On small subgraphs of random graphs

Alan Frieze
Carnegie Mellon University

Follow this and additional works at: http://repository.cmu.edu/math
§1. Introduction

Let \( H \) be some fixed graph with \( r \) vertices and \( s \) edges. \( H \) is assumed to be strictly balanced i.e.

\[
\frac{s}{r} > \frac{\mu(H')}{\nu(H')}
\]

for all non-trivial subgraphs \( H' \) of \( H \), \( H' \neq H \), where \( \nu(H') \), \( \mu(H') \) are the numbers of vertices, edges in \( H' \) respectively. (From now on \( H' \subset H \) will always mean such subgraphs).

Consider now the random graph \( G_{n,m} \) chosen uniformly from \( \mathcal{G}_{n,m} = \{ \text{graphs with vertex set } [n] = \{1,2,\ldots,n\} \text{ and } m \text{ edges} \} \) and let \( X_H \) denote the number of distinct copies of \( H \) in \( G_{n,m} \). Suppose now \( m = \frac{1}{2} \omega n^{2-r/s} \) where \( \omega = \omega(n) \). Erdős and Rényi [3] showed that

\[
\Pr(X_H = 0) = 1 - o(1) \quad \text{if } \omega \to 0
\]
\[
\Pr(X_H \neq 0) = 1 - o(1) \quad \text{if } \omega \to \infty.
\]

Here, as usual, we consider limits etc. as \( n \to \infty \). Using \( a(n) \sim b(n) \) to stand for \( a(n) = (1 - o(1)) b(n) \), we remark that

\[
E(X_H) \sim \frac{\omega^s}{\alpha} = \lambda, \text{ say},
\]

where \( \alpha \) denotes the number of automorphisms of \( H \).

Erdős and Rényi’s result has been refined in many ways. In particular, Bollobás [1] and Karonski and Rucinski [6] independently showed that if \( \omega \) tends to a constant and \( k \) is a fixed non-negative integer then
The aim of this paper is to show that the Poisson expression (1.1) is good for $\omega \to \infty$ reasonably fast. In particular we prove

**Theorem 1.1**

Let $H$ be strictly balanced and $\lambda$ be as defined above. Then there exists a positive real constant $\theta = \theta(H)$ such that if $\omega = o(n^\theta)$ then

$$\Pr(X_{H} = k) \sim e^{-\lambda} \frac{\lambda^k}{k!}$$

(1.2) \hspace{1cm} 0 \leq k \leq (1 + \epsilon_1)\lambda

where $\epsilon_1 = \frac{A_1(\log n)^{r/(2r-1)}}{\lambda^{(r-1)/(2r-1)}}$ for some constant $A_1 > 0$.

$$\Pr(X = k) \gg e^{-\lambda} \frac{\lambda^k}{k!} \hspace{1cm} (1 + \epsilon_2)\lambda \leq k \leq \lambda \log n$$

(1.3) \hspace{1cm} \text{where} \quad \epsilon_2 = A_2\left(\frac{\log n}{\lambda^{1-2/r}}\right)^{r/2(r-1)} \text{ for some constant } A_2 > 0, \text{ provided } \epsilon_2 \to 0.

(The notation $a(n) \gg b(n)$ is used for $a(n)/b(n) \to \infty$).

**Remarks**

1. We are not able to obtain the largest possible values for $\theta(H)$ although we hope to refine our analysis for particular graphs e.g. triangles.

2. Observe that $\epsilon_1 \lambda \gg \lambda^{1/2}$ and so (1.2) is valid into the tails of the Poisson distribution.
3. A somewhat stronger result for \( k = 0 \) and \( G_{n,p} \) has been proved independently by Boppanna and Spencer \[2\] and Jansen, Łuczak and Rucinski \[4\]. Jansen \[5\] has extended these result to estimate \( \Pr(X_H \leq k) \) for \( k \leq E(X_H) \).

4. See Rucinski \[7\] for a recent survey on the distribution of the number of copies of small subgraphs of random graphs.

§2. Proof of Theorem 1.1.

We will not specify \( \theta(H) \) immediately but upper bounds for it will be derived along with the proof. We will use \( A, A_1, A_2, \ldots \) to denote absolute constants whose values may or may not be explicitly stated.

We distinguish between isolated copies of \( H \) and non-isolated copies. Here a copy of \( H \) in \( G_{n,m} \) is isolated if it shares no edge with any other copy of \( H \).

Now let

\[
\tau_{k,\ell} = \Pr(G_{n,m} \text{ contains exactly } k \text{ isolated and } \ell \text{ non-isolated copies of } H)
\]

and

\[
q_\ell = \sum_{k=0}^{\infty} \tau_{k,\ell} = \Pr(G_{n,m} \text{ contains exactly } \ell \text{ non-isolated copies of } H)
\]
The main work involved in the proof is to justify the following inequalities:

\( p_k = \sum_{\ell=0}^{k} \pi_{k-\ell, \ell} = \Pr(G_{n,m} \text{ contains exactly } k \text{ copies of } H) \)

The inequalities are:

\[
\begin{align*}
(2.1) & \quad n^{-A_3 \ell^2/r} \leq q_{\ell} \leq n^{-A_4 \ell^1/r} & 0 \leq \ell < \lambda_0 = \lambda(\log n)^4 \\
(2.2) & \quad \Pr(G_{n,m} \text{ contains at least } \lambda_0 \text{ isolated copies of } H) = o(e^{-\lambda_0})
\end{align*}
\]

and more importantly:

\[
\begin{align*}
(2.3) & \quad \frac{\pi_{k, \ell}}{\pi_{k-1, \ell}} = (1 + \epsilon_{k, \ell}) \frac{\lambda}{k} & 0 \leq k-1, \ell \leq \lambda_0
\end{align*}
\]

where \( |\epsilon_{k, \ell}| = o(\lambda_0^{-1}) \).

We devote the remainder of this section to showing how our theorem follows from (2.1) – (2.3) and prove these inequalities later on.

Suppose now that \( 0 \leq \ell \leq \lambda_0 \). It follows from (2.3) that

\[
\begin{align*}
(2.4) & \quad \pi_{i, \ell} = (1 + o(1)) \pi_{0, \ell} \frac{\lambda^i}{i!} & 0 \leq i \leq \lambda_0
\end{align*}
\]

and so

\[
q_{\ell} = (1 + o(1)) \pi_{0, \ell} \sum_{i=0}^{\lambda_0} \frac{\lambda^i}{i!} + \sum_{i > \lambda_0} \pi_{i, \ell}
\]
\[ = (1 + o(1)) \pi_0, \ell (e^\lambda - o(e^{-\lambda_0})) + o(e^{-\lambda_0}) \]

on using (2.2). Hence

\[ \pi_0, \ell = (1 + o(1)) (q_\ell - o(e^{-\lambda_0})) e^{-\lambda} \]

and by (2.4)

\[ \pi_i, \ell = (1 + o(1)) q_\ell e^{-\lambda \frac{\lambda^i}{i!}} + o(e^{-\lambda_0}) \quad 0 \leq i \leq \lambda_0 \]

Thus

\[ p_k = (1 + o(1)) \sum_{\ell=0}^k q_\ell e^{-\lambda \frac{\lambda^k - \ell}{(k-\ell)!}} + o(e^{-\lambda_0}) \quad 0 \leq k \leq \lambda_0. \]

Now

\[ p_k \geq q_k \geq n \rightarrow e^{-\lambda_0 \lambda_0} \quad \text{since } r \geq 3 \]

and so

\[ p_k \sim \sum_{\ell=0}^k q_\ell e^{-\lambda \frac{\lambda^k - \ell}{(k-\ell)!}} \quad 0 \leq k \leq \lambda_0. \]
\begin{equation}
(2.5) \quad \frac{\lambda^k}{k!} (q_0 + \sum_{\ell=2}^{k} \frac{(k)_\ell}{\ell!} q_\ell)
\end{equation}

where \((k)_\ell = k(k-1)...(k-\ell+1)\).

To proceed from here we need \(q_0 = 1 - o(1)\). To prove this we need a lemma on the edge density of intersecting copies of \(H\). We need a general version of this to prove (2.1) and we prove this here. Let

\[ \theta_1 = \min_{H' \subset H} \left( \frac{2s - \mu(H')}{2r - \nu(H')} - \frac{s}{r} \right) > 0. \]

Note that \(\theta_1 > 0\) follows from the fact that \(H\) is strictly balanced. A collection \(H_1, H_2, ..., H_k\) of copies of \(H\) in \(G_{n,m}\) is said to be linked if for each \(i\) there is \(j \neq i\) such that \(H_i, H_j\) share an edge.

**Lemma 2.1**

Let \(H_1, H_2, ..., H_k\), \(k \geq 2\) be a linked collection of copies of \(H\). Let \(K = \bigcup_{i=1}^{k} H_i\). Then

\[ \mu(K) \geq (\theta_1 + \frac{s}{r})\nu(K). \]

**Proof**

Assume w.l.o.g. that \(H_i \notin \bigcup_{j \neq i} H_j\) for \(i = 1, 2, ..., k\). We prove the result by induction on \(k\). We discuss the base case and the inductive step in tandem. Let \(K' = \bigcup_{i=1}^{k-1} H_i\). Then

\begin{equation}
(2.6) \quad \frac{\mu(K)}{\nu(K)} = \frac{\mu(H_k) + \mu(K') - |E(H_k) \cap E(K')|}{\nu(H_k) + \nu(K') - |V(H_k) \cap V(K')|}.
\end{equation}
Furthermore

\[ uv \in E(H_k) \cap E(K') \rightarrow u, v \in V(H_k) \cap V(K') \]

and so if \( H' = (V(H_k) \cap V(K'), E(H_k) \cap E(K')) \)

then \( H' \) is a non-trivial proper subgraph of \( H \) and, by (2.6)

\[ \frac{\mu(K)}{\nu(K)} = \frac{s + \mu(K') - \mu(H')}{r + \nu(K') - \nu(H')} \]

**Base Case: \( k = 2 \)**

Here \( K' = H_2 \) and \( \mu(K)/\nu(K) \geq \theta_1 + \frac{s}{r} \) follows from the definition of \( \theta_1 \).

**Inductive Step**

Write

\[ \frac{\mu(K)}{\nu(K)} = \frac{2s - \mu(H') + (\mu(K') - s)}{2r - \nu(H') + (\nu(K') - r)} \]

and observe that

\[ (\mu(K') - s) - (\theta_1 + \frac{s}{r})(\nu(K') - r) \]

\[ = (\mu(K') - (\theta_1 + \frac{s}{r})\nu(K')) + r\theta_1 > 0 \]

by induction.
It is always more pleasant to do computation in the independent model $G_{n,p}$, $p = m/N$, $N = \binom{n}{2}$. We quote the following simple results (see Bollobas [], Section 2.1). Let $\mathcal{A}$ be any property of graphs. Then

$$(2.7) \quad \Pr(G_{n,m} \in \mathcal{A}) \leq 3m^{1/2} \Pr(G_{n,p} \in \mathcal{A})$$

and if $\mathcal{A}$ is monotone then

$$(2.8) \quad \text{a.e. } G_{n,p} \in \mathcal{A} \rightarrow \text{a.e. } G_{n,m} \in \mathcal{A}.$$ 

**Lemma 2.2**

If

$$(2.9) \quad \theta < \theta_1 r^2/(s^2 + \theta_1 rs)$$

then $q_0 = 1 - o(1)$.

**Proof**

If $G_{n,m}$ has a pair of edge intersecting copies of $H$ then it contains a set of $k \leq 2r-1$ vertices which span at least $[k(\frac{s}{r} + \theta_1)]$ edges. Now this property is monotone and

$$\Pr(G_{n,p} \text{ contains a pair of edge intersecting copies of } H)$$
Now use (2.8).

Referring to (2.5), suppose first that \(0 < k < \lambda\). then for \(\theta\) sufficiently small

\[
1 - o(1) \leq q_0 + \sum_{\ell=2}^{k} \frac{(k)}{\ell^2} q_{\ell} \leq q_0 + \frac{k}{\lambda^{r-1}(2r-1)} \leq 1
\]

Now let \(k = (1 + \varepsilon)\lambda\) where \(0 \leq \varepsilon \leq \varepsilon_1 = A_1 (\log n)^{r/(2r-1)}/\lambda^{(r-1)/(2r-1)}\). Then, using (2.1)

\[
u_{\ell} = \frac{(k)}{\lambda^2} q_{\ell} \leq 2\frac{(k)}{\lambda^2} e^{-\ell^2/2k} n - A_4 \ell^{1/r}
\]

\[
\leq 2 \exp\{\ell - \frac{\ell^2}{2k} - A_4 \ell^{1/r} \log n\}.
\]

Case 1: \(\ell \geq 3 \varepsilon \lambda\)

\[
u_{\ell} \leq 2n^{-A_4 \ell^{1/r}}
\]
Case 2: \( \ell < 3 \varepsilon \lambda \)

\[ u_\ell \leq 2 \exp\{\ell^{1/r}(\varepsilon \ell^{1-1/r} - A_4 \log n)\} \]

\[ \leq 2 \exp\{\ell^{1/r}(3^{1-1/r} \varepsilon^{2-1/r} \lambda^{1-1/r} - A_4 \log n)\} \]

\[ \leq 2 \exp\{\ell^{1/r} \log n(3^{1-1/r} A_1^{2-1/r} - A_4)\}. \]

So if we make \( A_1 \) small enough so that \( A_4 \geq 4A_1^2 \) then we have

\[ u_\ell \leq 2n^{-A_1^2 \ell^{1/r}} \]

which is also valid for Case 1.

Hence if \( \lambda \leq k \leq (1 + \varepsilon_1)\lambda \) and \( \theta \) is sufficiently small

\[ 1 - o(1) \leq q_0 + \sum_{\ell=2}^k \frac{(k)\ell}{\lambda^\ell} q_\ell \leq 1 + 2 \sum_{\ell=2}^\infty n^{-A_1^2 \ell^{1/r}} \]

\[ = 1 + o(1). \]

This together with (2.10) proves the first part of the theorem.

Suppose now that \( k = (1 + \varepsilon)\lambda \) where \( 1 \geq \varepsilon \geq \varepsilon_2 = A_2(\frac{\log n}{\lambda^{1-2/r}})^{r/2(r-1)}. \)

Then by (2.5)
11

\[
\frac{p^k}{k!} \geq \frac{1}{2} \frac{k!}{\lambda^{k-\lfloor k/\lfloor \lambda \rfloor \rfloor} q_{k-\lfloor \lambda \rfloor}} \lambda^{k-\lfloor k/\lfloor \lambda \rfloor \rfloor} q_{k-\lfloor \lambda \rfloor}
\]

\[
\geq A \frac{k}{e\lambda} \lambda^{k-\lfloor k/\lfloor \lambda \rfloor \rfloor} q_{k-\lfloor \lambda \rfloor} \lambda^{2/3} q_{k-\lfloor \lambda \rfloor} \lambda \log n
\]

\[
\geq A \exp\left(\frac{2}{3}(1 - 2A_3 \epsilon^2 - 2 \frac{2}{\lambda^2} - 1) \log n\right)
\]

\[
\geq A \exp\left(\frac{2}{3}(1 - 2A_3 A_2^2 - 2)\right).
\]

Now \( \epsilon^2 \lambda \to \infty \) and we are free to choose \( A_2 \) so that \( 1 - 2A_3 A_2^2 = \frac{1}{2} \) and the result is proved for this case.

When \( k \geq 2\lambda \) we use

\[
\frac{(k+1)!}{\lambda^{(k+1-s)} q_s} \geq \frac{k!}{\lambda^{(k-s)} q_s}
\]

to reduce to the previous case.

\[\square\]

§3. Proof of (2.1) and (2.2)

The upper bound in (2.1) follows fairly easily from Lemma 2.2. Indeed suppose \( G_{n,m} \) contains exactly \( \ell \) non-isolated copies of \( H \). Let \( K \) denote the graph induced by the union of these copies. If \( K \) has \( \rho \) vertices then, by Lemma 2.2, it has at least \( \tau \rho \) edges where
\[ \tau = \theta_1 + \frac{s}{r} \] Note that

\[ \ell^{1/r} \leq \rho \leq r\ell \leq r\lambda_0 \]

where the lower bound on \( \rho \) is from \((p)_x \geq \ell \) Hence, on using (2.7),

\[
q_x \leq 3m^{1/2} \sum_{\rho = \ell^{1/r}}^{r\ell} \left( \frac{n}{\rho} \right)^{\frac{\theta}{2}} \rho^{\tau \rho} \leq 3m^{1/2} \sum_{\rho = \ell^{1/r}}^{r\ell} \left( \frac{n e}{\rho} \right)^{\rho} \left( \frac{\theta}{2\rho^2} \right)^{\tau \rho} \\
\leq 3m^{1/2} \sum_{\rho = \ell^{1/r}}^{r\ell} \left( \frac{A \rho^{\tau - 1} \omega^{\tau}}{n^{\tau/8 - 1}} \right)^{\rho} \leq 3m^{1/2} \sum_{\rho = \ell^{1/r}}^{r\ell} \left( \frac{A \omega^{s(\tau - 1) + \tau (\log n)^4(\tau - 1)}}{n^{\theta_1/s}} \right)^{\rho}
\]

and the upper bound in (2.1) follows provided

\[ \theta(s(\tau - 1) + \tau) < r\theta_1/s. \]

It is convenient to stop and prove a similar inequality which is needed later.

Let \( \lambda_1 = \lfloor \omega^{rs(\log n)^4r + 1} \rfloor \). It follows from (3.1) that provided

\[ \theta(rs(\tau - 1) + \tau) < r\theta_1/s \]
that

\[ 2^{\lambda_1} \sum_{\ell=\lambda_1} q'_{\ell} = o(e^{-2\lambda_0}) \]  

(3.3)

where \( q'_{\ell} \) is the probability that \( G_{n,2m} \) contains precisely \( \ell \) non-isolated copies.

Furthermore, if \( G_{n,2m} \) contains more than \( 2\lambda_1 \) non-isolated copies of \( H \) then we can choose \( \lambda_1 \) of them. For each chosen copy of \( H \) that does not share an edge with another chosen copy we choose a further copy that does share an edge. In this way we build a linked collection of between \( \lambda_1 \) and \( 2\lambda_1 \) copies. It then follows by the calculations above that

\[ \sum_{\ell=2\lambda_1+1}^{\omega} q'_{\ell} = o(e^{-2\lambda_0}), \text{ also.} \]  

(3.4)

To prove the lower bound of (2.1) we consider the probability of the existence of a collection of disjoint complete subgraphs of specific sizes. Thus let \( \sigma_t = \binom{t}{r} \frac{r!}{\alpha} \) for \( t \geq r \) and observe that \( K_t \) contains \( \sigma_t \) distinct copies of \( H \). For a given \( \alpha \) define \( \tau = \tau(\alpha) \) by \( \sigma_{\tau+1} > \alpha \geq \sigma_\tau \).

Next let \( \ell_1 = \ell \) and \( \ell_{i+1} = \ell_i - \tau(\ell_i) \) and \( T_i = \sum_{j=1}^{i} \tau(\ell_j) \) for \( i = 1,2,...,k \) where

\[ \ell_k \geq \frac{(r+1)!}{\alpha} > \ell_{k+1}. \]

Now let \( \mathcal{E} \) denote the event that

\[ G_{n,m} \text{ contains complete subgraphs with vertex set } [T_1],[T_2][T_1],...,[T_k][T_{k-1}] \]  

(3.5a)

and
(3.5b) \( \ell_{k+1} \) copies of \( H \) containing the edge \{1,2\} but otherwise disjoint from all other copies. Let their vertices belong to \([T]\setminus[T_k]\) where \( T - T_k = (r-2)\ell_{k+1} \)

and

(3.5c) there are no other edges in \([T]\) (this assumption simplifies the calculations but may be a bit drastic!)

and

(3.6) there are no other non-isolated copies of \( H \) is \( G_{n,m} \).

Thus if \( \mathcal{E} \) occurs then \( G_{n,m} \) contains exactly \( \ell \) non-isolated copies of \( H \). We can write

\[
Pr(\mathcal{E}) = \pi_1 \pi_2
\]

where

\[
\pi_1 = Pr((3.5)) \text{ and } \pi_2 = Pr((3.6)|(3.5)).
\]

But

\[
\pi_1 = \left[ \frac{(N - \binom{T}{2})}{m - u} \right] / \binom{N}{m} = \frac{m}{N} u \left( 1 - \frac{T^4}{N} + \frac{u^2}{m} \right)
\]
where \( u = k \sum_{i=1}^{k} \left[ \frac{\tau_i}{2} \right] + (s-1)r_k+1 \). So

\[
\pi_1 = \left( \frac{\omega}{n^{r/s}} \right)^u (1 - 0(\frac{1}{n} + \frac{u}{m} + \frac{u}{n}))
\]

(3.7)

\[
= \left( \frac{\omega}{n^{r/s}} \right)^u (1 - o(1)),
\]

since we show later that

\[
\sum_{i=1}^{k} \tau_i^x = O(\frac{\omega^{x/r}}{r})
\]

for any fixed positive integer \( x \),

and we assume

(3.9) \[ \theta < r(2 - \frac{1}{s})/4s. \]

We show next that \( \pi_2 = 1 - o(1) \). Note that (3.6) given (3.5) is monotone and so we can use the \( G_{n,p} \) model to estimate \( \pi_2 \). Now by the FKG inequality

\[
\pi_2 \geq \pi'_2 \pi''_2
\]

where

\[
\pi'_2 = \Pr (\text{there are no non-isolated copies of } H \text{ in } [n]\setminus[T])
\]
and

\[ \pi''_2 = \Pr(\text{there are no extra copies of } H \text{ which share an edge with those defined in (3.5))}. \]

Now \( \pi'_2 = 1 - o(1) \) if (2.9) holds and

\[ \pi''_2 = 1 - E \text{ (number of such copies of } H) \]

\[ \geq 1 - \sum_{H' \in H} (r - \nu(H')) 2rp^{s-\mu(H')} \left( \sum_{i=1}^{k} \tau^i \right)^{\nu(H')}! + o(1) \]

\[ = 1 - O\left( \sum_{H' \in H} n^{r - \nu(H')} \frac{\omega^{s-\mu(H')}}{r - \frac{r}{s} \mu(H')} \tau^i \right) \]

on using (3.8) to simplify the second summation

\[ = 1 - o(1) \]

provided

\[ \theta < \min_{H' \in H} \frac{\nu(H') - \frac{r}{s} \mu(H')}{\mu(H') + \nu(H') \frac{s}{r}} \]

The proof of (2.1) is completed once we have proved (3.8). For then (3.7) implies
When $a$ is large we have, where $\tau = \tau(a)$,

$$a - \sigma_\tau \leq \sigma_{\tau+1} - \sigma_\tau$$

$$= r(\tau - 1) a^{-1}$$

$$< r \tau^{r-1}.$$

But

$$a \geq \sigma_\tau \Rightarrow \left(\frac{\tau}{r}\right) \leq a$$

$$\Rightarrow \left(\frac{\tau}{r}\right)^r \leq a$$

(3.11) $$\Rightarrow \tau \leq r a^{1/r}$$

and so

$$a - \sigma_\tau \leq r^{1-1/r}$$

which implies
(3.12) \[ \ell_i \leq r^i \left( \frac{1 - \frac{1}{r}}{1 - \frac{1}{r}} \right)^i \quad 1 \leq i \leq k \]

and

\[ \tau(\ell_i) \leq r^{i+1} \ell^i \left( 1 - \frac{1}{r} \right)^{i/r}. \]

Now let \( i_0 = \lfloor r \log r \rfloor \) and assume \( \ell \) is large enough that \( i_0 \leq k \) ((3.8) is trivial for bounded \( \ell \)). Then (3.12) implies

(3.13) \[ \ell_{i_0} \leq A \ell^{1/r} \]

where \( A = r^{i_0} \).

Now \( \tau(\ell_i) \leq r^{1/r} \) and \( \tau \) is monotone increasing and so

(3.14) \[ \sum_{i=1}^{i_0} \tau(\ell_i)^x \leq i_0 r^{x/r}. \]

On the other hand it is easy to see that

\[ \sigma \geq \tau \quad \text{for} \quad \tau \geq r + 1 \]

and thus

\[ \ell = (\ell_1 - \ell_2) + (\ell_2 - \ell_3) + \ldots + (\ell_k - \ell_{k+1}) + \ell_{k+1} \]
\[= \sigma_1 + \sigma_2 + \ldots + \sigma_k + \ell_{k+1}\]

and so replacing \( \ell \) by \( \ell_0 \) above

\[\tau(\ell_0 + 1) + \ldots + \tau(\ell_k) \leq \ell_{0+1}^x\]

Hence

\[(3.15) \quad \sum_{i=0}^{k} \tau(\ell_i)^x \leq \left( \sum_{i=0}^{k} \tau(\ell_i)^x \right)^x \leq \ell_{0+1}^x \]

\[= 0(\ell^x) \quad \text{by (3.13)}.\]

(3.8) follows from (3.14) and (3.16) and this completes the proof of (2.1).

We now turn to the proof of (2.2). For positive integer \( t \)

\[\Pr(\exists t \text{ isolated copies of } H \text{ in } G_{n,p}) \leq \frac{1}{t!} \left( \frac{ \ell }{r!} \cdot \frac{ r! }{ \alpha } \cdot p^s \right)^t \]

\[\leq \left( \frac{e \cdot n^r}{t!} \cdot \frac{ r! }{ \alpha } \cdot p^s \right)^t\]
Now put $t = \lambda_0$ and apply (2.7).

The same argument gives

$$\Pr(G_{n,2m} \text{ contains at least } \lambda_1 \text{ isolated copies}) = o(e^{-2\lambda_0})$$

and so, using (3.3), (3.4), we find

$$\Pr(G_{n,2m} \text{ contains } 2\lambda_1 \text{ or more copies of } \mathcal{H}) = o(e^{-2\lambda_0}).$$

§4. Proof of (2.3)

This section contains the main ideas of the proof of Theorem 1.1

Let $\mathcal{A}_{k\ell} = \{ G \in \mathcal{G}_{n,m} : G \text{ has } k \text{ isolated copies and } \ell \text{ non-isolated copies of } \mathcal{H} \}$. Let $a_{k\ell} = |\mathcal{A}_{k\ell}|$ so that (2.3) is actually concerned with the ratio $a_{k,\ell}/a_{k-1,\ell}$.

Now for $k > 0$, $\ell > 0$, let $BP_{k,\ell}$ denote the bipartite graph with vertex partition $\mathcal{A}_{k,\ell}$ and edge set $\mathcal{E}_{k,\ell}$ where $G_1, G_2 \in \mathcal{E}_{k,\ell}$, $G_1 \in \mathcal{A}_{k,\ell}$, $G_2 \in \mathcal{A}_{k-1,\ell}$ if the edge sets of $G_1, G_2$ are related by

$$E(G_2) = (E(G_1) \setminus \{e\}) \cup \{f\}$$

where $e$ is an edge of some isolated copy of $\mathcal{H}$ in $G_1$ and $f$ is some edge which does not create a new copy of $\mathcal{H}$ when added to $G_1/e$.

If $G \in \mathcal{A}_{k,\ell} \cup \mathcal{A}_{k-1,\ell}$ let $d(G)$ denote its degree in $BP_{k,\ell}$. Then
(4.1) \( G \in \mathcal{A}_{k,\ell} \) implies

\[ ks(N - m - \xi(G)) \leq d(G) \leq ks(N - m) \]

where \( \xi(G) \) = the number of copies in \( G \) of a graph of the form \( H - x \) for some edge \( x \in E(H) \).

This is because we have \( ks \) choices for edge \( e \) in an isolated copy of \( H \). Then of the \( N - m \) possible edge replacements \( f \) there are at most \( \xi(G-e) - 1 \) choices which create a new \( H \) when added. Finally observe that \( \xi(G-e) - 1 \leq \xi(G) \).

Also

(4.2) \( G \in \mathcal{A}_{k-1,\ell} \) implies

\[ (m - s(k+\ell))(\xi(G) - 2\xi(G)) \leq d(G) \leq m\xi(G) \]

where \( \zeta(G) \) = the number of subgraphs of \( G \) of the form \( (H_1 \cup H_2) - x \) where \( H_1, H_2 \) are copies of \( H \) which share \( x \) (so if e.g. \( H \) is a triangle then \( (H_1 \cup H_2) - x \) must be a 4-cycle).

To see this we overestimate the number of choices of \( f \) by \( m \) and the number of choices of \( e \) by \( \xi(G) \). To underestimate \( d(G) \) we underestimate the number of choices of \( f \) by \( m - s(k+\ell) \) since we do not wish to touch a copy of \( H \). The number of choices for \( e \), given \( f \), is at least \( \xi(G-f) - \xi(G) \geq \xi(G) - 2\xi(G) \) (crudely.)

The equation

\[ \sum_{G \in \mathcal{A}_{k,\ell}} d(G) = \sum_{G \in \mathcal{A}_{k-1,\ell}} d(G) \]
and (4.1), (4.2) lead to

\begin{equation}
\frac{(m-s(k-\ell))(\xi_{k-1, \ell} - 2 \tilde{\xi}_{k-1, \ell})}{ks(N-m)} \leq \frac{a_{k, \ell}}{a_{k-1, \ell}} \leq \frac{m\tilde{\xi}_{k-1, \ell}}{ks(N-m-\tilde{\xi}_{k, \ell})}
\end{equation}

where $\tilde{\xi}_{k, \ell}$ denote the expectations of $\xi(G)$, $\zeta(G)$ over $\mathcal{A}_{k, \ell}$. It only remains now to estimate these quantities. For $G \in \mathcal{A}_{k, \ell}$ and $e \in E(\bar{G})$ ($\bar{G}$ = complement of $G$) let $h_e$ denote the number of new copies of $H$ created when $e$ is added to $G$. Let $\mathcal{N}(G) = \{e \in E(\bar{G}) : h_e > 0\}$ and $\mathcal{N}(G) = |\mathcal{N}(G)|$. Let $\lambda_1$ be as in (3.2).

**Lemma 4.3**

Let $G = G_{n,m}$:

(a) $\Pr(\exists e \in E(\bar{G}) : h_e \geq 2\lambda_1) = o(n^2 e^{-2\lambda_0})$.

(b) $\Pr(\eta(G) \geq n^{1/8}s\lambda_1 \log n) = o(e^{-2\lambda_0})$.

**Proof**

Let $\mathcal{E}$ denote the event \{G_{n,2m} has at least $2\lambda_1$ copies of $H$\}. Think of $G_{n,2m}$ as $G_{n,m}$ plus $m$ random edges.

(a) Let $\mathcal{E}_a = \{\exists e \in E(\bar{G})$ s.t. $h_e \geq 2\lambda_1\}$. Then

\[ \Pr(\mathcal{E}) \geq \Pr(\mathcal{E} | \mathcal{E}_a) \Pr(\mathcal{E}_a) \]

\[ \geq \frac{m}{N} \Pr(\mathcal{E}_a). \]
Part (a) now follows from (3.17).

(b) 

Let $\lambda_2 = n^{r/s} \lambda_1 \log n$ and $\mathcal{E}_b = \{ \eta(G) \geq \lambda_2 \}$. Then

$$\Pr(\mathcal{E}) \geq \Pr(\mathcal{E} | \mathcal{E}_b) \Pr(\mathcal{E}_b)$$

and (b) follows if we show that $\Pr(\mathcal{E} | \mathcal{E}_b) \geq \frac{1}{2}$. But to see this observe that the expected number of copies of $H$ created by adding the second $m$ edges is at least $\frac{m}{N} \eta(G_{n,m})$ and

$$\frac{m}{N} \lambda_2 \approx \omega \lambda_1 \log n$$

$$\gg \lambda_1.$$

Note that we see now that the actual number added, given $\mathcal{E}_b$, majorizes a binomial with mean $\gg \lambda_1$.

\[ \square \]

Let us now return to the consideration of (4.3). Suppose $\ell \leq \lambda_0$. It follows from (2.1) and (2.2) that there exists $k_0$ such that

$$\pi_{k_0, \ell} \geq n^{-A_3 \ell^2 \ell / r} (2\lambda_0)^{-1}.$$

We prove that

$$\pi_{k, \ell} \geq (1 - \frac{1}{\lambda_0})|k-k_0| n^{-A_3 \ell^2 / r} (2\lambda_0)^{-1} \quad 0 \leq k \leq \lambda_0.$$

(4.13)
This is true for $k = k_0$ and assume inductively that it is true for some $0 < k \leq k_0$. $k > k_0$ will be dealt with subsequently and this is why we are assuming that $k_0 > 0$. We will be able to verify (2.3) as we proceed with the induction. We will estimate $\tilde{\tilde{\xi}}_{k,\ell} \tilde{\tilde{\xi}}_{k,\ell}$ by the same method and to do this we let $\Gamma$ denote a generic graph of the form $H - x$ or $H_1 \cup H_2 - x$. Let $\Gamma_0$ denote some fixed copy of $\Gamma$ with vertex set $\{1,2,...,t\}$, $t = \nu(\Gamma)$ and let $e_1, e_2, ..., e_u, u = \mu(\Gamma)$ be an enumeration of its edges.

Let $\mathcal{A}_{k,\ell}^* = \{G \in \mathcal{A}_{k,\ell} : \text{for } i = 1,2,...,u \text{ we have either (i) } e_i \in E(G) \text{ and } e_i \not\in H \text{ or (ii) } e_i \not\in E(G) \text{ and } e_i \not\in \mathcal{N}(G)\}$.

**Lemma 4.4**

$$1 - \frac{1}{N} \geq \frac{|\mathcal{A}_{k,\ell}^*|}{|\mathcal{A}_{k,\ell}|} \geq 1 - \frac{2 \lambda_2}{N}.$$  

**Proof**

By symmetry, we have

$$\frac{s_k}{N} \leq 1 - \frac{|\mathcal{A}_{k,\ell}^*|}{|\mathcal{A}_{k,\ell}|} \leq E_{k,\ell} \left(\frac{(2s-1)(\eta(G)) + s(k+\ell)}{N}\right)$$

where $E_{k,\ell}$ denotes expectation over $G$ in $\mathcal{A}_{k,\ell}$ (4.13) and Lemma 4.3(b) imply that $E_{k,\ell}(\eta(G)) \leq (1 + \frac{1}{2s})\lambda_2$ and the result follows.

So now let $\mathcal{A}_{k,\ell,i}^* = \{G \in \mathcal{A}_{k,\ell}^* : E(G) \cap \{e_1, ..., e_u\} = \{e_1, ..., e_i\}\}$ for $0 \leq i \leq u$ and consider the bipartite graph $BP_{k,\ell,i}^* \geq 0$, with bipartition $\mathcal{A}_{k,\ell,i}^*, \mathcal{A}_{k,\ell,i-1}^*$ and an edge $G_1 G_2$ for $G_1 \in \mathcal{A}_{k,\ell,i}^*, G_2 \in \mathcal{A}_{k,\ell,i-1}^*$ if $G_2$ can be obtained from $G_1$ by deleting $e_i$ and
adding a new edge $f$. Using $d$ to denote degree in $BP^*_{k,\ell,i}$ we have

\[(4.14) \quad G \in \mathcal{A}^*_{k,\ell,i} \implies N - m - \eta(G) \leq d(G) \leq N - m.\]

There are at most $N - m$ choices for $f$ which gives the upper bound. On the other hand, if $f \notin E(G) \cup \eta(G)$ then $G - e_i + f \in \mathcal{A}^*_{k,\ell,i-1}$. To see this we first note that $G + f$ has the same $k + \ell$ copies of $H$ as $G$. But then if $e_i \notin \mathcal{N}(G - e_i + f)$ we find that $e_i$ belongs to a copy of $H$ in $G + f$ and hence in $G$, which is disbarred by $G \in \mathcal{A}^*_{k,\ell,i}$

\[(4.15) \quad G \in \mathcal{A}^*_{k,\ell,i-1} \implies m - s(k + \ell) \leq d(G) \leq m.\]

There are at most $m$ choices for $f$ and if we choose to delete an $f$ which is not in any copy of $H$ then $G + e_i - f$ is in $\mathcal{A}^*_{k,\ell,i}$. The latter fact following from $e_i \notin \mathcal{N}(G)$. Hence if $a^*_{k,\ell,i} = |\mathcal{A}^*_{k,\ell,i}|$ we have, analogously to (4.3),

\[(4.16) \quad \frac{m - s(k + \ell)}{N} \leq \frac{a^*_{k,\ell,i}}{N} \leq \frac{m}{a^*_{k,\ell,i-1}} \leq \frac{m}{N - m - \eta_{k,\ell,i}}.\]

It follows from (4.13) and Lemma 4.4 that there exists $i_0$ such that

\[a^*_{k,\ell,i_0} \geq \frac{1}{6} \lambda_0^{-1} n - A_3^2 \lambda_0^{2/r} \left( \frac{N}{m} \right).\]

Now (4.16) implies that $a^*_{k,\ell,i}/a^*_{k,\ell,i-1} \geq \frac{m}{2N}$ and so if $i > i_0$
and hence we see from Lemma 4.3(b) that \( \overline{\eta}_{k,\ell,i} \leq 2\lambda_2 \) for \( i \geq i_0 \). But this then implies that for \( i > i_0 \)

\[
(1 - \frac{2s\lambda_0}{m} \cdot \frac{m}{N}) \leq \frac{a_{k,\ell,i}^*}{a_{k,\ell,i-1}^*} \leq (1 + \frac{3(m+\lambda_2)}{N} \cdot \frac{m}{N}).
\]

But if \( i_0 \geq 1 \) we see from (4.21) that \( a_{k,\ell,i_0-1}^* \geq \frac{m}{2N} a_{k,\ell,i_0}^* \). This puts a bound of \( 2\lambda_2 \) on \( \overline{\eta}_{k,\ell,i_0-1} \) and proves (4.18) for \( i = i_0 \). Clearly we can repeat this argument a further \( i_0 - 1 \) times to show that (4.17) holds for \( i \geq 1 \).

It follows that

\[
\Pr(G \text{ contains } \Gamma_0 \mid G \in \mathcal{X}_{k,\ell}^*) = (\frac{m}{N})^u (1 + \epsilon_{k,\ell,\Gamma})
\]

where \( |\epsilon_{k,\ell,\Gamma}| \leq A\omega^{rs}/n^{2-r/s} \).

Let us now deal with \( \xi \). Let \( \Lambda_\xi \) denote the set of possible graphs of the form \( H \rightarrow x \).

Then, from (4.18),

\[
E(\xi(G) \mid G \in \mathcal{X}_{k,\ell}^*) = \sum_{\Gamma \in \Lambda_\xi} \binom{m}{1} \frac{r!}{\alpha_{\Gamma}} (\frac{m}{N})^{s-1} (1 + \epsilon_{k,\ell,\Gamma})
\]

where \( \alpha_{\Gamma} = \text{the number of automorphisms of } \Gamma \).

To handle \( E(\xi(G) \mid G \in \mathcal{X}_{k,\ell} \setminus \mathcal{X}_{k,\ell}^*) \) we note that for such \( G \),
It follows now from Lemmas 4.3 and 4.4 that

(4.21) \( E(\xi(G)|G \in \mathcal{E}_k, \ell - \mathcal{E}_k^* \) \leq 3\lambda_1\lambda_2. \)

Lemma 4.4, (4.19) and (4.21) then imply that

\[
\sum_{\Gamma \in \Lambda_\xi} \frac{1}{\alpha_\Gamma} (1 + \epsilon_{k,\ell,\Gamma}) \leq A \omega^{3rs-s+1/n} 2^{-r/s}.
\]

Before looking at \( \zeta \) observe that

\[
\sum_{\Gamma \in \Lambda_\xi} \frac{r!}{\alpha_\Gamma} = \frac{s r!}{\alpha},
\]

since we obtain all copies of graphs of the form \( H - x \) in \( K_r \) by taking all copies of \( H \) and deleting an edge. Thus we can write

(4.21) \[
\bar{\zeta}_{k,\ell} = \frac{s \omega^{s-1}}{\alpha} n^{r/s} (1 + \epsilon_{k,\ell})
\]

where \(|\epsilon_{k,\ell}| \leq A \omega^{3rs-s+1/n} 2^{-r/s} \).
Analogously to (4.19) we have

\begin{equation}
E(\zeta(G) \mid G \in \mathcal{S}_{k,\ell}^*) = \sum_{\Gamma \in \Lambda_\zeta} \left( \nu(\Gamma) \frac{r!}{\alpha_\Gamma} \left( \frac{n}{\lambda + n} \right)^m \mu(\Gamma) \right) (1 + \epsilon_{k,\ell,\Gamma})
\end{equation}

where $\Lambda_\zeta$ denotes the set of possible graphs of the form $H_1 \cup H_2 - x$.

**Lemma 4.5**

$\Gamma \in \Lambda_\zeta$ implies $\frac{r}{s} (\mu(\Gamma) + 1) - \nu(\Gamma) \geq 1 + \frac{r \theta_1}{s}$.

**Proof**

If $\Gamma = H_1 \cup H_2 - x$ let $H' = H_1 \cap H_2$. Then

$$\mu(\Gamma) = 2s - \mu(H') - 1$$

and

$$\nu(\Gamma) = 2r - \nu(H').$$

The result now follows from the definition of $\theta_1$.

It follows from (4.22) and Lemma 4.5 that

\begin{equation}
E(\zeta(G) \mid G \in \mathcal{S}_{k,\ell}^*) \leq A n^{2s-1} \frac{r / s (1 - \theta_1)}{\lambda + n}. \quad (4.23)
\end{equation}

For $G \in \mathcal{S}_{k,\ell} - \mathcal{S}_{k,\ell}^*$ we write, analogously to (4.20)
It now follows from Lemmas 4.3 and 4.4 that

\[ E(\xi(G) \mid G \in \mathcal{A}_{k,\ell} - \mathcal{A}^*_{k,\ell}) \leq 3\lambda_1^2\lambda_2. \]

Combining this with (4.23) and \( \theta_1 \leq \frac{1}{r} \) and using Lemma 4.4 we obtain

(4.24)

\[ \tilde{\xi}_{k,\ell} \leq A\omega^{2s-1} n^{r(1-\theta_1)/s}. \]

**Remark:** the above analysis, between here and (4.13) could equally well have been done with (4.13) replaced by \( \pi_{k,\ell} \geq e^{-\lambda_0}. \) This would lead to slightly larger "hidden" constants \( A. \)

Now (4.3) implies

(4.25)

\[ a_{k-1,\ell} \geq a_{k,\ell} \frac{ks(N-m-\tilde{\xi}_{k,\ell})}{m \tilde{\xi}_{k-1,\ell}}. \]

But clearly \( \tilde{\xi}_{k-1,\ell} \leq n^{\tilde{r}} \) and so, using (4.13), \( \pi_{k-1,\ell} \geq e^{-\lambda_0} \) and by the above remark (4.21) and (4.24) hold with \( k \) replaced by \( k - 1. \) But using these estimates now in (4.3) gives
where, \(|\beta_{k,\ell}| = O(n)\) provided

\[
\begin{align*}
(4.26) & \quad \frac{a_{k,\ell}}{a_{k-1,\ell}} = \frac{\lambda}{k} (1 + \beta_{k,\ell}) \\
(4.27) & \quad \theta < \min\{\frac{r\theta_1}{2s}, \frac{2s-r}{s(3rs+1)}\}.
\end{align*}
\]

Note that (4.26) = (2.3) and that this completes the inductive step in the proof of (4.13) for \(k \leq k_0\). For \(k > k_0\) the only thing that changes is that we replace (4.23) by

\[
\begin{align*}
a_{k+1,\ell} \geq \frac{(m-s(k+\ell)) (\xi_{k,\ell}^2 - 2 \zeta_{k,\ell})}{ks(N-m)} a_{k,\ell}
\end{align*}
\]

which enables to use (4.21), (4.24) with \(k\) replaced by \(k+1\). The rest is as before. This completes the proof of (2.3) and the theorem.

**Remark:** we have identified 5 upper bounds (2.9), (3.2), (3.9), (3.10) and (4.27). It turns out that (2.9) and (3.9) are implied by the others.

**References**


