Almost universal graphs

Alan Frieze∗ Michael Krivelevich†

December 28, 2005

Abstract

We study the question as to when a random graph with $n$ vertices and $m$ edges contains a copy of almost all graphs with $n$ vertices and $cn/2$ edges, $c$ constant. We identify a "phase transition" at $c = 1$. For $c < 1$, $m$ needs to grow slightly faster than $n$, and we prove that $m = O(n \log \log n / \log \log \log n)$ is sufficient. When $c > 1$, $m$ needs to grow at a rate $m = n^{1+a}$, where $a = a(c) > 0$ for every $c > 1$, and $a(c)$ is between $1 - \frac{2}{(1-o(1))c}$ and $1 - \frac{1}{c}$ for large enough $c$.

1 Introduction

1.1 Problem statement and results

A graph $G$ is universal for a class of graphs $\mathcal{H}$ if for every $H \in \mathcal{H}$, there is a subgraph of $G$ which is isomorphic to $H$. The problem of constructing small $G$ which are universal for interesting classes $\mathcal{H}$ has attracted much attention as it arises in the study of VLSI circuit design. See for example [1] and the references there-in. This paper shows for example that if $\mathcal{H} = \mathcal{H}(c,n)$ is the class of graphs with vertex set $[n]$ and maximum degree $c$, then any $\mathcal{H}$-universal graph must contain $\Omega(n^{2-2/c})$ edges.

On the other hand, it is shown in [1] that almost every graph with $(1+\epsilon)n$ vertices and $An^{2-1/c}(\log n)^{1/c}$ edges is $\mathcal{H}(c,n)$-universal. Here $A$ depends only on $\epsilon$. Furthermore, the results of [2] prove the existence of an $\mathcal{H}(c,n)$ universal graph with $O(n)$ vertices and $O(n^{2-2/c}(\log n)^{1+8/c})$ edges.

∗Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh PA15213, U.S.A. Supported in part by NSF grant CCR-0200945.
†Department of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. E-mail: krivelev@post.tau.ac.il. Research supported in part by USA-Israel BSF Grant 2002-133, and by grant 64/01 from the Israel Science Foundation.
Suppose now we consider $H = H^*(c, n)$ where $H^*(c, n)$ is the set of all labeled graphs with vertex set $[n]$ and average degree at most $c$. Clearly $H^*(c, n) \supseteq H(c, n)$ but in fact an $H^*(c, n)$-universal graph requires $\Omega(n^{2-o(1)})$ edges since it must contain all graphs with $(1 - \epsilon)n$ isolated vertices and a $[c/\epsilon]$-regular graph on the remaining $\epsilon n$ vertices.

In this paper we relax the strict notion of universality. We say that $G$ is almost universal for the class of graphs $H$ if it contains subgraphs isomorphic to all but $o(|H|)$ graphs in $H$. In particular we consider $H = H^*(c, n)$. As customary we denote by $G_{n,m}$ the probability space of all labeled graphs on $n$ vertices with $m$ edges, where all such graphs are equiprobable, i.e., $Pr[G] = \binom{n}{2}^{-1}$. Similarly, $G_{n,p}$ stands for the probability space of all labeled graphs with vertex set $\{1, \ldots, n\}$, where each pair $1 \leq i \neq j \leq n$ is an edge of a sample graph $G_{n,p}$ with probability $p = p(n)$, independently of all other edges. We study

$$\Pr(G_{n,m} \text{ is almost universal for } H^*(c, n)).$$

More formally, we should perhaps consider

$$\Pi^*(c, m, \epsilon) = \Pr(\{|H \in H^*(c, n) : G_{n,m} \cong H\| \geq (1 - \epsilon)|H^*(c, n)|\})$$

where $\cong$ denotes “contains a subgraph isomorphic to”.

We can reduce this to estimating the probability

$$\Pi(c, m) = \Pr(G_{n,m} \cong G_{n, cn/2})$$

where $G_{n,m}$ and $G_{n, cn/2}$ are generated independently.

We can link definitions (2) and (3) via

$$1 - \epsilon^{-1}(1 - \Pi(c, m)) \leq \Pi^*(c, m, \epsilon) \leq \frac{\Pi(c, m)}{1 - \epsilon}.$$ 

To verify this, let $N = \binom{n}{2}$ and $M_1 = \binom{N}{m}$, $M_2 = \binom{N}{cn/2}$. Consider the $M_1 \times M_2$ matrix $A$ where $A(i, j) = 1_{G_i \cong H_j}$, assuming that $G_i, i \in [M_1]$ (resp. $H_j, j \in [M_2]$) is an enumeration of all graphs with vertex set $[n]$ and $m$ (resp. $cn/2$) edges. $\Pi^*(c, m, \epsilon)$ is the proportion of rows of $A$ with at least $(1 - \epsilon)M_2$ 1’s and $\Pi(c, m)$ is the proportion of entries of $A$ which are 1. Equation (4) is now easy to verify.

We show that there is a sharp difference in $\Pi(c, m)$ for the cases $c < 1$ and $c > 1$ respectively. We prove the following:

**Theorem 1**

(a) Suppose that $c < 1$ is constant. Then if $A$ is constant,

$$\Pi(c, m) = \begin{cases} 1 - (1 - e^{-c^3/6})e^{-A^3/6} + o(1), & m = An, \\ 1 - o(1), & m \geq \frac{C_0 \log \log n}{\log \log \log n}. \end{cases}$$
for some sufficiently large $C_0 = C_0(c)$.

(b) Suppose that $c > 1$ is constant. Then for some constants $C_1, C_2$,

$$
\Pi(c, m) = \begin{cases} 
  o(1), & m \leq C_1 n^{2-2/(c+x_c)} \\
  1 - o(1), & m \geq C_2 n^{2-1/(c-y_c)}, 
\end{cases}
$$

for some $x_c, y_c \to 0$ as $c \to \infty$.

\[\square\]

We doubt that the upper bound in (a) is tight:

**Conjecture** If $c < 1$ and $m/n \to \infty$ then $\Pi(c, m) = 1 - o(1)$.

It is more difficult to guess whether the upper or lower bound is correct in (b).

### 1.2 $m$ or $p$

In work on random graphs, it is usually more convenient to work in the independent model $G_{n,p}$ rather than in $G_{n,m}$. We therefore estimate

$$
\Pi^\#(p_1, p_2) = \Pr(G_{n,p_2} \supseteq G_{n,p_1})
$$

where $G_{n,p_1}$ and $G_{n,p_2}$ are generated independently.

Putting $p_1 = c/n$ and $p_2 = 2m/n$ we relate $\Pi$ and $\Pi^\#$ through

$$
\Pi(c, m) \geq \Pi^\# \left( p_1 + \frac{\log n}{n^{3/2}}, p_2 - \frac{m^{1/2} \log n}{n^2} \right) - o(1) 
$$

$$
\Pi(c, m) \leq \Pi^\# \left( p_1 - \frac{\log n}{n^{3/2}}, p_2 + \frac{m^{1/2} \log n}{n^2} \right) + o(1).
$$

The inequalities come from the fact that whp $G_{n,p_1-\log n/n^{3/2}}$ contains fewer than $cn/2$ edges etc.

We break the proof of Theorem 1 into 4 pieces:

### 2 $c < 1$, $m = An$.

This is straightforward. It is well known (see, e.g., Theorem 3.19 of [6]) that the probability of $G_{n,(c+o(1))/n}$ not to contain a triangle is asymptotically equal to $e^{-c^3/6}$.  

3
We thus get:

\[
\Pr(G_{n,p_1-\frac{\log n}{n^{3/2}}} \text{ contains a triangle and } G_{n,p_2+\frac{\log n}{n^{3/2}}} \text{ is triangle free})
\]

\[
= (1 - e^{-c^3/6}) e^{-A^3/6} + o(1).
\]

The upper bound in part (a) of the theorem now follows from this and (3)–(6). Having made this connection once, we will just focus on \(\Pi^\#(p_1, p_2)\) from now on.

3 \(c < 1, \ m = o(n \log \log n)\).

From now on let us shorten \(G_{n,p_i}\) to \(G_i\) for \(i = 1, 2\).

We first state some almost sure properties of the random graph \(G_1\) required in our proof.

**Lemma 1** Let \(0 < c < 1\) be a constant. The random graph \(G = G_{n,c/n}\) has whp the following properties:

1. All connected components of \(G\) have size at most \(L = (c - 1 - \log c)^{-1} \log n\);
2. All connected components of \(G\) are trees or unicyclic;
3. The number of vertices in unicyclic components of \(G\) does not exceed \(\frac{\log \log n}{\log \log \log n}\);
4. The number of isolated vertices in \(G\) is at least \(n/e\).

All of the above properties are quite standard and can be found, e.g., in [6].

Let \(p_1 = c/n, \ m = \omega n/2\) and \(p_2 = \omega/n\) where

\[
\omega = \frac{120 \log \log n}{\log \log \log n}.
\]

The following properties of \(G_1\) are more technical and pertain closely to our arguments. They aim to estimate from above the expected number of vertices of \(G_2\) needed to embed all large tree components of \(G_1\).

Let a vertex of \(G_1\) be **large** if its degree is at least \(\geq \omega/20\).

For a tree \(T\) of \(G_1\), let

\[
\sigma(T) = \prod_{i=1}^{a(T)} (d_i - 1)!
\]

where \(d_1, d_2, \ldots, d_{a(T)}\) are the degrees of the large vertices of \(T\).
Lemma 2 Let $\omega/20 \leq k \leq L$, $a > 0$, $\sigma$, $D$ be integers. Denote by $Y(k,a,\sigma,D)$ the number of isolated tree components $T$ of $G$ with $|V(T)| = k$, $a(T) = a$, $\sigma(T) = \sigma$ and $d_1 + \ldots + d_a = D$, where $d_1, \ldots, d_a$ are the degrees of large vertices in $T$. Let $0 < c < 1$ be a constant. Then the random graph $G = G(n,c/n)$ has whp the following property:

$$\sum_{k \leq L, a, \sigma, D} Y(k,a,\sigma,D) k(1 - e^{-\omega/10})^{-k} e^{2\omega a}3^{D-a} \sigma^{-D+a} < \frac{n}{\log^2 n}.$$ 

Proof Let $k^* = k - 2 - D + a$. The number of trees $T$ with parameters $k, a, \sigma, D$ in the complete graph $K_n$ can be estimated from above by:

$$\left(\begin{array}{c} n \\ k \end{array}\right) \sum_{d_1 + \ldots + d_a = D} \frac{k - 2}{d_1 - 1, d_2 - 1, \ldots, d_a - 1, k^*} (k)_a (k - a)^{k^*} \leq \frac{(1 + o(1)) n^k (k - 2)! k^a (k - a)^{k^*}}{k^*!} \sum_{d_1 + \ldots + d_a = D} \prod_{i=1}^a \frac{1}{(d_i - 1)!} < (1 + o(1)) n^k k^a \left(\frac{D - 1}{a - 1}\right) \sigma^{-1} e^{k-a}.$$

(In the first line above, we choose $k$ vertices of the tree in $\left(\begin{array}{c} n \\ k \end{array}\right)$ ways, then choose names of a large vertices in $(k)_a$ ways, then decide which positions in the Prüfer code of the tree are occupied by each of the large vertices, then fill the rest of the positions by remaining vertices in $(k-a)^{k^*}$ ways.) For each such tree, the probability that it is an isolated component in $G(n,c/n)$ is $(c/n)^{k-1}(1-c/n)^{k(n-k)} \leq (1 + o(1))(c/n)^{k-1}e^{-ck}$. Using the linearity of expectation we derive that the expected value of the random variable in the lemma’s formulation is at most:

$$\sum_{k,a,\sigma,D} n^k k^a \left(\frac{D - 1}{a - 1}\right) \sigma^{-1} e^{k-a} (c/n)^{k-1} e^{-ck} k(1 - e^{-\omega/10})^{-k} e^{2\omega a}3^{D-a} \sigma^{-D+a}$$

$$\leq (1 + o(1)) \sum_{k,a,\sigma,D} (ce^{1-c})^k (k/e)^a 2^D k(1 - e^{-\omega/9}) e^{2\omega a}3^{D-a} \omega^{-D+a}$$

$$< 2n \sum_{k,a,\sigma,D} k((1 + o(1))ce^{1-c})^k (ke^{2\omega w}/3e)^a (6/\omega)^D.$$

Next observe that $D \geq aw/20$ and that there are at most $\binom{D-1}{a-1} \leq 2^D$ possible values for $\sigma$, given $a, D$. Thus the last expression can be bounded by

$$2n \sum_{k,a,D} k((1 + o(1))ce^{1-c})^k ((ke^{2\omega w}/3e)^{20/\omega})^D (12/\omega)^D$$

$$\leq 2n \sum_{k,a,D} k((1 + o(1))ce^{1-c})^k (k^{20/\omega}e^{50})^D (12/\omega)^D.$$
Observe that $k^{20/\omega} \ll \omega^{1/2}$. Indeed, $k \leq L = O(\log n)$, while

$$w^{\omega/40} = \left(\frac{120 \log \log n}{\omega \log \log \log n}\right)^{\frac{120 \log \log n}{\omega \log \log \log n}} \approx (\log n)^{3+o(1)}.$$ 

Therefore, $(k^{20/\omega}e^{50})^D (12/\omega)^D \ll (12\omega^{1/2}/\omega)^D \omega^{-(1+o(1))\omega/20} = (\log n)^{-6+o(1)}$. Also, for $c < 1$ one has $ce^{1-c} < 1$, and thus $k(ce^{1-c})^k = O(1)$. Since $k \leq L = O(\log n)$, $D \leq 2k$, $a \leq k$, we have altogether $O(\log^3 n)$ summands, each at most $(\log n)^{-6+o(1)}$, hence the sum is at most $(\log n)^{-3+o(1)}$. It follows by the Markov Inequality that the random variable of the lemma whp has value at most $n/\log^2 n$.

\square

**Lemma 3** For an integer $k \geq 2$, denote $\pi_k = (1 - e^{-\omega/10})^{k-1}$. Let $\tau_k$ be the number of isolated tree components of size $k$ in $G$. Then, for every $0 < c < 1$, whp in $G(n, c/n)$

$$\sum_{k=(\log \log n)^2}^{L} \frac{k\tau_k}{\pi_k} \leq \frac{n}{\log^2 n}.$$ 

**Proof** First we estimate the expectation of the expression in question:

$$E \left( \sum_{k=(\log \log n)^2}^{L} \frac{k\tau_k}{\pi_k} \right) = \sum_{k=(\log \log n)^2}^{L} \binom{n}{k} k^{k-1} p_1^{k-1} (1 - p_1)^{(k(n-k))+(k-1)\frac{\log n}{\log \log n}} \bar{p}_1$$

where

$$\bar{p}_1 = p_1(1 - e^{-\omega/10})^{-1} \text{ and } \epsilon_1 = \frac{p_1e^{-\omega/10}}{1 - e^{-\omega/10} - p_1} = \frac{ce^{-\omega/10}}{n - c - ne^{-\omega/10}} \leq \frac{2ce^{-\omega/10}}{n}.$$ 

Thus, writing $\bar{p}_1 = \tilde{c}_1/n$ we see that

$$E \left( \sum_{k=(\log \log n)^2}^{L} \frac{k\tau_k}{\pi_k} \right) \leq \frac{n}{\tilde{c}_1} \sum_{k=(\log \log n)^2}^{L} (\tilde{c}_1 e^{1-\tilde{c}_1 + o(1)})^k,$$

and since $1 - \tilde{c}_1 e^{1-\tilde{c}_1 + o(1)}$ is bounded below by a positive function of $c$,

$$E \left( \sum_{k=(\log \log n)^2}^{L} \frac{k\tau_k}{\pi_k} \right) = o(n/\log^2 n).$$
Applying the Markov inequality, the lemma follows. \hfill \Box

Suppose now that $G_1$ consists of isolated trees $T_1, T_2, \ldots, T_s$ and unicyclic components $K_1, K_2, \ldots, K_t$. We will assume that $G_1$ satisfies the conditions stated in Lemmas 1–3. We will assume that $a(T_i) \neq 0$ iff $i \leq r$. Note that $r \leq 21cn/\omega = o(n)$.

We try to embed the trees one by one in $G_2$. Our strategy for embedding a tree $T_i$ into $G_2$ is as follows: Choose a vertex $v_1$ of degree one of $T_i$ and then let the vertices of $T_i$ be $v_1, v_2, \ldots, v_k$ where the order comes from some breadth first search of $T_i$. Let $d_j$ be the degree of $v_j$ in $T_i$. Suppose that the embedding of trees $T_1, T_2, \ldots, T_{i-1}$ has involved examining vertices $w_1, w_2, \ldots, w_\ell$. Let $w_{\ell+1}$ be the lowest numbered vertex in $U = \{w_1, w_2, \ldots, w_\ell\}$, the set of unexamined vertices. We use the following algorithm to try to embed $T_i$ into $G_2$. Basically, at each stage we try to embed $v_i$ as $w_{\ell+i}$ by finding $d_i^* = d_i - 1 + 1_{i=1}^a$ new neighbours for $w_{\ell+i}$ from which to continue the embedding.

\begin{verbatim}
Begin
  \lambda = \ell + 1.
  For i = 1, 2, \ldots, k do
    Begin
      If $w_{\ell+i}$ has $\geq d_i^*$ neighbours in $U$ then
        Begin
          Let $x_j, j = 1, 2, \ldots, d_i^*$, be the $d_i^*$ lowest numbered neighbours in $U$.
          $U \leftarrow U \setminus \{x_j : j = 1, 2, \ldots, d_i^*\}$
          $w_\kappa \leftarrow x_{i, \kappa = \lambda + 1, \lambda + 2, \ldots, \lambda + d_i^*}$
          $\lambda \leftarrow \lambda + d_i^*$.
        End
      Else FAIL
    End
  End
End
\end{verbatim}

If we fail then we repeat this procedure until either we succeed or $\ell$ reaches $n_F = n(1 - e^{-1})$, in which case we abandon the attempt to embed $G_1$ into $G_2$.

### 3.1 Phase 1: Embedding trees with large vertices

Consider first the embedding of the first $r$ trees, i.e. those with $a(T) > 0$. At the end of this phase of the embedding we expect $\ell$ to be at most $O(n/(\log n)^2)$ and consider the attempt to be a failure if $\ell$ reaches $n/(\log n)^{1/2}$ before $T_r$ has been embedded.

Suppose the tree $T_i, i \leq r$, has the same parameters as in Lemma 2 and the degrees of the large vertices are $d_1, d_2, \ldots, d_a$. The probability of success in any attempt is at
least

\[(1 - e^{-\omega/10})^k \prod_{i=1}^a \psi(d_i)\]

where \(\psi(d_i)\) is the probability that a vertex has \(\geq d_i - 1\) \(G_2\) neighbours in \(U\). Let us condition on having consumed at most \(n/2\) vertices before starting to embed \(T_i\). Then

\[\psi(d_i) \geq \left(\frac{n/2}{d_i - 1}\right) p_2^{d_i-1} (1 - p_2)^{n} \geq \frac{\omega^{d_i-1}}{3^{d_i-1}(d_i - 1)!} e^{-2\omega}\]

and then

\[\prod_{i=1}^a \psi(d_i) \geq \frac{\omega^{D-a}}{3^{D-a}\sigma(T)} e^{-2\omega}a.\]

So the expected time (increase in \(\ell\)) while embedding this tree is at most

\[k(1 - e^{-\omega/10})^{-k} e^{2\omega} 3^{D-a} \sigma(T) \omega^{-D+a}.\]

This explains the expression we put earlier in Lemma 2.

Let \(X_i\) be the number of vertices consumed while embedding \(T_i\) (or reaching a failure while embedding \(T_i\)). Conditioning on \(X_1 + \ldots + X_{i-1} \leq n/2\), the random variable \(X_i\) is dominated by \(k\) times the geometric random variable with probability of success

\[\rho_i = (1 - e^{-\omega/10})^k \prod_{i=1}^a \psi(d_i) \geq (1 - e^{-\omega/10})^k \frac{\omega^{D-a}}{3^{D-a}\sigma(T)} e^{-2\omega}a.\]

Denote \(\mu_i = k/\rho_i\). Then

\[\Pr[X_i \geq (\log^{1.5} n)\mu_i | X_1 + \ldots + X_{i-1} \leq \frac{n}{2}] \leq (1 - \rho_i)^{\log^{1.5} n/\rho_i} = o(1/n).\]

Observe that by Lemma 2, whp,

\[\sum_{i=1}^r \mu_i \leq \frac{n}{\log^{2} n}.\]

It thus follows:

\[\Pr[X_1 + \ldots + X_r \geq \frac{n}{\log^{1/2} n}]\]

\[\leq o(1) + \Pr[X_1 + \ldots + X_r \geq \log^{1.5} n \sum_{i=1}^r \mu_i] \]

\[\leq o(1) + \Pr \left[ \bigcup_{i=1}^r \left( X_i \geq (\log^{1.5} n)\mu_i | X_1 + \ldots + X_{i-1} \leq \frac{n}{2} \right) \right] \]

\[= o(1) + o(r/n) = o(1).\]

Thus, whp Phase 1 completes successfully, having used at most \(n/(\log n)^{1/2}\) vertices, and the remaining trees of \(G_1\) have their maximum degree at most \(\omega/20\).
3.2 Phase 2: Embedding trees with no large vertices

We use the same embedding algorithm as before and a similar argument to analyze it. Note that \( \sum_i s_i = r_1 + |V(T_i)| \leq (1 - 2\alpha)n \) where \( \alpha = e^{-1/2} \). We declare Phase 2 to be a failure if it uses at least \((1 - \alpha)n\) vertices. At each stage of the embedding \(|U| \geq \alpha n/2\) for a tree \( T_i \) with \( k \) vertices, the probability of a successful attempt is at least

\[
\Pr\left( \text{Bin}\left( \frac{\alpha n}{2}, p_2 \right) > \omega/20 \right)^{k-1} \geq \pi_k = (1 - e^{-\omega/10})^{k-1}.
\]

(7)

Let \( \tau_k \) denote the number of trees of size \( k \) in \( G_1 \) and let \( Z_{i,k} \) denote the number of vertices used in attempts to embed \( T_{i,k} \), the \( i \)th tree of size \( k \). Then

1. \( Z_{i,k} \) is dominated by \( k \) times the geometric random variable \( \Gamma_{i,k} \) which has probability of success \( \pi_k \), see (7). We can couple \( Z_{i,k} \) with a copy of \( \Gamma_{i,k} \) so that \( Z_{i,k} \leq \Gamma_{i,k} \).
2. If \( Z = \sum_{k=2}^{\log \log n} \sum_{i=1}^{\tau_k} Z_{i,k} < (1 - \alpha)n \) then the embedding succeeds.

Let \( Z_2 = \sum_{k=2}^{\log \log n} \sum_{i=1}^{\tau_k} Z_{i,k} \). Observe that by Lemma 3, the expectation of \( Z_2 \) does not exceed \( n/\log^2 n \). Thus we can apply the argument similar to that of Phase 1 to show that \( \text{whp} \) it takes \( O(n/\log^{1/2} n) \) vertices to embed tree components of \( G_1 \) with at least \( (\log \log n)^2 \) vertices.

We now turn to \( Z_1 = \sum_{k=2}^{\log \log n} \sum_{i=1}^{\tau_k} Z_{i,k} \). Consider the number of failures as we try to embed trees with at most \( (\log \log n)^2 \) vertices. Let \( \nu = \frac{n}{(\log \log n)^2} \). The probability of failing at any attempt is at most \( 1 - \pi_{(\log \log n)^2} \leq (\log \log n)^2 e^{-\omega/10} < e^{-\omega/20} \). Furthermore, there will be less than \( n \) attempts embedding a tree and so the probability that there are \( \nu \) failed attempts or more is at most

\[
\left( \frac{n}{\nu} \right)^{(e^{-\omega/20})^\nu} \leq \left( \frac{en}{\nu} e^{-\omega/20} \right)^\nu = o(1).
\]

Each failed attempt consumes at most \( (\log \log n)^2 \) vertices. Thus \( \text{whp} \)

\[
Z_1 = \sum_{k=2}^{(\log \log n)^2} \sum_{i=1}^{\tau_k} kZ_{i,k} = \sum_{k=2}^{(\log \log n)^2} k\tau_k + O(n/\log \log n) < (1 - 2\alpha)n.
\]

Hence we conclude that \( \text{whp} \) \( Z = Z_1 + Z_2 < (1 - \alpha)n \) and the embedding succeeds.

3.3 Phase 3: Embedding unicyclic components

Recall that by Lemma 1 \( G_1 \) contains at most \( \omega/120 \) vertices altogether in unicyclic components. We divide the unused vertices of \( U \) into two sets \( U_1, U_2 \) of approximately
equal size, which \textbf{whp} are at least \( an/2 \). Then \textbf{whp} \( U_1 \) will, in \( G_2 \), contain at least \( \omega \) cycles of size \( j \) for each \( 3 \leq j \leq \omega \). So, we will \textbf{whp} be able to choose a collection of cycles in \( U_1 \) to match with the cycles of \( G_1 \). We can then use \( U_2 \) to embed the trees attached to these cycles which go to make up each of the unicyclic components. Since these trees are without large vertices, we use the analysis of the previous section and argue that the expected time to embed these trees is \( o(n) \). This completes the analysis for \( c < 1 \).

\[ \square \]

4 \quad \text{\( c > 1, \, m \leq C_1 n^{2-2/(c+x_c)} \)}

Let \( x = x_c \) be the unique solution in \((0,1)\) to \( xe^{-x} = ce^{-c} \). It is easy to verify that \( c + x_c > 2 \) for all \( c > 1 \). Let \( \alpha = 1 - \frac{x}{c} \) and \( \beta = \frac{x}{2} \left( 1 - \frac{x^2}{c^2} \right) \). \textbf{Whp} \( G_{n,cn/2} \) contains a giant component with \( A \) vertices and \( B \) edges, where \( |A - \alpha n|, |B - \beta n| = O(n^{1/2} \log n) \) vertices, [3].

We associate with each graph \( G_1 \) in \( G_{n,cn/2} \), containing such a giant component, a graph \( H \) with \( A \) vertices and \( B \) edges. Then

\[
\Pr[G_{n,m} \supseteq G_1] \leq \Pr[G_{n,m} \supseteq H] \leq (n)_A \left( (1 + o(1)) \frac{2m}{n^2} \right)^B \\
\leq n^{\alpha n + O(\sqrt{n} \log n)} \left( \frac{2m}{n^2} \right)^{\beta n + O(\sqrt{n} \log n)} \\
= \left[ n^{(1 + o(1)) \alpha} \left( \frac{2m}{n^2} \right)^{(1 + o(1)) \beta} \right]^n.
\]

Hence if \( m \leq c_1 n^{2-2/(c+x_c)} \) for \( c_1 > 0 \) small enough, the above expression tends to 0 as \( n \) grows, implying that \textbf{whp} \( G_{n,m} \) does not contain most of the graphs from \( G_{n,cn/2} \).

\textbf{Comments.} (a) Since \( c + x_c > 2 \), we get that for every \( c > 1 \) there exists \( \epsilon = \epsilon(c) > 0 \) such that \( \Pi(c, m) = o(1) \) for \( m \leq n^{1+\epsilon} \). This, combined with the result of Theorem 1 (a), shows that the function \( \Pi(c, m) \) has sharp threshold at \( c = 1 \);

(b) As pointed out by the referee, we can get a somewhat weaker bound \( \Pi(c, m) = o(1) \) for \( m = n^{2 - \frac{2}{c} - o(1)} \) by the following simple counting argument: the number of subgraphs with \( \frac{cn}{2} \) edges of a graph \( G_2 \) with \( m \) edges is \( \left( \frac{m}{n/2} \right) \). The total number of non-isomorphic graphs on \( n \) vertices with \( \frac{cn}{2} \) edges is at least \( \left( \left( \frac{n}{2} \right) \right)^{-1} \). Thus, if

\[
\left( \frac{m}{cn/2} \right) = o \left( \left( \frac{n}{2} \right)^{-1} \right),
\]

\( any \) graph \( G \) with \( m \) edges does not contain most of the graphs with \( cn/2 \) edges. Solving the above inequality for \( m \) we get the claimed bound.
5 Upper bound

We now show that if \( c > 1 \) and \( p_2 = n^{-1/(1+o(1))c} \) then \( \text{whp} \ G_{n,p_2} \supseteq G_{n,p_1} \).

We note first that Pittel, Spencer and Wormald [7] have shown that the threshold probability \( c_k/n \) for having a \( k \)-core satisfies \( c_k = k + \sqrt{k \log k} + O(\log k) \) and so for large \( c \) we find that \( \text{whp} \ G_{n,c/n} \) is \( d \)-degenerate for \( d = c - \sqrt{c \log c} + O(\log c) \). (A graph is \( d \)-degenerate if its vertices can be ordered as \( v_1, v_2, \ldots, v_n \) so that \( v_i \) has at most \( d \) neighbours in \( \{v_1, v_2, \ldots, v_{i-1}\} \).) Also, \( G_1 \) is \( K_{2,3} \)-free \( \text{whp} \). Finally, \( \text{whp} \ G_1 \) has \((e^{-c} + o(1))n \) isolated vertices. We will therefore be able to use the following:

**Theorem 2** Let \( d > 3, 0 < c_0 < 1 \) be constants. Let \( \delta = \max \left\{ \frac{1}{d-1}, \frac{1}{d(d-3)} \right\} \). Let \( p(n) = An^{(-1+\delta)/d} \). Let \( H \) be a \( d \)-degenerate \( K_{2,3} \)-free graph on \( c_0 n \) vertices, of maximum degree \( \Delta(H) \leq \Delta_0 = 1/(4dp) \). If \( A = A(c_0, d) \) is large enough, then \( \text{whp} \) the random graph \( G(n, p) \) contains a copy of \( H \).

**Proof** Fix an ordering \( \sigma = (v_1, \ldots, v_{c_0n}) \) of the vertices of \( H \) such that every vertex of \( H \) has at most \( d \) neighbors preceding it in \( \sigma \). Denote \( N_H^{-}(v_i) = \{v_j : j < i, (v_i, v_j) \in E(H)\} \). We will embed \( H \) vertex by vertex, according to \( \sigma \).

Partition the vertex set of \( G(n, p) \) into two parts: \( V_0 \) of size \( |V_0| = \frac{1-c_0}{2} n \), and \( V_1 \). We will use \( V_0 \) to embed the first \( v_0 = A_1 n^{\delta} \) vertices of \( \sigma \), and \( V_1 \) to embed the rest, the value of \( A_1 = A_1(d) \) will be chosen later to satisfy inequality (8).

Observe that if \( A \) is large enough then \( \text{whp} \) in \( G(n, p) \) every \( d \) vertices of \( V_0 \) have at least \( A_1 n^{\delta} \) common neighbors in \( V_0 \). Thus we can easily embed the first \( A_1 n^{\delta} \) vertices of \( H \) according to \( \sigma \) in \( V_0 \).

Now, we will use Theorem 2 of Fernandez de la Vega and Manoussakis [4] (or rather its proof) to embed the rest of \( H \) in \( V_1 \). We start by adding some edges to it to form a new graph \( H' \). Specifically, for each \( i > A_1 n^{\delta} \), if \( v_i \in V(H) \) has less than \( d \) neighbors preceding it in \( H \), we add to \( H \) \( d - |N_H^{-}(v_i)| \) edges connecting \( v_i \) to random vertices before it in the order. Let \( H' \) be the (random) graph obtained. Denote by \( U_i \) the set of neighbors of \( v_i \) in \( H' \) preceding \( v_i \).

Let \( A_1 n^{\delta} < i < j \). We will estimate the probability of \( U_i = U_j \). Assume \( |N_H^{-}(v_i)| = s_1, \) \( |N_H^{-}(v_j)| = s_2, |N_H^{-}(v_i) \cap N_H^{-}(v_j)| = t \). We first expose \( d - s_1 \) random neighbors of \( v_i \); for \( U_i = U_j \) they should contain all \( s_2 - t \) vertices of \( N_H^{-}(v_j) \setminus N_H^{-}(v_i) \), this happens with probability at most:

\[
\left( \frac{(i-1-s_2+t)}{d-s_1-s_2+t} \right)^{s_2-t} < \left( \frac{i-1}{d-s_1} \right)^{s_2-t} \leq \frac{(1 + o(1))(d - s_1)^{s_2-t}}{i^{s_2-t}}.
\]

Now expose \( d - s_2 \) random neighbors of \( v_j \), they should coincide with \( U_i \setminus N_H^{-}(v_j) \), the
probability of this to happen is 
\[
\frac{1}{(j-1)^{d-s}} = \frac{(1 + o(1))(d - s)!}{j^{d-s}!}.
\]

So altogether the probability that \( U_i = U_j \) is less than \( \frac{d^d}{j^{d-s}} \).

Recall that \( H \) is \( K_{2,3} \)-free and thus \( t \leq 2 \). Also, \( t > 0 \) only for those \( i < j \) who have a common neighbor in front of them in \( \sigma \), and this happens for at most \( d\Delta(H) \) values of \( j \), for a given \( i \). Therefore the probability that there exist \( i, j \) such that \( U_i = U_j \) is at most:

\[
\sum_{A_1n^d < i < j} \frac{d^d}{j^{d-2}} + \sum_{A_1n^d < i < j} \frac{d^d}{i^{d-1}}
\]

\[
\leq d^d \left[ \frac{d\Delta(H)}{(d-1)(A_1n^d)^{d-3}} + \frac{n}{(d-1)(A_1n^d)^{d-1}} \right]
\]

\[
< \frac{1}{2}
\]

for large enough \( A_1 = A_1(d) \).

Now we argue that **whp**

each \( U_k \) intersects at most \( 2d\Delta_0 \) of the sets \( U_j \) with \( j < k \).  \( \quad \text{(9)} \)

Clearly \( N_H^-(v_k) \) intersects at most \( d\Delta_0 \) sets \( N_H^-(v_j) \) with \( j < k \). We thus need to show that **whp** the random edges of \( H' - H \) do not add \( d\Delta_0 \) sets \( U_j \) intersecting \( U_k \), for any given \( k \). To do so, we fix \( v_k \), condition on \( U_k \), fix \( u \in U_k \). Then the number \( Z_u \) of vertices \( v_j \) that choose to add a random edge \( (v_j, u) \) is dominated by \( X_1 + X_2 + \ldots + X_n \) where \( X_1, X_2, \ldots, X_n \) are independent and \( X_i \) is 0/1 and \( \Pr(X_i = 1) \leq \frac{d}{n_0 + i - 1} \) for \( i = 1, 2, \ldots, n \). Then \( \mathbf{E}(X_1 + X_2 + \ldots + X_n) \leq O(\log n) \) and by Theorem 1 of Hoeffding [5] we see \( \Pr(Z_u \geq \Delta_0) \) is exponentially small. We conclude then that the probability that (9) is violated is less than 1/2, too. Thus, when \( A > 0 \) is large enough, there exists a supergraph \( H' \supseteq H \) satisfying

(i) For each pair \( A_1n^d \leq i < j \), \( U_i \setminus U_j \neq \emptyset \).

(ii) Each \( U_k \) intersects at most \( 2d\Delta_0 \) sets \( U_j, j \neq k \).

Now consider continuing an embedding of \( v_1, v_2, \ldots, v_{j-1} \) in \( G_2 \) where \( j > n_0 \). As in [4] we define, for each \( i < j \) a vertex \( x_i \in U_i \setminus U_j \). Suppose our embedding has assigned
\(v_i \to w_i\) for \(i < j\) and that \(W = [n] \setminus \{w_1, w_2, \ldots, w_{j-1}\}\). Let the correspondence \(v_i, w_i\) map \(U_i\) to \(W_i\) and for \(w \in W\) let \(N_2(w)\) be the neighbours of \(w\) in \(G_2\). Suppose our embedding algorithm checks each \(w \in W\) as a candidate in increasing value of \(w\). Let \(J = \{i < j : U_i \cap U_j \neq 0\}\). Then arguing as in [4] we see

\[
\Pr(N_2(w) \supseteq W_j \mid \text{history of process}) = \Pr(N_2(w) \supseteq W_j \mid N_2(w) \not\supseteq W_i, i \in J')
\]

for some \(J' \subseteq J\)

\[
\geq \Pr(N_2(w) \supseteq W_j \text{ and } x_i \notin N_2(w), i \in J')
\]

\[
\geq p^d(1 - p)^{2d\Delta_0}
\]

\[
\geq p^d/2.
\]

Thus,

\[
\Pr(\exists w : N_2(w) \supseteq W_j \mid \text{history of process}) \leq (1 - p^d/2)(1 - c_0)n/2 \leq e^{-An^d/4}.
\]

Thus \textit{whp} the embedding succeeds for all \(j \leq c_0n\).

We apply the above theorem with \(H\) equal to \(G_1\) minus its isolated vertices. Then \textit{whp} \(H\) has at most \((1 - e^{-c} + o(1))n\) vertices. Finally, we will need

\[
p_2 = n^{-c^{-1}(1+\sqrt{\log c}/c+O(c^{-1}))}
\]

which is somewhat better than claimed in Theorem 1(b).

**Comment 1** As pointed out by the referee, the techniques of [2] can be used to get an explicit construction of a graph \(G\) with \(\tilde{O}(n^{2-1(c'+1)})\) edges, where \(c' = \lceil c \rceil\), containing almost all graphs with \(cn/2\) edges. This is about the same as the upper bound of \(n^{2-(1+o(1))}/c\) derived in this paper, our argument shows however that almost all graphs with that many edges are almost universal.

**Comment 2** At a point quite late in the reviewing process, Andrzej Ruciński became cognisant of and reminded us of the relevance of the paper by Riond [8]. Theorem 2.1 of that paper is similar to Theorem 1(b). He shows that if

\[
\gamma(G) = \max_{H \subseteq G}\{|E(H)|/(|V(H)| - 2)|\}
\]

then our \(1/(c - y_c)\) can be replaced by the likely value of \(1/\gamma(G_{n,p})\). Now the value for \(\gamma(G_{n,p})\) is not easy to estimate although it does seem likely that \(\gamma(G_{n,p}) = c + o(c)\). Furthermore, Riond’s proof is very different relying on the second moment method, whereas ours is constructive.

**Acknowledgement.** The authors wish to thank Itai Benjamini for posing this problem. They are also thankful to the anonymous referee and to Andrzej Ruciński for their helpful criticism on the first version of the paper.
References


