 \) such that \( \{u, y_1\} \in M_i \). We show that \( y_1 \in U_i \) which will prove (3.4a). 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 1.1

We use a coloring argument that was introduced in Fenner and Frieze [5]. 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 recolourings are done independently of each
other.

Let $E^b$, $E^g$ denote the blue and green edges respectively and let $G^b = (V_n, E^b)$, $H_1 = H_1(G)$ and $H_1^b = H_1(G^b)$.

Remark 3.1

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 $p_1 = p p'/(1-p(1-p'))$ and excluded with probability $1-p_1$.

Consider next the following 2 events:

$G = G_{n,p}$ satisfies the conditions of Lemmas 2.1, 2.2 and $u(H_1) < |V(H_1)|/2$.

$E \equiv (a) \not\exists S \subseteq A_1(G^b), |S| \leq n/14 \implies |N_{H_1^b}(S)| \geq |S|$; 

(b) $u(H_1^b) < \lfloor |V(H_1^b)|/2 \rfloor$;

(c) there does not exist $e = \{v, w\} \in E^g, e \subseteq A_1(G^b)$ such that some maximum cardinality matching of $H_1^b$ leaves both $v$ and $w$ isolated.

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

(3.5) \[ \lim_{n \to \infty} \Pr(G) = 0. \]

To prove (3.5) we shall prove

(3.6a) \[ \Pr(E \mid G) \geq (1 - o(1))(1-p')^{2n/3} \]

(3.6b) \[ \Pr(E) \leq (1-p)^n 2/392 \]

which together imply (3.5).

Proof of (3.6a)

Let $G_0 \in G$ be fixed and let $M_0$ be any fixed maximum cardinality matching of $H_1$. We prove
(3.7) \( \Pr(E| G_{n,p} = G_0) \geq (1-p')^{2n/3} - 16(\log n)^4/c^2n. \)

We can readily verify this once we have shown that

(3.8) \( E \cap G \supseteq E_1 \cap E_2 \cap E_3 \cap G \)

where

- \( E_1 = E^g \) is a matching of \( G_0 \);
- \( E_2 \) = no green edge meets any vertex of degree less than \( c/10 + 2 \) in \( G_0 \) or any vertex in \( W \times Y_1 \);
- \( E_3 = M_0 \cap E^g = \emptyset \)

For \( E_1 \cap E_2 \) implies

(3.9) \( A_1(G_0^b) = A_1(G_0) \)

and then \( E_1 \) implies (see Lemma 2.3) that (2.23) holds, which verifies \( E(a) \). \( E(b) \) follows directly from (3.9) and \( G_0 \in G \). \( E_3 \) implies \( \mu(H_1^b) = \mu(H_1) \) and \( E(c) \).

Now it follows from (2.3) that

(3.10) \( \Pr(\bar{E}_1) \leq 16(\log n)^4/c^2n. \)

From Lemmas 2.1, 2.2 and (3.3) we find that the total number of edges of \( G_0 \) that are excluded by the conditions in \( E_2 \), \( E_3 \) is no more than

\[ n((c/10 + 1)e^{-2c/3} + 4nce^{-cE})n + n/2 \leq 2n/3 \]

Thus

\[ \Pr(\bar{E}_1 \cup \bar{E}_2 \cup \bar{E}_3) \leq 1-(1-p')^{2n/3} + 16(\log n)^4/c^2n \]

which proves (3.7).

Proof of (3.6b)

Now

(3.11) \( \Pr(E) = \sum_r \Pr(E| G^b = r) \Pr(G^b = r) \)
where \( r \) is an arbitrary graph with vertices \( V_n \).

Now if \( H_1(r) \) fails to satisfy \( \mathcal{E}(a), \mathcal{E}(b) \) then \( \Pr(\mathcal{E}|G^b = r) = 0 \). So let us assume that \( \mathcal{E}(a), \mathcal{E}(b) \) hold.

Now if \( U, U_1, \ldots, U_t \) are as defined in Lemma 3.2 with \( H = H_1 \), then each set is of size at least \( n/14 \) and for \( \mathcal{E}(c) \) to hold no green edge can join \( u_i \in U \) to \( w \in U_1 \). But then in view of Remark 3.1 and \( \mathcal{E}(a) \) we have

\[
\Pr(\mathcal{E}(c) | G^b = r) \leq (1-p_1)^n/392
\]

which implies (3.6b).

We have thus shown that

\[
u(G) \geq n(1 - (1+\varepsilon_1(c))e^{-c})/2 \quad \text{a.s.}
\]

On the other hand (3.2) implies

\[
u(G) \leq n(1+o(1))(1 - e^{-c})/2 \quad \text{a.s.}
\]

and Theorem 1.1 follows.

If we put \( c = \log n + \omega \) where \( \omega \to \infty \) then we have \( a(c) = 1 - (1+o(1))e^{-\omega}n^{-1} \) and then \( G_{n,p} \) a.s. has a matching of size at least \( (n - (1+o(1))e^{-\omega})/2 \). This is Erdos and Rényi's result [4], (what we have proved is that \( H_1 \) a.s. has a matching of size \( \lfloor |V(H_1)|/2 \rfloor \) and one can see that when \( c = \log n + \omega \), \( H_1 = G_{n,p} \) a.s.).
4. **Cycles**

Let $H_2$ be the subgraph of $G$ defined in Lemma 2.3. We are going to prove that $H_2$ a.s. has a hamiltonian cycle. The proof is very similar to that of section 3 and as such we will only give the essential differences.

**Lemma 4.1**

(4.1) \[ |V(H_2)| = n(1 - (1+\epsilon_2(c))ce^{-c}) \quad \text{a.s.} \]

where $\epsilon_2(c) \to 0$ as $n \to \infty$

**Proof**

\[ |V(H_2)| \geq |V_2(G)| - |W| - |X| - |Y_2 - W \cup X| \]

Now \[ |Y_2 - W \cup X| \leq |X| \]

follows by a similar argument to (3.3). Now let $Z_0$ be the set of vertices of degree 0 or 1 in $G$ and let $Z_1$, $Z_2$, ... be the sequence of sets removed in each iteration of the 2-core finding algorithm of section 2. Now, corresponding to (3.2), it is also well known that

\[ Z_0 = (1-o(1))n(1-ce^{-c}) \quad \text{a.s.} \]

We complete the proof of the lemma by showing that

\[ Z_i \subseteq X \cup W_1 \cup Y_2 \quad i=1,2,\ldots \]

Thus assume inductively that $Z_1$, $Z_2$, ..., $Z_{i-1} \subseteq X \cup W_1 \cup Y_2$ for some $i \geq 1$ (true vacuously for $i=1$) and let \[ T = \bigcup_{t=0}^{i-1} Z_t. \]

Then $y \in Z_i$ implies $d_G(y) \geq 2$ but $|N_G(y) - T|$ \leq 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 \(d_G(y) = 2\) implies \(y \in X \cup W_1 \cup Y_2\).

Lemma 4.2

If \(c\) is large enough and \(G\) satisfies the conditions in Lemmas 2.1, 2.2 then \(H_2\) is connected.

Proof

If \(H = H_2\) is not connected then there exists a nonempty \(S \subseteq V(H)\) such that \(N_H(S) = \emptyset\). We show that this is not possible for \(c\) large enough. (2.23) implies that \(|S| \geq n/14\). (4.1) implies that, for \(c\) large, fewer than \(2ce^{-c}n\) vertices are deleted from \(G\) in producing \(H\). Then (2.2) implies that at most \(8c^2e^{-c}n\) edges are lost in the construction. But then (2.6) implies that not all edges with one vertex in \(S\) have been deleted.

The analogue of Lemma 3.2 is

Lemma 4.3

Let \(H\) be a connected graph which is non-hamiltonian. Then

(a) (4.2) no edge of \(H\) joins the endpoints of any longest path of \(H\).

(b) Let \(U = \{u_1, u_2, \ldots, u_t\}\) be the set of vertices which are endpoints of longest paths of \(H\). For \(i = 1, 2, \ldots, t\) there exists \(U_i \subseteq U\) satisfying

(4.3a) \(|N_H(U_i)| < 2|U_i|\);

(4.3b) \(w \in U_i\) implies \([u_i, w] \in E(H)\) and there is some longest path of \(H\)
that joins \( u_i \) to \( w \).

**Proof**

(4.2) is straightforward and (4.3) is from Posà [11].

We can now give an outline of the

**Proof of Theorem 1.2**

We define \( E^b, E^g \) and \( G^b \) as in the proof of Theorem 1.1 and let \( H^b_2 = H_2(G^b) \). Let now

\[ G = G_{n,p} \]

satisfies the conditions of Lemma's 2.1, 2.2 and \( H_2 \) is not hamiltonian, which implies that (4.2) holds with \( H=H_2 \).

We have only to show that (3.5) holds with this definition of \( G \). Let now

\[ E = (a) \emptyset \neq S \subseteq A_2(G^b), |S| \leq n/14 \text{ implies } |H^b_2(S)| \geq 2 |S|; \]

(b) there does not exist \( e=(v,w) \in E^b \cup E^g \) such that \( v, w \) are the endpoints of some longest path of \( H^b_2 \).

We replace (3.6) by

\[ (4.3a) \ Pr(E|G) \geq (1-o(1))(1-p)^{3n/2}; \]

\[ (4.3b) \ Pr(E) \leq (1-p)^{n^2/392}. \]

This will prove the theorem.

To prove (4.3a) let \( G_0 \in G \) be fixed and let \( P_0 \) be some longest path of \( H_2 \).

We define \( E_1, E_2 \) as before and define \( E_3 \equiv P_0 \cap E^g = \emptyset \).

Now \( E_1 \cap E_2 \) implies that \( A_2(G^b_0) = A_2(G_0) \) and then (3.8) and (4.3a) will
follow in the same way as (3.8) and (3.6a) previously.

To prove (4.3b) we use (3.11) and concentrate on the case where \( H_2(r) \) satisfies \( E(a) \). We note that for \( E(b) \) to hold there is no \( \{v,w\} \in E^q, v \in U, w \in U_i \) where \( U, U_1, U_2, \ldots U_t \) are defined by (4.3) w.r.t. \( H=H_2(r) \). (4.3b) follows from Remark 3.1 and \( E(a) \) as before.

We note that if we put \( c=\log n+\log \log n+\omega \) where \( \omega \to \) then we obtain the result of Komlós and Szemerédi [8] and Korsunov [9].

Finally note that our Corollary follows from the Percolation Theorem of McDiarmid [10].
References


