AN EXPONENTIAL BOUND FOR THE PROBABILITY OF NONEXISTENCE OF A SPECIFIED SUBGRAPH IN A RANDOM GRAPH

By

Svante Janson

Tomasz Łuczak

and

Andrzej Ruciński

IMA Preprint Series # 393

March 1988

AN EXPONENTIAL BOUND FOR THE PROBABILITY OF NONEXISTENCE OF A SPECIFIED SUBGRAPH IN A RANDOM GRAPH

SVANTE JANSON*, TOMASZ ŁUCZAK** AND ANDRZEJ RUCIŃSKI***

Abstract. Let $K(n,p)$ be the random graph in which each edge is present independently with the probability $p = p(n)$ and let, for a graph G, $\mathcal{A} = \mathcal{A}(n,p,G)$ be the event that no subgraph of $K(n,p)$ is isomorphic to G. We shall present two proofs of the fact that for positive numbers c_1 and c_2 depending on G only and for all n

$$-(1-p)^{-1}c_1 M \le \log P(\mathcal{A}) \le -c_2 M\ ,$$

where $M = M(n,p,G) = \min\{n^{|H|}p^{e(H)} : H \subseteq G,\ e(H) > 0\}$, and $|H|$, $e(H)$ are the number of vertices and edges in graph H, respectively.

In other words, $\mathrm{P}(\mathcal{A})$ can be bounded below and above exponentially, and both exponents are of the same order of magnitude as the expected number of subgraphs of $K(n,p)$ isomorphic to the least likely to appear in $K(n,p)$ subgraph of G.

1. Introduction. Let $K(n,p)$ be a graph obtained from the complete graph $K(n)$ on vertex set $[n] = \{1,2,\ldots,n\}$ by deleting each edge independently with probability $1-p$. Our aim is to find a fair estimate to the probability that $K(n,p)$ contains no copy of a given graph G, a problem which arises naturally when considering certain properties of $K(n,p)$ (see [5], [9]).

A lower bound can be found at hand. Let $G_1, G_2, \ldots, G_N$, $N = N(n,G)$, be all copies of G in $K(n)$ and let $\mathcal{A}_i$ be the event that $G_i \subseteq K(n,p)$, $i = 1,2,\ldots,N$. Then by the FKG-inequality (see [4], for instance), setting $\overline{\mathcal{A}} = \bigcup \mathcal{A}_i$,

$$\mathrm{P}(\mathcal{A}) \ge \prod_{i=1}^{N} \mathrm{P}(\overline{\mathcal{A}}_i) = \left(1-p^{e(G)}\right)^N \ge \exp\left(-(1-p)^{-1}N(n,G)p^{e(G)}\right)\ ,$$

which can be immediately strengthen to

$$\mathrm{P}(\mathcal{A}) \ge \max_{H\subseteq G,\, e(H)>0} \exp\left(-(1-p)^{-1}N(n,H)p^{e(H)}\right)\ . \tag{1.1}$$

The above inequality can be also expressed in terms of random variables. Let I_i be the indicator of event $\mathcal{A}_i$, $i = 1,2,\ldots,N$, and let $X_{n,p}(G)$ count the copies of G in $K(n,p)$. Then (1.1) becomes

$$\mathrm{P}(X_{n,p}(G) = 0)) \ge \exp\left(-(1-p)^{-1} \min_{H\subseteq G,\, e(H)>0} \mathrm{E}\, X_{n,p}(H)\right)\ . \tag{1.2}$$

---

*Department of Mathematics, Uppsala University, Uppsala, Sweden.

**Institute for Mathematics and its Applications, University of Minnesota, Minneapolis, U.S.A. and Institute of Mathematics, Adam Mickiewicz University, Poznań, Poland.

***Institute of Mathematics, Adam Mickiewicz University, Poznań, Poland.

However for many random graph problems a similar upper bound would be welcome. For instance the probability of coloring the vertices of $K(n,p)$ with r colours so that no triangle is monochromatic is bounded from above by $r^n \, \mathrm{P}(K(\lfloor n/r \rfloor, p) \not\supset K(3))$; the probability that one can find at least cn disjoint copies of G in $K(n,p)$ is less than $2^n \, \mathrm{P}(K(n - \lfloor |G| cn \rfloor, p) \not\supset G)$; the probability that every vertex of $K(n,p)$ is in a copy of G is at least $1 - n\,\mathrm{P}(X^1_{n,p}(G) = 0)$, where $X^1_{n,p}$ counts copies of G which contain a given vertex.

The hope for such a bound was planted on the following small piece of evidence. Let $d(G) = e(G) / |H|$ and call G strictly balanced if $d(H) < d(G)$ for all $H \subsetneq G$.

THEOREM 0 [1,3]. *If G is strictly balanced and $np^{d(G)} \to c > 0$ then $X_{n,p}(G)$ converges to a Poisson distribution and, in particular,*

$$\lim_{n\to\infty} \mathrm{P}(X_{n,p}(G) = 0) = \exp\left(-\lim_{n\to\infty} \mathrm{E}\, X_{n,p}(G)\right) . \tag{1.3}$$

The breakthrough came during the Seminar "Random Graphs '87" held in Poznań, when two quite different methods of deriving an exponential upper bound for $\mathrm{P}(X_{n,p}(G) = 0)$ were discovered. We present them both here, in Sections 3 and 4.

Our main result is the following:

THEOREM 1. *For every graph G with at least one edge there is a constant $c = c(G) > 0$ such that for all p and all n*

$$\mathrm{P}(X_{n,p}(G) = 0) \le \exp\left(-c \min_{H \subseteq G,\, e(H) > 0} \mathrm{E}\, X_{n,p}(H)\right) .$$

In the next section we shall give a reformulation of Theorem 1 together with an asymptotic formula for $\log \mathrm{P}(X_{n,p}(G) = 0)$ under more restrictive conditions. This will be followed by a geometric description of the set of subgraphs H of G which minimize $\mathrm{E}\, X_{n,p}(H)$ for some range of $p = p(n)$.

Let us mention that each of the authors is responsible for one of Sections 2–4 and for those familiar with the random graphs mini-world it will not be hard to solve this puzzle. For others it is of no importance at all.

Actually, it is even a more collective paper, since other people contributed to it as well. They are Michał Karoński, the chief-organizer of RG'87, Béla Bollobás, whose opening lecture at RG'87 contained the ideas utilized in Section 4, and Joel Spencer, a participant of RG'87, who discussed with us the details of the paper for several times. (We joke that if our paper is a child of RG'87 then Joel Spencer should be its Godfather.)

Finally, we should acknowledge that Béla Bollobás, during his talk in Egĕr, on the Hungarian conference which was held week after that in Poland, announced he was able to prove our bound by martingale method when G is complete. Recently Boppana and Spencer [2] proved our Corollary 1 directly from the FKG-inequality. In fact, Spencer developed that argument further to gain our main result (a personal communication).

2. Leading overlaps. First, let us formulate the equivalent asymptotic version of our main result. Given a sequence $p = p(n)$, we call $H \subseteq G$ a *leading overlap* of G if $e(H) > 0$ and

$$\mathrm{E}\, X_{n,p}(H) = O(\mathrm{E}\, X_{n,p}(H')) \quad \text{for all } H \neq H' \subseteq G,\ e(H') > 0\ . \tag{2.1}$$

THEOREM 1'. *If H is a leading overlap of G then there exists a constant $c = c(G)$ such that for all n*

$$\mathrm{P}(X_{n,p}(G) = 0) \leq \exp(-c\,\mathrm{E}\, X_{n,p}(H))\ .$$

The name "leading overlap" comes from the following calculation, which shows that the leading term in an expansion of $\mathrm{Var}X_{n,p}(G)$ corresponds to the pairs of copies of G that overlap in a leading overlap. With I_i defined in Introduction, we write $i \sim j$ when G_i and G_j have a common edge. Then

$$\mathrm{Var}X_{n,p}(G) = \sum\sum_{i \sim j} \mathrm{Cov}\,(I_i, I_j) \leq \sum\sum_{i \sim j} \mathrm{E}\, I_i I_j\ . \tag{2.2}$$

We partition the last sum according to the shape of $G_i \cap G_j$, the graph spanned by the common edges in G_i and G_j.

If H is a subgraph of G, let $\#\{H \subseteq G\}$ denote the number of copies of H in G. There are $N(n, G)\#\{H \subseteq G\}$ pairs (G', H') with $H' \subseteq G' \subseteq K(n)$, where H' and G' are copies of H and G, respectively. Hence, each of the $N(n, H)$ copies of H in $K(n)$ is contained in exactly $\nu = N(n, G)\#\{H \subseteq G\} \,/\, N(n, H)$ copies of G and, consequently, if H' is one of them, there are at most ν^2 pairs (i, j) with $G_i \cap G_j = H'$. Since for each such a pair $\mathrm{E}\, I_i I_j = p^{2e(G)-e(H)}$, we obtain

$$\sum\sum_{i \sim j} \mathrm{E}\, I_i I_j \leq \sum_H N(n, H)\nu^2 p^{2e(G)-e(H)} = \sum_H \frac{(\mathrm{E}\, X_{n,p}(G))^2}{\mathrm{E}\, X_{n,p}(H)} \#\{H \subseteq G\}^2\ , \tag{2.3}$$

where we sum over all pairwise nonisomorphic subgraphs of G with at least one edge. If we let $n \to \infty$ then the inequality in (2.3) can be replaced by "$= (1 + o(1))$". We write $a_n \asymp b_n$ if $a_n = O(b_n)$ and $b_n = O(a_n)$ simultaneously. The inequality in (2.2) can be replaced by "$\asymp$", provided $\limsup p(n) < 1$ and by "$= (1 + o(1))$", provided $p(n) \to 0$. Thus, if $\limsup p(n) < 1$, H is a leading overlap of G iff

$$\mathrm{Var}(X_{n,p}(G)) \asymp \frac{(\mathrm{E}\, X_{n,p}(G))^2}{\mathrm{E}\, X_{n,p}(H)}\ . \tag{2.4}$$

Moreover, if $p(n) \to 0$ then H is a *unique leading overlap* of G (*u.l.o.*), i.e. (2.1) holds with $o(\cdot)$ instead of $O(\cdot)$, iff

$$\mathrm{Var}(X_{n,p}(G)) = (1 + o(1))\frac{(\mathrm{E}\, X_{n,p}(G))^2}{\mathrm{E}\, X_{n,p}(H)}\ . \tag{2.5}$$

(Note that if H is a u.l.o. then $\#\{H \subseteq G\} = 1$.) In this special case we obtain a considerably sharper result.

THEOREM 2. *If H is a unique leading overlap of G and $p = p(n) \to 0$ as $n \to \infty$ then*

$$\mathrm{P}(X_{n,p}(G) = 0) = \exp\left(-(1 + o(1))\,\mathrm{E}\,X_{n,p}(H)\right) . \tag{2.6}$$

We leave it as an open problem to find a constant $c = c(G)$ such that (2.6) holds in general case with $1 + o(1)$ replaced by $c + o(1)$. To get a flavour of possible solution to it, consider the case when G is a triangle plus a pendant edge hanging from it and $np = 1$. Then both G and the triangle are leading overlaps and $\mathrm{E}\,X_{n,p}(G) = \frac{1}{2} + o(1)$, whereas it follows from [6, Cor. 1b, Section 5] that $\mathrm{P}(X_{n,p}(G) = 0) = \exp\left(-(1 - e^{-1})/6\right) + o(1)$.

In view of our results it is important, and also interesting on its own, to know how the leading overlaps of G change as p increases (i.e. as the rate of decay of $p(n)$ decreases). We call the ordered collection of them the *spectrum* of G. It is, maybe, a bit surprising that the spectrum of G can be derived from the structure of G without any reference to the probabilistic space.

Let $\Omega = \Omega(G)$ be the set $\{(|H|, e(H)) : H \subseteq G,\ |H| \geq 2\}$ of points in the plane. Note that $\Omega = \{(k, l) : 2 \leq k \leq v,\ 0 \leq l \leq e_k\}$, where $v = |G|$ and $e_k = \max\{e(H) : H \subseteq G,\ |H| = k\}$. The points (k, e_k) correspond to some induced subgraphs of G. We denote these "top" points by T_k.

Let $C(\Omega)$ be the convex hull of Ω and let Γ be the top part of the boundary of $C(\Omega)$. Thus Γ is a concave polygonal curve with endpoints $T_2 = (2, 1)$ and $T_v = (|G|, e(G))$.

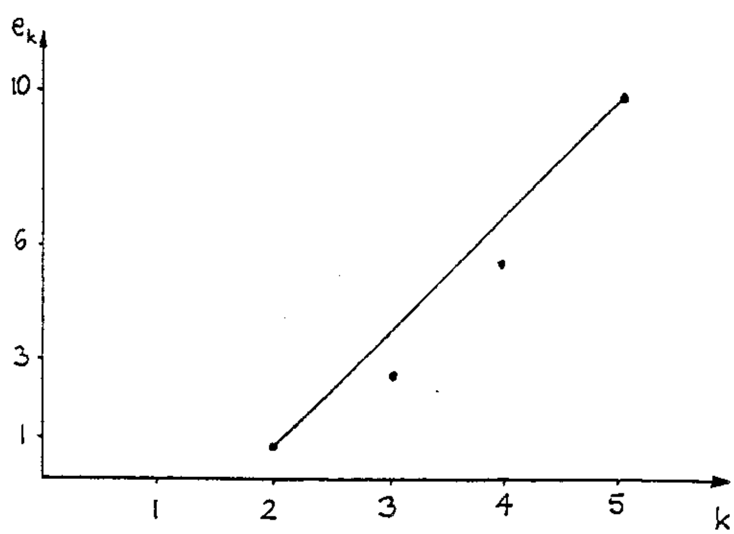
Let $\Gamma_1 = \{k : T_k \in \Gamma\}$ and let $\mathcal{H}_k = \{H : H \subseteq G,\ |H| = k,\ e(H) = e_k\}$. Then, the spectrum of G is the sequence $\{\mathcal{H}_k\}$, where k ranges over Γ_1 in the reverse order. If two (or more) points T_j and T_k lie in the interior of the same line segment of Γ, then we contract $\mathcal{H}_j$, $\mathcal{H}_k$ to $\mathcal{H}_j \cup \mathcal{H}_k$ (see Example 3 below). Note that the first and the last elements of the spectrum always are $\mathcal{H}_v = \{G\}$ and $\mathcal{H}_2$ – the edge-set of G.

Example 1. $K(5)$.

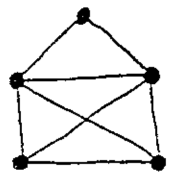
$G = K(5)$

$e_k = \binom{k}{2}$, $k = 2, 3, 4, 5.$

$\mathrm{Spectrum}(G) = (G, \mathcal{H}_2)$

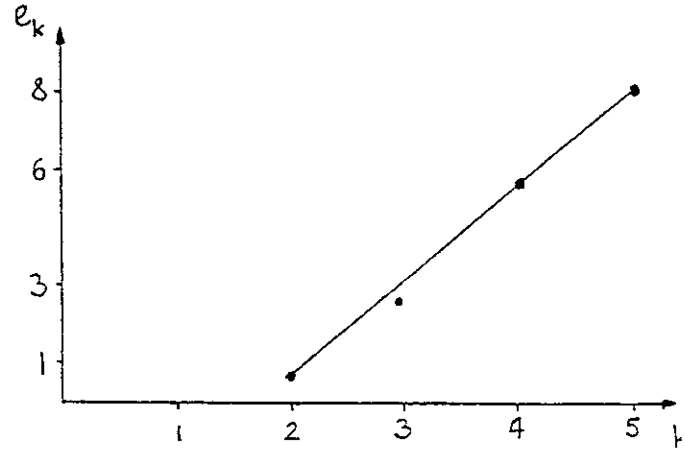


Example 2. The house of Santa Claus.

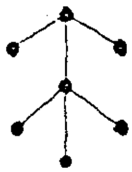
$G =$ 

$e_2 = 1,\ e_3 = 3,\ e_4 = 6,\ e_5 = 8.$

Spectrum$(G) = (G, K(4), \mathcal{H}_2)$

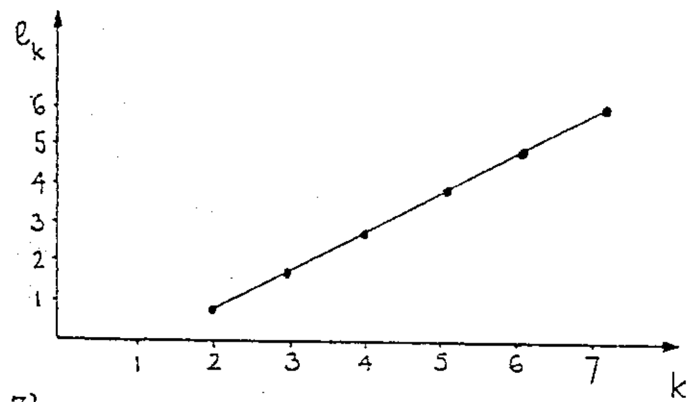


Example 3. Christmas tree.

$G =$ 

$e_k = k - 1,\ k = 2, 3, \ldots, 7.$

Spectrum$(G) = (G, \mathcal{T}, \mathcal{H}_2)$,
where $\mathcal{T} = \{H \subseteq G : H \text{ is a tree, } 2 < |H| < 7\}$.



Actually, the above diagrams bring more information about leading overlaps than just the spectrum. For $k \in \Gamma_1$ let a_k^+ $[a_k^-]$ be the slope of the line segment in Γ that joins T_k to the next point T_j to the right [left] in Γ, set $a_2^- = \infty$ and $a_v^+ = 0$, for convenience. Thus $a_k^- \geq a_k^+ \geq 0$ for all $k \in \Gamma_1$.

The numbers a_k^- and a_k^+ tell us precisely for what range of $p = p(n)$ the elements of the spectrum become leading overlaps. Indeed, given $p = p(n)$, H is a leading overlap iff $H \in \mathcal{H}_k$ for some $k \in \Gamma_1$ such that

$$np^{a_k^-} = O(1) \qquad \text{and} \qquad \left(np^{a_k^+}\right)^{-1} = O(1)$$

(here $p^\infty = 0$).

Calculating the slopes in our examples one can see now that:

— in Example 1, the leading overlaps are G iff $np^3 = O(1)$ and all edges iff $1 = O(np^3)$;

— in Example 2, the leading overlaps are G iff $np^2 = O(1)$, $K(4)$ iff $1 = O(np^2)$ but $np^{5/2} = O(1)$, and all edges iff $1 = O\left(np^{5/2}\right)$;

— in Example 3, the leading overlaps are only G when $np \to 0$, all subtrees when $np \asymp 1$, and all edges when $1 = O(np)$.

Further properties of the spectrum can be read from Ω and Γ in the same manner; we mention now a few. Let

$$\alpha = \alpha(G) = a_v^- = \min\left\{\frac{e(G)-e(H)}{|G|-|H|} \ : \ H \subsetneq G,\ e(H) > 0\right\} . \tag{2.7}$$

G itself remains a u.l.o. as long as $np^\alpha = o(1)$. If $np^\alpha \asymp 1$ then G is still a leading overlap but no longer unique. If $np^\alpha \to \infty$ arbitrarily slowly (i.e. $np^{\alpha+\epsilon} = o(1)$ for all $\epsilon > 0$), then the smallest H which yield the minimum in (2.7) are leading overlaps and only these.

Note that $d(G)$ is the slope of the line from the origin to T_v. Thus G is strictly balanced iff $\Omega \setminus \{T_v\}$ lies below that line, the latter equivalent to $\alpha(G) < d(G)$. Hence Theorem 2 implies (1.3) from Theorem 0. Also for G strictly balanced the range of $p = p(n)$ for which (1.3) is valid coincides with that in which $X_{n,p}(G)$ is Poisson convergent [7].

Let $m = m(G) = \max\{d(H) : H \subsetneq G\}$ and call H *extreme* if $d(G) = m(G)$. Then $m(G)$ is the slope of the upper tangent from the origin to $C(\Omega)$ and the extreme subgraphs correspond to the points T_k lying on the tangent.

Bollobás [1] has shown (see [6] for several proofs) that

$$\mathrm{P}(X_{n,p}(G) = 0) \to \begin{cases} 1 & \text{if} \quad np^{m(G)} \to 0 \\ 0 & \text{if} \quad np^{m(G)} \to \infty . \end{cases} \tag{2.8}$$

It is easy to observe that if $np^m \to 0$ [∞] arbitrarily slowly then the largest [smallest] extreme subgraph of G is a leading overlap. At the threshold, i.e. when $np^{m(G)} \asymp 1$, all extreme subgraphs are leading overlaps, and so, Theorem 1′ implies (2.8).

Let, assuming $|G| > 2$,

$$\beta = \beta(G) = a_2^+ = \max\left\{\frac{e(H)-1}{|H|-2} \ : \ H \subseteq G\ ,\ |H| > 2\right\} .$$

Then, as soon as $1 = O\left(np^\beta\right)$, all edges of G become leading overlaps and remain such till the very end of the evolution of $K(n,p)$, i.e. when $p \equiv 1$. Always $\alpha \le \beta$, and we call G *K_2-balanced* if $\alpha(G) = \beta(G)$, i.e. if Γ is a straight line between T_2 and T_v. If no other point T_k lies on this line, i.e. if the spectrum is $(G, \mathcal{H}_2)$, we say that G is *strictly K_2-balanced*. If this is not case (but still $\alpha = \beta$) there may be other leading overlaps besides G and the edges of G but only in the short range of $p = p(n)$ when $np^\alpha = np^\beta \asymp 1$.

Complete graphs and cycles are strictly K_2-balanced, whereas trees are K_2-balanced but not strictly K_2-balanced. Generally, with the exception of vertex-disjoint unions of copies of K_2, a K_2-balanced graph is strictly balanced, but the converse is not true as Example 2 shows.

3. Main result – Proof no.1. We formulate the main estimate for a somewhat more general situations with less symmetry.

Let $\{I_i\}_{i\in J}$ be independent 0-1 variables where J is an arbitrary index set. (The probabilities $\mathrm{P}(I_i = 1)$ may be different.) Define, for every subset α of J, $I_\alpha = \prod_i I_i$.

Let A be a collection of subsets of J and set $S = \sum_{\alpha\in A} I_\alpha$. The random graph problem studied in this paper is obviously of this type (assuming, as we may, that G has no isolated vertices); J is the set of edges in $K(n)$ and A is the set of copies of G.

We write $\alpha \sim \beta$ if $\alpha\cap\beta \neq \emptyset$.

LEMMA 1. *With notation as above,*

$$\sum_\alpha \log(1-\mathrm{E}\, I_\alpha) \le \log \mathrm{P}(S=0) \le -\sum_\alpha \mathrm{E}\ \frac{I_\alpha}{\sum\limits_{\beta\sim\alpha} I_\beta}\ . \tag{3.1}$$

(We interpret 0/0 as 0 in the right hand side.)

Proof. The left inequality follows by the FKG-inequality. To prove the right one we use a version of Stein's method. Let $\psi(t) = \mathrm{E}\, e^{-tS}$, $t \ge 0$. As t increases, $\psi(t)$ decreases from $\psi(0) = 1$ to $\psi(\infty) = \mathrm{P}(S=0)$, and

$$-\frac{d\psi(t)}{dt} = \mathrm{E}\, S e^{-tS} = \sum_\alpha \mathrm{E}\, I_\alpha e^{-tS}$$

We write, for each α, $S'_\alpha = \sum\limits_{\beta\sim\alpha} I_\beta$ and $S''_\alpha = \sum\limits_{\beta\not\sim\alpha} I_\beta$. Thus S''_α is independent of I_α. Since $S = S'_\alpha + S''_\alpha$, we have

$$\mathrm{E}(I_\alpha e^{-tS}) = \mathrm{P}(I_\alpha = 1)\,\mathrm{E}(e^{-tS'_\alpha} e^{-tS''_\alpha} | I_\alpha = 1).$$

The condition $I_\alpha = 1$ fixes $I_i, i\in\alpha$. Since $e^{-tS'_\alpha}$ and $e^{-tS''_\alpha}$ both are decreasing functions of the remaining $I_i, i\notin\alpha$, the FKG-inequality yields

$$\begin{aligned}\mathrm{E}(I_\alpha e^{-tS}) &\ge \mathrm{P}(I_\alpha=1)\,\mathrm{E}(e^{-tS'_\alpha}|I_\alpha=1)\,\mathrm{E}(e^{-tS''_\alpha}|I_\alpha=1) = \\ &= \mathrm{E}(I_\alpha e^{-tS'_\alpha})\,\mathrm{E}(e^{-tS''_\alpha}|I_\alpha=1) = \\ &= \mathrm{E}(I_\alpha e^{-tS'_\alpha})\,\mathrm{E}(e^{-tS''_\alpha}) \ge \mathrm{E}(I_\alpha e^{-tS'_\alpha})\psi(t)\ .\end{aligned} \tag{3.2}$$

Summing over a we obtain

$$\frac{d(-\log\psi(t))}{dt} = -\frac{1}{\psi(t)}\,\frac{d\psi(t)}{dt} \ge \sum_\alpha \mathrm{E}(I_\alpha e^{-tS'_\alpha})$$

and thus

$$-\log \mathrm{P}(S=0) = \int_0^\infty \frac{d(-\log\psi(t))}{dt}\,dt \geq \sum_\alpha \mathrm{E} \int_0^\infty I_\alpha e^{-tS'_\alpha}\,dt = \sum_\alpha \mathrm{E}\,\frac{I_\alpha}{S'_\alpha}\,.$$

This completes the proof of the lemma. □

For applications, it is convenient to replace the right hand side by a more easily computable expression. For example, since $I_\alpha = 1$ implies that

$$\left(\sum_{\beta\sim\alpha} I_\beta\right)^{-1} \geq 1 - \frac{1}{2} \sum_{\substack{\beta\sim\alpha\\ \beta\neq\alpha}} I_\beta$$

we have the following estimate.

Corollary 1. *With notation as above*

$$\log \mathrm{P}(S=0) \leq -\sum_\alpha (\mathrm{E}\,I_\alpha - \frac{1}{2}\,\mathrm{E} \sum_{\substack{\beta\sim\alpha\\ \beta\neq\alpha}} I_\alpha I_\beta) = -\,\mathrm{E}\,S + \frac{1}{2} \sum\sum_{\substack{\alpha\sim\beta\\ \alpha\neq\beta}} \mathrm{E}\,I_\alpha I_\beta\;. \tag{3.3}$$

(A simpler proof of this result is given in [2]).

For our theorem this is, however, not enough (unless G is a unique leading overlap). Instead we use Jensen's inequality, and obtain

$$\begin{aligned}\mathrm{E}\,\frac{I_\alpha}{\sum_{\beta\sim\alpha} I_\beta} &= \mathrm{E}\,I_a\,\mathrm{E}\left(\frac{1}{\sum_{\beta\sim\alpha} I_\beta}\,|I_\alpha = 1\right)\\ &\geq \mathrm{E}\,I_\alpha \frac{1}{E(\sum_{\beta\sim\alpha} I_\beta | I_\alpha = 1)} = \frac{(\mathrm{E}\,I_\alpha)^2}{\sum_{\beta\sim\alpha} \mathrm{E}\,I_\alpha I_\beta}\,.\end{aligned}$$

In symmetric situations where $\mathrm{E}\,I_\alpha$ is independent of α this yields (using the arithmetic-harmonic inequality if the denominator in the right hand side of (3.4) depends on α)

$$\log \mathrm{P}(S=0) \leq -\frac{(\mathrm{E}\,S)^2}{\sum\sum_{\alpha\sim\beta} \mathrm{E}\,I_\alpha I_\beta} \tag{3.5}$$

Returning to the random graphs, we obtain by (3.5) and (2.3)

$$\log \mathrm{P}(X_{n,p}(G) = 0) \leq -\left(\sum_H \frac{1}{\mathrm{E}\,X_{n,p}(H)}\,\#\{H \subseteq G\}^2\right)^{-1} \tag{3.6}$$

Thus the main theorem follows.

4. Main result – Proof no.2. We shall only treat the case when H is a u.l.o. of G and so $p(n) \to 0$ – one can easily modify our arguments to obtain the general case.

Let us start with introducing some notation. For $b = b(n)$, $r = r(n)$, $0 \le r \le b \le 1$ the *blue-red pattern* $K(n, b, r)$ is a structure obtained by colouring edges of a graph $K(n, b)$ blue with probability $1 - r/b$ and red with probability r/b. A subgraph G' is contained in $K(n, b, r)$ if the set of edges of G' is contained in the set of all edges of $K(n, b, r)$. We shall call such a subgraph 1-*blue* if exactly one from its edges is blue and $e(G') - 1$ other are red.

Now suppose that H is a u.l.o. of G for $r = r(n)$. Then H is also a u.l.o. for itself so (2.5) implies that the expectation $\mathrm{E}' X_{n,r}(H)(X_{n,r}(H) - 1)$ of the number of pairs of different subgraphs isomorphic to H which have at least one edge in common is of the order less than $\mathrm{E}\, X_{n,r}(H)$. Thus we can always find a function $b = b(n)$ such that $b/r \to \infty$ and

$$\frac{\mathrm{E}' X_{n,r}(H)(X_{n,r}(H) - 1)b}{\mathrm{E}\, X_{n,r}(H) r} \to 0 \ . \tag{4.1}$$

The following result is crucial for our arguments.

Lemma 2. *Let $r = r(n)$, $b = b(n)$ be such that $\frac{b}{r} \to \infty$ a graph H is a u.l.o. of G for $r(n)$ and (4.1) holds. Let $X_{n,b,r}$ denote the number of blue edges in $K(n, b, r)$ which are contained in some 1-blue copy of $G \subseteq K(n, b, r)$. Then for a function $\omega = \omega(n)$ which tends to infinity slowly enough we have*

$$\mathrm{P}\left(\left|X_{n,b,r} - \frac{e(H)b}{r}\,\mathrm{E}\, X_{n,r}(H)\right| \ge \frac{e(H)b}{\omega r}\,\mathrm{E}\, X_{n,r}(H)\right) \le 2\exp\left(-\frac{e(H)b}{\omega^4 r}\,\mathrm{E}\, X_{n,r}(H)\right) \ .$$

Proof of Lemma 2. Let $\mathcal{E} = \mathcal{E}_n$ be a the partition of the edges of complete graph K_n into sets of $k = k(n) = \lfloor \frac{\binom{n}{2} r}{\omega^2\, \mathrm{E}\, X_{n,r}(H) b} \rfloor$ or $k + 1$ elements and let $Y = Y_{n,b,r}$ be the number of elements of the maximal family $\mathcal{H}$ of subgraphs of $K(n, b, r)$ with three following properties.

(1) each $H' \in \mathcal{H}$ is 1-blue copy of H;

(2) each $H' \in \mathcal{H}$ is contained in some 1-blue copy of $G \subseteq K(n, b, r)$;

(3) all members of family $\mathcal{H}$ are $\mathcal{E}-$*disjoint* i.e. there are no sets $E \in \mathcal{E}$ which have common edges with two different sets $H_1', H_2' \in \mathcal{H}$.

We shall prove that $\mathrm{E}\, Y = (1 + o(1))\frac{e(H)b}{r}\,\mathrm{E}\, X_{n,r}(H)$. Let us find first the expectation of the variable $\overline{Y}$ which counts the numbers of all 1-blue copies of H in $K(n, b, r)$. The expectation of the number of all copies of H in $K(n, b)$ is $\mathrm{E}\, X_{n,b}(H)$. For all pairs of vertices $\{v, w\}$ and all copies of H containing $\{v, w\}$ probability that in $K(n, b, r)$ only

$\{v,w\}$ remains blue whereas all other edges of the copy of H are red is $(1-r/b)(r/b)^{e(H)-1}$. Summing for all pairs of vertices we count each copy of H $e(H)$ times so

$$\mathrm{E}\,\overline{Y} = e(H)\left(1-\frac{r}{b}\right)\left(\frac{r}{b}\right)^{e(H)-1}\mathrm{E}\,X_{n,b}(H) = (1+o(1))\frac{e(H)b}{r}\,\mathrm{E}\,X_{n,r}(H)\ .$$

Thus, it is enough to show that conditions (2) and (3) do not change this expectation very much.

Let fix 1-blue copy $H' \subseteq K(n,b,r)$ and let Z be the number of all 1-blue copies of G contained in $K(n,b,r)$ which contained H'. Then

$$\mathrm{E}\,Z = (1+O(1/n))\#\{H \subseteq G\} n^{|G|-|H|} r^{e(G)-e(H)}$$

Since H is u.l.o. of G then $\mathrm{E}\,Z \to \infty$ as $n \to \infty$. Now observe that the number of "disjoint" 1-blue copies G_1, G_2 of G i.e. such that $G_1 \cap G_2 = H'$ is equal $(\mathrm{E}\,Z)^2(1+o(1))$ whereas the number of pairs of not disjoint G-extensions of H' can be estimated from above by

$$O(1)\sum_{|J|} n^{|G|-|H|} r^{e(G)-e(H)} n^{|G|-|H\cup J|} r^{e(G)-e(H\cup J)}$$

where the sum is taken over all subgraphs J of $G\setminus H$. Since $\mathrm{E}\,X_{n,r}(H) = o(\mathrm{E}\,X_{n,r}(H\cup J))$ all terms of this sum are of the order less than $(\mathrm{E}\,Z)^2$, so $\mathrm{E}\,Z^2 = (\mathrm{E}\,Z)^2(1+o(1))$ and

$$\mathrm{P}(Z>0) \geq \frac{(\mathrm{E}\,Z)^2}{\mathrm{E}\,Z^2} \to 1\ .$$

Now let for fixed 1-blue $H' \subseteq K(n,b,r)$ W be the number of 1-blue $H'' \subseteq K(n,b,r)$ such that H' and H'' are not $\mathcal{E}$–disjoint. Let us denote by W' the number of 1-blue $H'' \subseteq K(n,b,r)$ which has an edge in common with H' and set $W'' = W \setminus W'$. Then from (4.1) we get $\mathrm{E}\,W' \to 0$ and for $\mathrm{E}\,W''$ we obtain immediately

$$\mathrm{E}\,W'' \leq (1+o(1))e(H)|E|\binom{n}{|H|-2} br^{e(H)-1} \leq \frac{O(1)}{\omega^2\,\mathrm{E}\,X_{n,r}(H)} n^{|H|} r^{e(H)} \to 0\ .$$

Thus any fix 1-blue H' a.s. is contained in some 1-blue copy of G in $K(n,b,r)$ and a.s. there are no such 1-blue H'' in $K(n,b,r)$ that H' and H'' are not $\mathcal{E}$–disjoint, so

$$\mathrm{E}\,Y = (1+O(\omega^{-2}))\frac{e(H)b}{r}\,\mathrm{E}\,X_{n,r}(H)$$

for some function $\omega(n) \to \infty$.

Now for $i = 1,2,\ldots,m$, where $m = \binom{n}{2}/k$, let $\overline{X}_i$ be the expectation of the size of the maximal family $\mathcal{H}_i$ for which (1), (2) and (3) are fulfilled in a blue-red pattern $K_i(n,b,r)$ under the condition that both $K_i(n,b,r)$ and $K(n,b,r)$ choose from the set $\cup_{j=1}^{i} E_j$ the same edges and colour them with the same colour (by E_j we denote here the r-th set of the partition $\mathcal{E}$). It is not hard to check that the sequence of random variables $\overline{X}_0 = \mathrm{E}\,Y, \overline{X}_1, \ldots, \overline{X}_m = Y$, is a martingale (see e.g. [8]). Moreover, due to (3), we have

$$|\overline{X}_i - \overline{X}_{i-1}| \leq 1 \quad \text{for} \quad i = 1,\ldots,\frac{\binom{n}{2}}{k}\ .$$

Now apply the following result

LEMMA 3 [8]. *Let a sequence $X_0, X_1, \ldots, X_m = X$ be a martingale with the property*

$$|X_i - X_{i-1}| \leq 1 \quad \text{for} \quad i = 1, \ldots, m$$

and let $\lambda \geq 0$. Then

$$\mathrm{P}(|X - \mathrm{E}\,X| \geq \lambda\sqrt{m}) \leq 2\exp\left(-\frac{\lambda^2}{2}\right) .$$

Thus, setting $\lambda = \frac{e(H)}{\omega^2}\sqrt{\frac{b}{r}\,\mathrm{E}\,X_{n,r}(H)}$ completes the proof of Lemma 2. □

Proof of Theorem 2. Let $p(n)$ be such a function of n that for $b = \omega^6 p$, $r(i) = ip/\omega$, $i = 1, 2, \ldots, \omega$, the assumptions of Lemma 2 are fullfilled. Furthermore, let $\mathcal{A}_i$ denote the event that $K(n, b, r(i))$ contains no red copy of G. Since $p = r(\omega)$ we have

$$\mathrm{P}(K(n,p) \not\supseteq G) = \mathrm{P}(\mathcal{A}_1) \prod_{i=1}^{\omega-1} \mathrm{P}(\mathcal{A}_{i+1}|\mathcal{A}_i) .$$

From Lemma 2 probability that $X_{n,b,r(i)} \geq \left(1 - \frac{1}{\omega}\right) \frac{e(H)b}{r(i)}\,\mathrm{E}\,X_{n,r(i)}(H)$ is at least $1 - \exp\left(-\frac{e(H)b}{\omega^4 r(i)}\,\mathrm{E}\,X_{n,r(i)}(H)\right)$ for all $i = 1, 2, \ldots, \omega$. Thus, $K(n, b, r(i+1))$ can be obtained from $K(n, b, r(i))$ by recolouring blue edges with probability

$$\frac{p}{\omega b - ip} = \frac{p}{\omega b}\left(1 + O\left(\frac{1}{\omega^2}\right)\right) = \frac{\Delta r}{b}(1 + o(1)) \quad \text{where} \quad \Delta r = \frac{p}{\omega} .$$

Hence

$$\mathrm{P}(K(n,p) \not\supseteq G) \leq \prod_{i=1}^{\omega-1} \left[\left(1 - \frac{\Delta r}{b}(1 + o(1))\right)^{(1-\frac{1}{\omega})\frac{e(H)b}{r(i)}\,\mathrm{E}\,X_{n,r(i)}(H)} + 2\exp\left(-\frac{e(H)b}{\omega^4 r(i)}\,\mathrm{E}\,X_{n,r(i)}(H)\right)\right]$$

$$\leq \exp\left(-(1 + o(1)) \sum_{i=1}^{\omega-1} \frac{e(H)}{r(i)}\,\mathrm{E}\,X_{n,r(i)}(H)\Delta r\right) .$$

Since $\omega \to \infty$ we can replace the sum in the first term by the integral. Hence, keeping in mind that $\mathrm{E}\,X_{n,r(i)}(H) = N_H n^{|H|} r(i)^{e(H)}$ we have

$$\sum_{i=1}^{\omega-1} \frac{e(H)}{r(i)}\,\mathrm{E}\,X_{n,r(i)}\Delta r \to \int_0^p \frac{e(H)}{r(i)}\,\mathrm{E}\,X_{n,r(i)}\,dr(i) = \mathrm{E}\,X_{n,p}(H) .$$

Thus

$$\log \mathrm{P}(K(n,p) \not\supseteq G) \leq -(1 + o(1))\,\mathrm{E}\,X_{n,p}(H)$$

and Theorem 2 follows. □

REFERENCES

[1] B. Bollobás, *Threshold functions for small subgraphs*, Math. Proc. Camb. Phil. Soc., 90 (1981), pp. 197–206.

[2] R. Boppana and J. Spencer, *A useful elementary correlation inequality*, draft.

[3] M. Karoński and A. Ruciński, *On the number of strictly balanced subgraphs of a random graph*, In *Graph theory* (Lagów, 1981). Lecture Notes in Math. 1018, Springer Verlag, Berlin–New York (1983), pp. 79–83.

[4] H. Kesten, *Percolation Theory for Mathematicians*, Birkhäuser, Boston, 1982.

[5] T. Luczak, A. Ruciński and B. Voigt, *Ramsey properties of random graphs*, in preparation.

[6] A. Ruciński, *Small subgraphs of random graphs: a survey*, submitted.

[7] A. Ruciński, *When small subgraphs of random graphs are normally distributed*, Prob. Th. & Rel. Fields, to appear.

[8] E. Shamir and J. Spencer, *Sharp concentration of the chromatic number on random graphs $G_{n,p}$*, Combinatorica, 7 (1987), pp. 121–129.

[9] J. Spencer, *Extension properties in random graphs*, in preparation.