Exponential inequalities for the number of triangles in the Erdős-Rényi random graph

A.A. Bystrov\textsuperscript{1} and N.V. Volodko\textsuperscript{2}

Abstract

Upper exponential inequalities for the tail probabilities of the centered and normalized number of triangles in the Erdős-Rényi graph are obtained, where the probability of every edge is fixed. The result is formulated in terms of random fields.

keywords: Erdős-Rényi graph, number of triangles, Höffding-type exponential inequalities

1 Introduction

The Erdős-Rényi graph $G(n, p)$ is a graph of $n$ nodes where each edge is included with probability $p$ independently from every other edge. These objects are widely studied since pioneer work \cite{Erdos60} published sixty years ago. One of the problems of interest is the distribution of the number of certain subgraphs in $G(n, p)$, and, in particular, the number of triangles. The necessary and sufficient conditions for the asymptotic normality of the number of subgraphs can be found in \cite{Janson00}. A great number of authors are interested in the large deviation principle (LDP) for the number of triangles. The rough LDP is derived in \cite{Durrett96} for $p(n) \to 0$, and in \cite{Durrett99} for fixed $p$. See also \cite{Kesten96},\cite{Kesten97}.

In a number of recent works, so-called generalized random graphs are studied, where the edge probability is random and depends on the weights of vertices. In \cite{Kolchin90}, a convergence in distribution of the number of triangles in the generalized random graph to the Poisson law is obtained.

We are interested in obtaining so-called Höffding-type exponential inequalities for the probability tails of the number of triangles in $G(n, p)$.

The classical Höffding’s inequality was firstly formulated for sums of independent bounded random variables in \cite{Hoeffding56}:

\[
\mathbb{P}(n^{-1}(S_n - \mathbb{E}S_n) \geq t) \leq e^{-2nt^2/(b-a)^2}
\]

for all $t$, where $S_n$ is a sum of independent random variables with values in $[a, b]$. Then, similar results were obtained for more complicated objects – $U$- and $V$-statistics, based on the sequences of independent random variables, later – for the sequences, satisfying weak dependence conditions. The technique for deriving

\textsuperscript{1}Novosibirsk State University, 1 Pirogova str. 630090 Novosibirsk Russia, email: bystrov@ngs.ru

\textsuperscript{2}Sobolev Institute of Mathematics, 4 Acad. Koptyug avenue, 630090 Novosibirsk Russia, email: nvolodko@gmail.com
such inequalities for the independent or weakly dependent random variables is well-known, but for the number of subgraphs we need to investigate a model where the elements of the discrete random field with strong dependence of certain form are summed.

The result, most close to our aim, can be found in [7] but only for deviations of the order $n$ (after normalizing) and without exact constant in the exponent:

$$
P(T_n \geq ET_n + \varepsilon n^3 p^3) \leq \exp(-\alpha(\varepsilon)n^2 p^2).
$$

Here $T_n$ is a number of triangles in the graph $G(n, p)$, $\alpha(\varepsilon)$ has no explicit form.

The close result is obtained in [4], for $p$ bounded from below.

We formulate the result in general terms of random fields and the particular case of the number of triangles is deduced as a corollary.

2 Main result

Let $\{\xi_{ijk}\}_{i<j<k}$ be a field of bounded random variables, not necessarily identically distributed. Two different variables are independent if there is only one common index (or no common indices) in triples. If exactly two of three indices in the triples coincide then the covariance has a positive lower bound:

$$
\xi_{i_1j_1k_1}, \xi_{i_2j_2k_2} \text{ are independent if } \text{card}\{i_1, j_1, k_1\} \bigcap \{i_2, j_2, k_2\} \leq 1.
$$

$$
\text{cov}(\xi_{i_1j_1k_1}, \xi_{i_2j_2k_2}) \geq \rho > 0 \text{ if } \text{card}\{i_1, j_1, k_1\} \bigcap \{i_2, j_2, k_2\} = 2;
$$

$$
|\xi_{ijk} - E\xi_{ijk}| \leq C \text{ a.s.} \quad (2.2)
$$

We obtain the exponential inequalities for the distribution tails of the centered and normalised sum

$$
R_n = b_n^{-1/2} \sum_{1 \leq i < j < k \leq n} (\xi_{ijk} - E\xi_{ijk}), \quad (2.3)
$$

where

$$
b_n = \text{Var} \sum_{1 \leq i < j < k \leq n} \xi_{ijk}. \quad (2.4)
$$

**Theorem 2.1.** Under conditions (2.2) the following upper bound holds for $n \geq 7$ and all positive $x$:

$$
P(|R_n| > x) \leq \exp\left\{ -\frac{1}{2e} \left( \frac{x}{C_0} \right)^2 \right\}, \quad (2.5)
$$

where

$$
C_0 = C\left( \frac{3\sqrt{2}}{\rho} \right)^{1/2}. \quad (2.6)
$$
3 Proof

To estimate the two-sided distribution tail of $R_n$ we use the standard method involving power Chebyshev inequality. Our main goal is to establish an appropriate upper bound for an arbitrary even moment $E R_n^{2m}$, $m = 1, 2, \ldots$. First, we estimate the value $b_n$ from below:

$$b_n = \sum_{i_1 < j_1 < k_1, i_2 < j_2 < k_2} \text{cov}(\xi_{i_1j_1k_1}, \xi_{i_2j_2k_2}) \geq$$

$$\geq \sum_{1 \leq i < j < k \leq n} \text{Var}(\xi_{ijk}) + \rho \frac{n(n - 1)(n - 2)(n - 3)}{2}.$$  

Here $\frac{n(n-1)(n-2)(n-3)}{4}$ is the number of summands with exactly two common indices. Indeed, the number of ways to choose the common pair is $C_n^2$, the third index can be chosen by $(n - 2)$ ways for the first triple and by $(n - 3)$ ways for the second triple.

Further, if $\text{Var}(\xi_{ijk_0}) < \rho$ for some $k_0$ then $\text{Var}(\xi_{ijl}) > \rho$ for all $l \neq k_0$, since

$$\rho \leq \text{cov}(\xi_{ijk_0}, \xi_{ijl}) \leq \sqrt{\text{Var}(\xi_{ijk_0}) \text{Var}(\xi_{ijl})}.$$  

Moreover,

$$\text{Var}(\xi_{ijk_0}) + \text{Var}(\xi_{ijl}) \geq 2 \text{cov}(\xi_{ijk_0}, \xi_{ijl}) \geq 2 \rho.$$

Obviously, it yields that

$$\sum_{1 \leq i < j < k \leq n} \text{Var}(\xi_{ijk}) \geq \rho \frac{C_n^3}{n},$$

and then

$$b_n \geq \rho \left( \frac{n^4}{2} - \frac{17}{6} n^3 + 5n^2 - \frac{8n}{3} \right) = \rho \frac{n^4}{6} + \rho \left( \frac{n^4}{3} - \frac{17}{6} n^3 + 5n^2 - \frac{8n}{3} \right).$$

It is not difficult to prove that the second summand in the last sum is positive when $n \geq 7$. Thus,

$$b_n \geq \frac{\rho n^4}{6}. \quad (3.7)$$

Denote

$$\tilde{\xi}_{ijk} = \xi_{ijk} - E \xi_{ijk},$$

$$\mathbb{E} R_n^{2m} = b_n^{-m} \sum_{i_s < j_s < k_s} \mathbb{E} \tilde{\xi}_{i_1j_1k_1} \cdots \tilde{\xi}_{i_2m_j2m} \cdots \tilde{\xi}_{k_2m_k}, \quad (3.8)$$

Consider a single summand in the sum above. It does not vanish if only for every triple $(i_s, j_s, k_s)$ from the mixed moment another triple can be found with
at least two common indices (otherwise, due to independence a factor of the form \( \mathbb{E}\xi_{i,j,k} \) "kills" the summand). Let us estimate the number of non-zero summands in (3.8).

It follows from the above that, for each triple \((i_s, j_s, k_s)\), a pair of indices must coincide with a pair in another triple and one index can be chosen arbitrarily. We are going to estimate the number of non-zero summands by estimating, firstly, the number of ways to choose “arbitrary” indices, secondly, the number of ways to choose a number \( l \) of different pairs of indices (later we will sum up over \( l \)), thirdly, the number of ways to choose the multiplicity of each pair (not less than two), and fourthly, the number of ways to choose factors in the summand of (3.8) for each such pair.

The number of ways to choose an arbitrary index and its position in each vector is not greater than \( n^{2m} \). Further we consider only repeating pairs in every triple \((i_s, j_s, k_s)\). Denote by \( l \) the number of different pairs among \( 2m \) pairs of indices, \( l \leq m \). The number of possible pairs of indices equals \( n(n-1)/2 \). Then \( l \) different pairs can be chosen in \( \binom{n}{l} \) ways (hence, \( l \) can not be more than \( n(n-1)/2 \) also). We can order chosen pairs somehow. Further, we need to place these different pairs on \( 2m \) positions so that every pair repeats at least twice.

The number of ways to divide \( 2m \) indistinguishable elements into \( l \) groups of size not less than two equals

\[
\binom{2m-1}{2m-l-1}.
\]

Here the cardinality of each of \( l \) groups is a number of times the corresponding pair of indices (we ordered them somehow, see above) occurs in summand (3.8) (i.e. “the multiplicity” of this pair), not taking into account arbitrary indices. Denote the vector of these multiplicities by \((t_1, ..., t_l)\), \( t_1 + ... + t_l = 2m \).

Now we need to estimate the number of ways to place different \( 2m \) elements (positions, corresponding to indices \( s \) in (3.8)) into the groups. This number equals to the polynomial coefficient

\[
\frac{(2m)!}{t_1!...t_l!},
\]

and this coefficient reaches its maximum when all multiplicities \( t_s \) are equal, or, if \( 2m \) cannot be exactly divided by \( l \), the difference between them does not exceed 1. Denote by \([x]\) the integer part of \( x \), and \( k = 2m - \left\lfloor \frac{2m}{l} \right\rfloor \cdot l \). Then the maximum of (3.9) w.r.t. \( t_1, ..., t_l \) equals to

\[
\frac{(2m)!}{\left(\left\lfloor \frac{2m}{l} \right\rfloor + 1\right)! \cdot \left(\frac{2m}{l}\right)!^{l-k}}.
\]
Note that
\[(2m)! = \prod_{i=1}^{k} (i+l) (i+2l) \cdots (i+\lfloor 2m/l \rfloor l) \prod_{i=k+1}^{l} (i+l) (i+2l) \cdots (i+(\lfloor 2m/l \rfloor -1)l) \]
\[\leq \prod_{i=1}^{k} \left( l^{\lfloor 2m/l \rfloor} (\lfloor 2m/l \rfloor +1)! \right) \prod_{i=k+1}^{l} \left( l^{\lfloor 2m/l \rfloor -1} \lfloor 2m/l \rfloor ! \right) =
\]
\[= \prod_{i=1}^{k} \left( \lfloor 2m/l \rfloor +1 \right)! \left( \lfloor 2m/l \rfloor ! \right)^{l-k}, \]

hence the maximum in (3.10) does not exceed \(l!/l^{2m-l}\).

Thus, the number of non-zero summands in (3.8), in factors of which there are exactly \(l\) different pairs \((l \leq C_n^2)\), is bounded by
\[n^{2m} \left( \frac{n(n-1)}{2} \right)^l \frac{(2m)!}{l!} \left( \lfloor 2m/l \rfloor +1 \right)! \left( \lfloor 2m/l \rfloor ! \right)^{l-k} C_2^{l-1} \leq n^{2m+2l} \cdot 2^{2m-2l-1} \cdot l^{2m-l} \leq \frac{4^m}{2^{2l+1}} \cdot n^{2m+2l} \cdot l^{2m-l}. \]

Here we use elementary inequalities \(C_a^b \leq a^b\) and \(C_a^b \leq 2^n\).

As it was explained, \(l\) can not exceed neither \(C_n^2\) nor \(m\). Denote \(M = \min\{m, C_n^2\}\).

Taking into consideration (2.2), (3.7) and the normalization in (3.8), we have an estimate for \(n \geq 7\):
\[\mathbb{E} R_n^{2m} \leq \left( \frac{24C^2}{\rho} \right)^m \sum_{l=1}^{M} \frac{l^{2m-l}}{2^{2l+1} n^{2m-2l}} \leq \left( \frac{12C^2}{\rho} \right)^m \sum_{l=1}^{M} \frac{l^m}{2^{2l+1} n^2} \left( \frac{2l}{n^2} \right)^{m-l}. \quad (3.11) \]

This estimate is sufficient to derive the exponential inequalities for probability tails of \(R_n\), but we are going to make a few additional steps to get a sharper constant.

Divide the sum from 1 to \(M\) into three subsums:
\[\sum_{l=1}^{M} = \sum_{l=1}^{[M/2]} + \sum_{l=\lfloor M/2 \rfloor +1}^{[m/2]} + \sum_{l=\lfloor m/2 \rfloor +1}^{M} \cdot \]

\([\cdot]\) here is the ceiling function. Note that the middle sum vanishes if \(M = m\) and \(m\) is even. In the first subsum \(2l/n^2\) is less than 1/2, in the second and third sums we can use an estimate \(2l/n^2 \leq 1\). Continue estimating of (3.11):
\[\mathbb{E} R_n^{2m} \leq \left( \frac{12C^2}{\rho} \right)^m \left[ \sum_{l=1}^{[M/2]} \frac{l^m}{2^{2l+1}} + \sum_{l=\lfloor M/2 \rfloor +1}^{[m/2]} \frac{l^m}{2^{2l+1}} + \sum_{l=\lfloor m/2 \rfloor +1}^{M} \frac{l^m}{2^{2l+1}} \right] \leq \]
\[
\left( \frac{12C^2}{\rho} \right)^m \left[ \frac{m^m}{2^{2m+1}} + \left( \left\lceil \frac{m}{2} \right\rceil \right)^m \cdot \frac{1}{2^{\lfloor M/2 \rfloor + 1}} + \frac{m^m}{2^{\lceil m/2 \rceil + 1}} \right].
\]

(3.12)

If \( m \) is even then
\[
\left( \left\lceil \frac{m}{2} \right\rceil \right)^m = \left( \frac{m}{\sqrt{2}} \right)^m \frac{1}{2^{m/2}} \quad \text{and} \quad \frac{m^m}{2^{\lceil m/2 \rceil + 1}} = \frac{1}{2} \left( \frac{m}{\sqrt{2}} \right)^m.
\]

If \( m \geq 3 \) is odd then
\[
\left( \left\lceil \frac{m}{2} \right\rceil \right)^m \leq \left( \frac{m}{\sqrt{2}} \right)^m \quad \text{and} \quad \frac{m^m}{2^{\lceil m/2 \rceil + 1}} = \frac{1}{2^{3/2}} \left( \frac{m}{\sqrt{2}} \right)^m.
\]

Thus, the second factor of (3.12) can be estimated by \( \left( \frac{m}{\sqrt{2}} \right)^m \) excluding the case \( M = m = 1 \), but in this case the estimate follows immediately from (3.11).

Finally,
\[
\mathbb{E} R^2 \leq \left( \frac{12C^2}{\rho} \right)^m \left( \frac{m}{\sqrt{2}} \right)^m = (C_0)^2 m (2m)^m,
\]

where constant \( C_0 \) depends only on \( C \) and \( \rho \) and is defined in (2.6).

Taking into account an estimate obtained for the moment of an arbitrary even order we can derive the inequality for the two-sided distribution tails of \( R \). The corresponding standard argument can be found, for example, in [1] (Lemma 1, Remark 1):

Let \( \zeta \) be an arbitrary random variable with the finite moments of all orders \( r \geq 0 \), which satisfy the following relations:
\[
\mathbb{E} |\zeta|^r \leq AC_0^r r^{r/2} \quad \text{for all even} \quad r,
\]

(3.13)

where constants \( A \geq 1 \) and \( C_0 > 0 \) do not depend on \( r \). Then for all \( x \geq 0 \) the following estimate is valid:
\[
P(|\zeta| \geq x) \leq A e^{-\frac{x}{2C_0^2}}.
\]

(3.14)

We can use this result with \( \zeta = R \), \( A = 1 \), \( r = 2m \). Theorem is proved.

**Partial case: a number of triangles.**

The number of triangles \( T \) in the Erdős-Rényi graph of \( n \) nodes can be considered as
\[
T = \sum_{1 \leq i < j < k \leq n} \xi_{ijk},
\]

where \( \xi_{ijk} = X_{ij}X_{jk}X_{ik} \) with independent Bernoulli’s \( X_{ij} \) – indicators of the corresponding edges. In this case, for \( k \neq l \)

\[
\text{Var} \, \xi_{ijk} = p^3(1-p^3), \quad \text{cov}(\xi_{ijk}, \xi_{ijl}) = \rho = p^5(1-p), \quad C = \max\{p^3, 1-p^3\}.
\]

(3.15)

**Corollary 3.1.** If \( T \) is a number of triangles in the Erdős-Rényi graph of \( n \) nodes, where each edge is included with probability \( p \) independently from every other edge, then the assertion of Theorem holds for \( R = \left( \text{Var} \, T \right)^{-1/2} (T - \mathbb{E}T) \) with constants \( C \) and \( \rho \) from (3.15).
References

[1] Borisov, I.S., Bystrov, A.A., 2019. Exponential inequalities for the distributions of canonical multiple partial sum processes. Theory Prob. Appl. 64(2), 171–185.

[2] Chatterjee, S., 2011. The missing log in large deviations for triangle counts. Random Struct. Algorithms 40 (4), 437–451.

[3] Chatterjee, S. and Varadhan, S. R. S., 2011. The large deviation principle for the Erdös-Rényi random graphs. Eur. J. Comb. 32 (7), 1000–1017.

[4] DeMarco, B. and Kahn, J., 2012. Upper tails for triangles. Random Struct. & Algorithms 40 (4) 452–459 (2012).

[5] Erdös, P. and Rényi, A., 1960. On the evolution of random graphs. Publ. Math. Inst. Hungar. Acad. Sci. 5, 209–227.

[6] Höffding, W., 1963. Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc. 58, 13–30.

[7] Janson, S., Oleszkiewicz, K. and Ruciński, A., 2004. Upper tails for subgraph counts in random graphs. Israel J. Math. 142, 61–92.

[8] Kim, J. H. and Vu, V. H., 2004. Divide and conquer martingales and the number of triangles in a random graph. Random Struct. Algorithms 24 (2), 166–174.

[9] Liu, Q. and Dong, Z., 2020. Limit laws for the number of triangles in the generalized random graphs with random node weights. Stat. & Probab. Letters 161, 1–6.

[10] Logachov, A. and Mogulskii, A., 2020. Exponential Chebyshev inequalities for random graphons and their applications. Siberian Math. J. 61(4), 880–900.

[11] Ruciński, A., 1960. When are small subgraphs of a random graph normally distributed? Prob. Th. Rel. Fields 78, 1–10.