

Distributions of sparse spanning subgraphs in random graphs

Pu Gao*

Max-Planck-Institut für Informatik
janegao@mpi-inf.mpg.de

Abstract

We describe a general approach of determining the distribution of the number of certain types of spanning subgraphs in the random graph $\mathcal{G}(n, p)$. Using this approach, we reprove the distribution of the number of Hamilton cycles, with a proof that is much shorter than previously known proofs. We also achieve new results on determining the distribution of the number of spanning triangle-free subgraphs and the number of triangle-factors.

1 Introduction

The distributions of subgraphs with fixed sizes in various random graph models have been investigated by many authors. A general approach by Ruciński [6, 7] showed that the numbers of subgraphs with fixed sizes in the binomial model $\mathcal{G}(n, p)$ are asymptotically normal for a large range of p . On the other hand, studies of distributions of subgraphs of sizes growing with n , for example, the spanning subgraphs, are much less common. The first breakthrough is perhaps due to Robinson and Wormald [8, 9] on proving that random regular graphs are a.a.s. Hamiltonian. Based on their work, Janson [3] deduced the limiting distribution of the number of Hamilton cycles in random regular graphs. The distributions of some types of spanning subgraphs (perfect matchings, Hamilton cycles, spanning trees) in random graphs $\mathcal{G}(n, p)$ and $\mathcal{G}(n, m)$ were determined by Janson [4]. These distributions behave significantly differently in $\mathcal{G}(n, m)$ and $\mathcal{G}(n, p)$. It was shown that within a big range of m , the numbers of these spanning subgraphs are asymptotically normally distributed in $\mathcal{G}(n, m)$, whereas in the corresponding $\mathcal{G}(n, p)$ with $p = m/\binom{n}{2}$, these random variables are asymptotically log-normally distributed. This is because the expectations of these variables in $\mathcal{G}(n, m)$ grow very fast as m grows. Therefore, even though the number of edges in $\mathcal{G}(n, p)$ has small deviation, the deviation of these random variables (e.g. the number of perfect matchings) can eventually be very large. This same phenomena was observed by the author [2] while studying the distribution of the number of d -factors in $\mathcal{G}(n, p)$.

In this paper, we take the technique that was used in [2] (for the study of d -factors) and extend and generalise it into a method for studying a broader class of large (spanning) subgraphs. In Section 2, we describe the general method (Theorems 1 and 3) and give conditions under which the distribution of the random variable under investigation will follow a pattern of concentration

*Research supported by the Humboldt Foundation

in $\mathcal{G}(n, m)$ and log-normal distribution in $\mathcal{G}(n, p)$, which we call the *log-normal paradigm* in this paper. The method is also extended to cope with probability spaces of random directed graphs (See Theorem 5). To show the power of the method, we reprove the distribution of the number of Hamilton cycles. The problem on the number of Hamilton cycles has been studied in the past by a few authors. The first investigation was done by Wright for the directed Hamilton cycles in [11] and then the undirected Hamilton cycles in [10]. Even though both proofs in [11] and [10] are based on a similar counting trick, the analysis for the undirected version is much more complicated. The proof for the directed Hamilton cycles was redone by Frieze and Suen [1], probably unaware of the existing work of Wright, using basically the same approach. In [4], Janson reproved the same result for both the undirected and directed versions, using the method of graph decomposition and projection. In this paper, we present a much shorter proof, using our method, for both the directed and undirected versions.

We also present two new results: the distribution of the number of triangle free subgraphs in Section 5, and the distribution of the number of triangle-factors (the spanning subgraphs isomorphic to a collection of vertex disjoint triangles) in Section 6. Their distributions are determined by verifying the conditions given in the theorems in Section 2. In Theorem 1, we state a general approach for proving concentration of any large subgraphs in $\mathcal{G}(n, m)$. The proofs in [2, Theorems 2.3 and 2.4] implicitly follow the approach as described in Theorem 1, though in [2] the setting is only for examining the d -factors. The idea of the proof of Theorem 3 is essentially the same as the proof of [4, Theorem 6] (with necessary modifications), except that we relax some constraints in [4, Theorem 6] so that it is applicable to the study of a larger class of subgraphs. The proofs of both Theorems 1 and 3 are presented in Section 7.

2 A general approach

Let \mathcal{S} denote a set of vertex-labelled graphs on a set $S = [n]$ of n vertices. For two graphs H_1 and H_2 both on vertex set S , let $H_1 \cap H_2$ ($H_1 \cup H_2$) denote the set of edges contained in both (either of) H_1 and H_2 . For any integer $j \geq 0$, let $F_j(\mathcal{S})$ denote the set of ordered pairs $(H_1, H_2) \in \mathcal{S} \times \mathcal{S}$ such that $|H_1 \cap H_2| = j$. Let $f_j = f_j(\mathcal{S}) = |F_j(\mathcal{S})|$ and let $r_j = f_j/f_{j-1}$ for any $j \geq 1$, as long as $f_{j-1} \neq 0$. Let $X_n = X_n(\mathcal{S})$ denote the number of members of \mathcal{S} that are contained in a random graph ($\mathcal{G}(n, p)$ or $\mathcal{G}(n, m)$, defined on the same vertex set S) as (spanning) subgraphs. Here S , p and m refer to sequences $(S(n))_{n \geq 1}$, $(p(n))_{n \geq 1}$ and $(m(n))_{n \geq 1}$. Assume every graph in \mathcal{S} has the same number $h = h(n)$ of edges. Let $N(n) = \binom{n}{2}$. We drop n from all these notations when there is no confusion. All asymptotics in this paper refer to $n \rightarrow \infty$. For any real x and any integer $\ell \geq 0$, define the ℓ -th falling factorial $[x]_\ell$ to be $\prod_{i=0}^{\ell-1} (x - i)$. Let

$$\mu_n = |\mathcal{S}| \binom{N-h}{m-h} / \binom{N}{m}, \quad \lambda_n = |\mathcal{S}| p^h. \quad (2.1)$$

Clearly,

$$\mathbf{E}_{\mathcal{G}(n, m)} X_n = \mu_n, \quad \mathbf{E}_{\mathcal{G}(n, p)} X_n = \lambda_n.$$

A simplification of μ_n (readers can also refer to Lemma 17 by taking $\ell = h$) gives

$$\mathbf{E}_{\mathcal{G}(n, m)} X_n = |\mathcal{S}| \cdot \frac{\binom{N-h}{m-h}}{\binom{N}{m}} = |\mathcal{S}| \cdot \frac{[m]_h}{[N]_h} = |\mathcal{S}| (m/N)^h \exp\left(-\frac{N-m}{mN} \frac{h^2}{2} + O(h^3/m^2)\right). \quad (2.2)$$

Theorem 1 Let μ_n be defined as in (2.1). Assume that $h^3 = o(m^2)$, $h^2 = \Omega(m)$ and for $\rho(n) = h^2/m$ and some function $\gamma(n)$, the following conditions hold:

(a) for all $K > 0$ and for all $1 \leq j \leq K\rho(n)$,

$$r_j = \frac{h^2}{Nj}(1 + o(m/h^2));$$

(b) $r_j \leq m/2N$ for all $4\rho(n) \leq j \leq \gamma(n)$;

(c) $t(n) := \sum_{j > \gamma(n)} f_j = o(\mu_n |\mathcal{S}|)$.

Then in $\mathcal{G}(n, m)$,

$$X_n / \mathbf{E}_{\mathcal{G}(n, m)}(X_n) \xrightarrow{p} 1,$$

as $n \rightarrow \infty$.

Remark: The ratio r_j in condition (a) looks quite restrictive. However, as we will see in the next section, this ratio appears naturally if the edges in \mathcal{S} are distributed randomly (see examples in Sections 3.1 and 3.2). In some cases, for instance, if we take \mathcal{S} to be the set of graphs isomorphic to a given unlabelled graph on n vertices, the edges in \mathcal{S} are likely to still distribute in some kind of “random-like” way and thus having r_j as expressed in condition (a) is expected. If we are lucky, we might have condition (b) satisfied for $\gamma(n) = h$. Then $t_n = 0$ and condition (c) is satisfied trivially. See the example in Section 5. But usually this is not the case, as the sequence r_j might decrease first and increase at its tail. Normally, in these cases, condition (c) is not difficult to verify. See examples in Sections 4 and 6.

Theorem 1 and its proof also gives the following proposition.

Proposition 2 Assume all conditions (a)–(c) of Theorem 1 are satisfied with $m = N$. Then, for all $j = O(h^2/m)$,

$$f_j(n) \sim |\mathcal{S}|^2 \exp(-h^2/N)(h^2/N)^j / j!.$$

The following theorem gives conditions under which X_n will be asymptotically log-normally distributed in $\mathcal{G}(n, p)$ if all conditions in Theorem 1 are satisfied by taking $m = pN$.

Theorem 3 Assume $h^3 = o(p^2 n^4)$. Let $\beta_n = h\sqrt{(1-p)/pN}$ and λ_n as defined in (2.1). Assume further that $\liminf_{n \rightarrow \infty} \beta_n > 0$. If for all $m = pN + O(\sqrt{pN})$, $X_n / \mathbf{E}_{\mathcal{G}(n, m)}(X_n) \xrightarrow{p} 1$, then

$$\frac{\ln(e^{\beta_n^2/2} X_n / \lambda_n)}{\beta_n} \xrightarrow{d} \mathcal{N}(0, 1), \quad \text{as } n \rightarrow \infty,$$

where $\mathcal{N}(0, 1)$ is the standard normal distribution.

Combining Theorems 1 and 3, we immediately have the following corollary.

Corollary 4 Assume $h^3 = o(p^2 n^4)$ and $h^2 = \Omega(p n^2)$. Let $\beta_n = h\sqrt{(1-p)/pN}$ and λ_n as defined in (2.1). Assume further that $\liminf_{n \rightarrow \infty} \beta_n > 0$. If for all $m = pN + O(\sqrt{pN})$, conditions (a)–(c) in Theorem 1 are satisfied, then

$$\frac{\ln(e^{\beta_n^2/2} X_n / \lambda_n)}{\beta_n} \xrightarrow{d} \mathcal{N}(0, 1), \quad \text{as } n \rightarrow \infty,$$

where $\mathcal{N}(0, 1)$ is the standard normal distribution.

Hence, in order to prove that a subgraph count has a log-normal distribution in $\mathcal{G}(n, p)$, it is enough to check conditions (a)–(c) in Theorem 1 by taking $m = pN + O(\sqrt{pN})$ if the value of p and the number of edges in the subgraph h satisfy the hypotheses in Corollary 4. This method is particularly powerful if we can estimate r_j without knowing f_j . This is usually the case when we apply the switching method developed by McKay [5]. As we will see in the later examples, our method is easy to be applied by making extensive use of the switching method.

We can generalise the results to random digraphs. Define $\mathcal{D}(n, m)$ to be the random digraph on n vertices with m directed edges chosen uniformly at random from the $2N$ ordered pairs of vertices. Define $\mathcal{D}(n, p)$ to be the random digraph on n vertices, which includes every directed edge independently with probability p . In this paper, we again define $\mathcal{D}(n, m)$ and $\mathcal{D}(n, p)$ on the vertex set S . With almost the same proofs of Theorems 1 and 3 we have the following theorem.

Theorem 5 The same conclusions of Theorems 1 and 3 hold if we replace $\mathcal{G}(n, m)$, $\mathcal{G}(n, p)$, N by $\mathcal{D}(n, m)$, $\mathcal{D}(n, p)$ and $2N$.

3 Two trivial examples

The purpose of this section is to provide simple demonstrations of our method and to convince readers that the behavior of the ratio r_j as given in Theorem 1 (a) shall be well expected.

3.1 The first trivial example

Take \mathcal{S}_1 to be the set of all graphs on vertex set S with h edges. Then $|\mathcal{S}_1| = \binom{N}{h}$. The conclusion of Theorem 1 should hold trivially in this case as $X_n(\mathcal{S}_1)$ is constant in $\mathcal{G}(n, m)$ (depending only on m and h). Nevertheless we verify conditions (a) and (b), also for later use in the next section. For all $0 \leq j \leq h$,

$$f_j = \binom{N}{j} \binom{N-j}{h-j} \binom{N-h}{h-j}.$$

Then for all $1 \leq j \leq h$,

$$r_j = \frac{(h-j+1)^2}{j(N-2h+j)} = \frac{h^2}{jN} (1 + O(j/h + h/n^2)).$$

This verifies conditions (a) and (b) (for $\gamma(n) = h$).

3.2 Another trivial example

Let $0 < \hat{p} < 1$. Consider the set of graphs \mathcal{S}_2 that is obtained by including each element in \mathcal{S}_1 independently with probability \hat{p} . Note here that \mathcal{S}_2 itself is a random variable. Then we have the following.

Theorem 6 *Assume $0 < \hat{p} \leq 1$, $0 < p < 1$ are reals and h is an integer that satisfy $m = p\binom{n}{2}$, $h^3 = o(m^2)$, $h^2 = \Omega(m)$, $m^2\hat{p}^2N^h \gg h^{3h+4} \ln n$. Let $\mu_n(\mathcal{S}_2)$ and $\lambda_n(\mathcal{S}_2)$ be defined as in (2.1) and let $\beta_n = h\sqrt{(1-p)/pN}$. Then $X_n(\mathcal{S}_2)/\mu_n(\mathcal{S}_2) \xrightarrow{p} 1$ in $\mathcal{G}(n, m)$, and*

$$\frac{\ln(e^{\beta_n^2/2} X_n(\mathcal{S}_2)/\lambda_n(\mathcal{S}_2))}{\beta_n} \xrightarrow{d} \mathcal{N}(0, 1), \quad \text{in } \mathcal{G}(n, p),$$

provided $\liminf_{n \rightarrow \infty} \beta_n > 0$.

Proof. By the definition of \mathcal{S}_2 , we have $s_2 = |\mathcal{S}_2| \sim \mathcal{B}(\binom{N}{h}, \hat{p})$ and for all $0 \leq j < h$, $f_j \sim \mathcal{B}(M_j, \hat{p}^2)$, $f_h \sim \mathcal{B}(M_h, \hat{p})$, where $M_j = \binom{N}{j} \binom{N-j}{h-j} \binom{N-h}{h-j}$. Let A_n denote the family of \mathcal{S}_2 which satisfies

$$\forall 0 \leq j \leq h-1, f_j = (1 + o(m/h^2)) \hat{p}^2 M_j, \quad f_h < 2M_h \hat{p}, \quad s_2 > \binom{N}{h} \hat{p}/2.$$

The Chernoff bound gives that

$$\mathbf{P}(|f_j - \hat{p}^2 M_j| > 2\sqrt{3(\ln n)\hat{p}^2 M_j}) < \exp(-3 \ln n) = n^{-3}, \quad \forall 0 \leq j \leq h-1,$$

and

$$\mathbf{P}(f_h < 2M_h \hat{p}) = 1 - o(1), \quad \mathbf{P}\left(s_2 > \binom{N}{h} \hat{p}/2\right) = 1 - o(1).$$

Therefore, with probability at least $1 - hn^{-3} - o(1) = 1 - o(1)$, for all $0 \leq j \leq h-1$,

$$f_j = \left(1 + O\left(\sqrt{\ln n / \hat{p}^2 M_j}\right)\right) \hat{p}^2 M_j.$$

Note that for all j , $M_j > [N]_h / (h!)^3 > (N/h^3)^h$ and \hat{p} satisfies

$$\hat{p}^2 \gg \frac{h^{4+3h} \ln n}{m^2 N^h}.$$

Thus, a.a.s. for all $0 \leq j \leq h-1$,

$$f_j = (1 + o(m/h^2)) \hat{p}^2 M_j,$$

i.e. $\mathbf{P}(S_2 \in A_n) = 1 - o(1)$. For every $S_2 \in A_n$, by the calculations in Section 3.1, both conditions (a) and (b) (for $\gamma(n) = h-1$) are satisfied whereas condition (c) can be easily verified by noting that $t_n = f_h < 2M_h \hat{p} = o(\mu_n | \mathcal{S}_2)$. The theorem thereby follows. ■

The following is a corollary of Theorem 6 by letting $\hat{p} = 1/2$. Here \mathcal{S}'_2 are no longer random variables. We may consider \mathcal{S}'_2 as elements in A_n in the proof of Theorem 6.

Corollary 7 *Assume $0 < p < 1$ is a real and h is an integer that satisfy $m = p\binom{n}{2}$, $h^3 = o(m^2)$, $h^2 = \Omega(m)$, $m^2 N^h \gg h^{3h+4} \ln n$. Then for almost all subsets \mathcal{S}'_2 of \mathcal{S}_1 , the same conclusions of Theorem 6 hold when \mathcal{S}_2 is replaced by \mathcal{S}'_2 .*

4 A new approach – Hamilton cycles

The most interesting examples of \mathcal{S} are perhaps taking \mathcal{S} as the set of graphs that are isomorphic to a given unlabelled graph H on a set of n vertices. In this section, we investigate the number of Hamilton cycles. In literature, computing the second moment of the number of Hamilton cycles involves heavy analysis, as done by Wright [10, 11], using the inclusion and exclusion and some recursive functions, and by Janson [4], using the graph decomposition and projection. Here, we present a new and much shorter proof.

Let H (H') be a cycle (directed cycle) with length n and \mathcal{S}_3 (\mathcal{S}'_3) to be the set of graphs (directed graphs) on S that are isomorphic to H (H'). Thus, $X_n(\mathcal{S}_3)$ and $X_n(\mathcal{S}'_3)$ count the numbers of undirected and directed Hamilton cycles respectively. It is well known that

$$|\mathcal{S}_3| = (n-1)!/2, \text{ and } |\mathcal{S}'_3| = (n-1)!. \quad (4.1)$$

We have the following theorem for the undirected version.

Theorem 8 *Let $0 < p < 1$ be a real and $0 < m < N$ an integer satisfying $m = pN$ and $p \gg n^{-1/2}$. Let X_n denote the number of Hamilton cycles in $\mathcal{G}(n, m)$ (or $\mathcal{G}(n, p)$). Let $\mu_n = \mathbf{E}_{\mathcal{G}(n, m)} X_n$ and let $\lambda_n = \mathbf{E}_{\mathcal{G}(n, p)} X_n$. Then $X_n/\mu_n \xrightarrow{p} 1$ in $\mathcal{G}(n, m)$. Assume further that $\limsup_{n \rightarrow \infty} p(n) < 1$, then*

$$\frac{\ln(e^{\beta_n^2/2} X_n/\lambda_n)}{\beta_n} \xrightarrow{d} \mathcal{N}(0, 1), \text{ in } \mathcal{G}(n, p),$$

where $\beta_n = \sqrt{2(1-p)/p}$.

Proof. We define two switching operations as follows.

h-switching: Choose an edge $xy \in G_1 \cap G_2$. Then choose edges $x_1y_1 \in G_1 \setminus G_2$, $x_2y_2 \in G_2 \setminus G_1$ such that xyx_1y_1 and xyx_2y_2 are in a cyclic order in G_1 and G_2 respectively. Replace xy and x_1y_1 by xx_1 and yy_1 in G_1 , and replace xy and x_2y_2 by xx_2 and yy_2 in G_2 . The *h-switching* is applicable if and only if

- (a) the six vertices x, y, x_i and y_i for $i = 1, 2$ are all distinct;
- (b) the edges xx_1 and yy_1 are not in G_2 and the edges xx_2 and yy_2 are not in G_1 .

inverse h-switching: Choose a pair of vertices $\{x, y\}$ such that $xy \notin G_1 \cup G_2$. For $i = 1, 2$, choose x_i and y_i such that $xx_i \in G_i$ and $yy_i \in G_i$ and xx_iyy_i is in a cyclic order in G_i . The inverse *h-switching* replaces xx_i and yy_i by xy and x_iy_i in G_i for $i = 1, 2$. The operation is applicable if and only if

- (a') the six vertices x, y, x_i and y_i for $i = 1, 2$ are all distinct;
- (b') the edges xx_i and yy_i are not in $G_1 \cap G_2$ for $i = 1, 2$;
- (c') $x_1y_1 \notin G_2$ and $x_2y_2 \notin G_1$.

For $g \in F_j$, let $N(g)$ be the number of *h-switchings* applicable on g . There are $2j$ ways to choose and label the end vertices of the edge $xy \in G_1 \cap G_2$. For any chosen xy , there are $n - j + O(1)$ ways to choose and label the end vertices of the edge $x_iy_i \in G_i$, where $j + O(1)$ accounts for the

case that $x_i y_i \in G_1 \cap G_2$ and the case that condition (a) is violated. Thus, a rough estimation of $N(g)$ is $2j(n - j + O(1))^2$. The only miscounts are those xy and $x_i y_i$ such that condition (b) is violated. Clearly, the miscount due to the violation of condition (b) is $O(jn)$ because for any chosen xy , there are exactly two choices for $x_1 y_1$ (equivalently $x_2 y_2$), such that either xx_1 or yy_1 is in G_2 (equivalently, either xx_2 or yy_2 is in G_1). Thus, $N(g) = 2jn^2(1 - j/n + O(n^{-1}))^2$.

On the other hand, for $g' \in F_{j-1}$, let $N'(g')$ denote the number of inverse h -switchings applicable on g' . There are $n^2 - O(n)$ ways to choose and label vertices x and y such that $xy \notin G_1 \cup G_2$. For any chosen xy , there are two ways to choose x_i and y_i from G_i for $i = 1, 2$ respectively, such that $xx_i, yy_i \in G_i$ and $xx_i yy_i$ is in a cyclic order in G_i . Thus, $N'(g')$ is approximately $4(n^2 - O(n))$. The only miscounts are those choices that violate conditions (a') or (b') or (c'). There are only $O(n)$ choices of xy so that (a') or (c') can possibly be violated, and there are only $O(jn)$ choices of xy so that (b') can possibly be violated. Therefore, $N'(g') = 4n^2(1 + O(j/n))$.

Hence for all $1 \leq j \leq n/2$,

$$r_j = \frac{4n^2}{2jn^2}(1 + O(j/n)) = \frac{2}{j}(1 + O(j/n)),$$

from which we can easily verify Theorem 1 (a), (b) (for $\gamma(n) = n/2$). The proof will be completed by verifying condition (c). Let G be a Hamilton cycle, and let $\kappa_j(G)$ denote the number of Hamilton cycles that share at least j edges with G . There are $\binom{n}{j}$ ways to choose j edges from G . These chosen edges form $r \leq j$ disjoint paths. Contract each path into a special vertex. The total number of vertices including these special vertices is then $n - j$. There are $(n - j - 1)!/2$ Hamilton cycles on these vertices. For every such Hamilton cycles, expand each special vertex by its corresponding path (there are two ways to expand each special vertex). Then each expanded Hamilton cycle corresponds to a Hamilton cycle that shares at least j edges with G . Thus, for every G ,

$$\kappa_j(G) \leq \binom{n}{j} \frac{(n - j - 1)!}{2} \cdot 2^j < n!2^j/j!.$$

It is then straightforward to verify that

$$\sum_{j \geq n/2} f_j \leq |\mathcal{S}_3| n!2^{n/2}/(n/2)! = o(|\mathcal{S}_3| \mu_n). \blacksquare$$

The same proof, with only slight modification of the switchings that cope with directed edges, works for the directed version (Theorem 9).

Theorem 9 *If all assumptions with N , $\mathcal{G}(n, p)$ and $\mathcal{G}(n, m)$ replaced by $2N$, $\mathcal{D}(n, p)$ and $\mathcal{D}(n, m)$ in Theorem 8 hold, then the same conclusion of Theorem 8 holds (for $\beta_n = \sqrt{(1-p)/p}$ by the definition of β_n in Theorem 3).*

5 triangle-free subgraphs

In this section, we consider another example where \mathcal{S}_4 is the set of all triangle-free graphs on S with h edges and maximum degree at most $\Delta = \Theta(h^{1/3})$. Then $X_n(\mathcal{S}_4)$ counts the number of triangle-free subgraphs with h edges and maximum degree at most Δ .

Theorem 10 Let $0 < p < 1$ be a real and $0 < m < N$ an integer satisfying $m = pN$, $h^2 = \Omega(m)$, $h^3 = o(m^2)$ (or equivalently $h^3 = o(p^2n^4)$) and $h^{8/3} = o(pn^3)$. Assume $\Delta = \Theta(h^{1/3})$ is an integer. Let X_n denote the number of triangle-free subgraphs with h edges and maximum degree at most Δ . Let μ_n and λ_n be defined as in (2.1) and let $\beta_n = h\sqrt{(1-p)/pN}$. Then $X_n/\mu_n \xrightarrow{p} 1$ in $\mathcal{G}(n, m)$, and

$$\frac{\ln(e^{\beta_n^2/2} X_n/\lambda_n)}{\beta_n} \xrightarrow{d} \mathcal{N}(0, 1), \quad \text{in } \mathcal{G}(n, p),$$

provided $\liminf_{n \rightarrow \infty} \beta_n > 0$.

Proof. Recall that $F_j(\mathcal{S}_4) = \{(G_1, G_2) \in \mathcal{S}_4 \times \mathcal{S}_4 : |G_1 \cap G_2| = j\}$. Consider $j \geq 1$ and the classes $F_j(\mathcal{S}_4)$ and $F_{j-1}(\mathcal{S}_4)$. Let K_n denote the complete graph on S . We define two other switchings operating on $\mathcal{S}_4 \times \mathcal{S}_4$ as follows.

s₂-switching: Let x be an edge in $G_1 \cap G_2$. Choose y and z from $K_n \setminus G_1 \cup G_2$, such that

- (a) $G_1 \cup y$ and $G_2 \cup z$ are triangle-free;
- (b) y (z) is not incident with a vertex with degree equal to Δ in G_1 (G_2).

Replace x by y in G_1 and replace x by z in G_2 .

inverse s₂-switching: Let x be an edge in $K_n \setminus G_1 \cup G_2$ such that

- (a') $G_1 \cup x$ and $G_2 \cup x$ are triangle-free;
- (b') In both G_1 and G_2 , x is not incident with a vertex with degree equal to Δ .

Let $y \in G_1 \setminus G_2$ and $z \in G_2 \setminus G_1$. Replace y by x in G_1 and replace z by x in G_2 .

Clearly, an *s₂-switching* converts an element $g \in F_j(\mathcal{S}_4)$ to an element $g' \in F_{j-1}(\mathcal{S}_4)$ and an *inverse s₂-switching* converts an element $g' \in F_{j-1}(\mathcal{S}_4)$ to an element $g \in F_j(\mathcal{S}_4)$ for some $j \geq 1$. For any $g \in F_j(\mathcal{S}_4)$, let $N(g)$ denote the number of *s*-switchings that are applicable on g . Note that in both G_1 and G_2 , the number of vertices with degree equal to Δ is $O(h^{2/3})$. There are j ways to choose x . Given x , the number of ways to choose y and z is $N - O(h + nh^{2/3} + T_1(g))$ and $N - O(h + nh^{2/3} + T_2(g))$ respectively, where $T_i(g)$ denotes the number of 2-paths in G_i , and $O(nh^{2/3})$ bounds the number of forbidden choices such that y (or z) is incident to a vertex with degree equal to Δ . Let $T(g) = \max\{T_1(g), T_2(g)\}$. Clearly $T(g) = O(n\Delta^2) = O(nh^{2/3})$. So $N(g) = j(N - O(nh^{2/3}))^2$. Then $N(g) = jN^2(1 + O(h^{2/3}/n))$. For any $g' \in F_{j-1}(\mathcal{S}_4)$, let $N'(g')$ denote the number of *inverse s'₂-switchings* applicable on g' . Then $N'(g') = (N - O((2h - j + 1) + nh^{2/3} + T(g')))(h - j + 1)^2 = Nh^2(1 + O(h^{2/3}/n + j/h))$. Since $\sum_{g \in F_j(\mathcal{S}_4)} N(g) = \sum_{g' \in F_{j-1}(\mathcal{S}_4)} N'(g')$, we have that for all $j \geq 1$,

$$r_j = \frac{Nh^2}{jN^2}(1 + O(h^{2/3}/n + j/h)) = \frac{h^2}{jN}(1 + o(m/h^2) + O(j/h)). \quad (5.1)$$

Note that $O(h^{2/3}/n) = o(m/h^2)$ because $h^{8/3} = o(pn^3)$. Next we verify conditions (a) and (b) of Theorem 1. For all $j = O(h^2/m)$, $j/h = O(h/m) = o(m/h^2)$ since $h^3 = o(m^2)$. Thus

$$r_j = \frac{h^2}{jN}(1 + o(m/h^2)),$$

which verifies condition (a). By (5.1), for all $j \geq 3h^2/m$,

$$r_j = \frac{h^2}{jN}(1 + o(1)) + O(h/N) \leq \frac{m}{2N},$$

which verifies condition (b) (for $\gamma(n) = h$). ■

6 Triangle-factors

Given a graph G on n vertices where n is a multiple of 3, a subgraph of G consisting of $n/3$ vertex disjoint triangles is called a triangle-factor of G . In this section, we assume $n \equiv 0 \pmod{3}$ and consider H (H') to be the unlabelled graph on n vertices consisting of $n/3$ vertex disjoint triangles (directed triangles). Let \mathcal{S}_5 (\mathcal{S}'_5) denote the set of graphs on S that are isomorphic to H (H'). Then $X_n(\mathcal{S}_5)$ counts the number of triangle-factors and

$$|\mathcal{S}_5| = \frac{n!}{6^{n/3}(n/3)!}, \quad |\mathcal{S}'_5| = \frac{n!}{3^{n/3}(n/3)!}. \quad (6.1)$$

The following theorem determines the limiting distribution of $X_n = X_n(\mathcal{S}_6)$.

Theorem 11 *Let $0 < p < 1$ be a real and $0 < m < N$ an integer satisfying $m = pN$ and $\liminf_{n \rightarrow \infty} p(n) > 0$. Let X_n denote the number of subgraphs that are isomorphic to a set of $n/3$ vertex disjoint triangles. Let $\mu_n = \mathbf{E}_{\mathcal{G}(n,m)} X_n$ and let $\lambda_n = \mathbf{E}_{\mathcal{G}(n,p)} X_n$. Then $X_n/\mu_n \xrightarrow{p} 1$ in $\mathcal{G}(n, m)$. Assume further that $\limsup_{n \rightarrow \infty} p(n) < 1$, then*

$$\frac{\ln(e^{\beta_n^2/2} X_n/\lambda_n)}{\beta_n} \xrightarrow{d} \mathcal{N}(0, 1), \quad \text{in } \mathcal{G}(n, p),$$

where $\beta_n = \sqrt{2(1-p)/p}$.

Remark: Indeed, the condition of $\liminf_{n \rightarrow \infty} p(n) > 0$ can be replaced by $p(n) \geq n^{-\delta}$, for some small constant δ . For instance, we checked that $\delta = 1/16$ works and there is still room for further improvement. However, $p \gg n^{-1/2}$ does not seem to be sufficient. For the purpose of a cleaner presentation, we only consider $\liminf_{n \rightarrow \infty} p(n) > 0$ in the proof. For readers who are interested in improving the condition of p , we give quite tight bounds in Lemmas 13 and 14, and we also point out here that there is plenty of room in the proofs of Lemma 16 and Theorem 11 to improve the range of p .

Almost the same proof of the previous theorem, with slight modifications of the switchings defined in the proof of Theorem 11, concerning the directions of edges, yields the following corresponding theorem for the number of directed triangle-factors.

Theorem 12 *If all assumptions with N , $\mathcal{G}(n, p)$ and $\mathcal{G}(n, m)$ replaced by $2N$, $\mathcal{D}(n, p)$ and $\mathcal{D}(n, m)$ in Theorem 11 hold, then the same conclusion of Theorem 11 holds (for $\beta_n = \sqrt{(1-p)/p}$ by the definition of β_n in Theorem 3).*

For any $(G_1, G_2) \in \mathcal{S}_6 \times \mathcal{S}_6$, the edges in G_1 and G_2 can intersect in two ways. We say $e \in G_1 \cap G_2$ is of type 1 if the triangles $T_i \in G_i$ with $e \in T_i$ for $i = 1, 2$ are distinct. We say e is of type 2 if T_1 and T_2 are on the same vertex set.

Let $F_{\ell,t}$ denote the set of $(G_1, G_2) \in \mathcal{S}_6 \times \mathcal{S}_6$ such that number of edges in $G_1 \cap G_2$ of type 1 and 2 is ℓ and t respectively. Clearly $F_{\ell,t}$ is non-empty only if t is a multiple of 3. Clearly $F_j(\mathcal{S}_6) = \cup_k F_{j-3k,3k}$. Let $f_{\ell,t} = |F_{\ell,t}|$. Then $f_j = \sum_{k=0}^{\lfloor j/3 \rfloor} f_{j-3k,3k}$.

Lemma 13 For any $t \geq 0$ and $\ell \geq 1$ such that $n - 4\ell - 3t - 1 > 0$ and $n - 3\ell - 3t - 12 > 0$,

$$\frac{2(n - 4\ell - 3t - 1)^2}{\ell(n - 3\ell - 3t)^2} \leq \frac{f_{\ell,3t}}{f_{\ell-1,3t}} \leq \frac{2(n - 4\ell - 3t + 4)^2}{\ell(n - 3\ell - 3t - 12)^2}.$$

Proof. We define two switchings operating on $\mathcal{S}_6 \times \mathcal{S}_6$ as shown in Figure 1.

t₁-switching: Take an edge of type 1 in $G_1 \cap G_2$ and label the end vertices x and y . Let u (v) be the vertex that is adjacent to both x and y in G_1 (G_2). Take a triangle T_1 (T_2) in G_1 (G_2) that is distinct from xyu (xyv) which does not contain any edge in $G_1 \cap G_2$. Label the vertices of T_1 (T_2) as $u_1u_2u_3$ ($v_1v_2v_3$). Replace these four triangles in $G_1 \cup G_2$ by $xuu_1, yu_2u_3 \in G_1$ and $xvv_1, yv_2v_3 \in G_2$. The t_1 -switching is applicable only if $v \notin T_1$, $u \notin T_2$ and $T_1 \cap T_2 = \emptyset$. See Figure 1.

inverse t₁-switching: A vertex x is pure if both triangles containing x in G_1 and G_2 do not contain any edge in $G_1 \cap G_2$. Choose a pure vertex x and label its neighbours in G_1 (G_2) as u and u_1 (v and v_1). Then choose another pure vertex y that is distinct from x, u_i and v_i for $i = 1, 2$. Label the neighbours of y in G_1 (G_2) as u_2 and u_3 (v_2 and v_3). Replace these four triangles under consideration by $xyu, u_1u_2u_3 \in G_1$ and $xyv, v_1v_2v_3 \in G_2$.

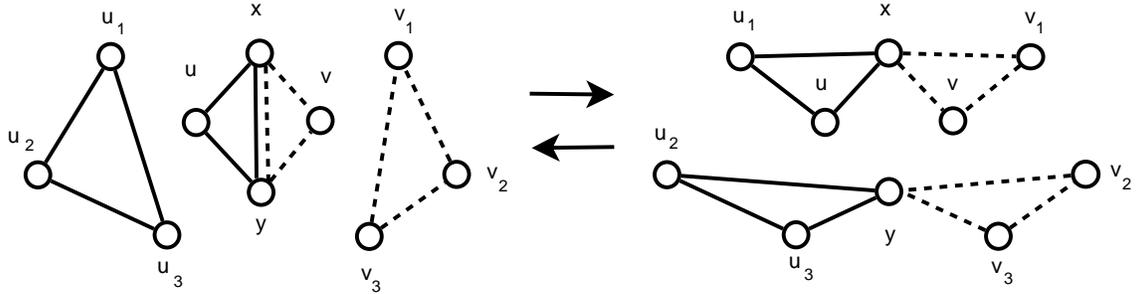


Figure 1: t_1 -switching and its inverse

For any $g = (G_1, G_2) \in F_{\ell,3t}$, let $N(g)$ be the number of t_1 -switchings that are applicable on g . Clearly $N(g) \leq 2\ell(6(n/3 - (\ell + t)))^2$, as there are 2 ways to label x and y for a chosen edge from $G_1 \cap G_2$, and in G_1 (G_2) there are at most $n/3 - (\ell + t)$ choices for the triangle $u_1u_2u_3$ ($v_1v_2v_3$) and for each choice there are 6 ways to label the vertices. We also have

$$N(g) \geq 2\ell \cdot 6(n/3 - (\ell + t) - 1) \cdot 6(n/3 - (\ell + t) - 4),$$

because for any chosen xy , the number of triangles in G_1 which contain no edges in G_2 and do not contain v is at least $n/3 - (\ell + t) - 1$, whereas given the triangle $u_1u_2u_3$, the number of triangles in G_1

which contain no edges in G_1 and do not contain any of $u, u_i, i = 1, 2, 3$ is at least $n/3 - (\ell + t) - 4$. On the other hand, for any $g' = (G_1, G_2) \in F_{\ell-1, 3t}$, let $N'(g')$ be the number of inverse t_1 -switchings applicable on g' . The number of pure vertices is exactly $n - 4(\ell - 1) - 3t$. Hence the number of ways to choose x is $n - 4(\ell - 1) - 3t$ and for any chosen x , the number of ways to label u, u_1, v, v_1 is 4. The number of ways to choose y is $n - 4(\ell - 1) - 3t - \delta$, where δ counts the number of pure vertices among x, u, u_1, v and v_1 . Therefore, $1 \leq \delta \leq 5$ always. Hence,

$$\frac{16(n - 4(\ell - 1) - 3t - 5)^2}{2\ell \cdot (6(n/3 - (\ell + t)))^2} \leq \frac{f_{\ell, 3t}}{f_{\ell-1, 3t}} \leq \frac{16(n - 4(\ell - 1) - 3t)^2}{2\ell \cdot 36(n/3 - (\ell + t) - 4)^2}. \blacksquare$$

Lemma 14 For any $\ell \geq 0$ and $t \geq 1$,

$$\frac{f_{\ell, 3t}}{f_{\ell, 3(t-1)}} = \frac{32(n - 4\ell - 3t)^3}{3(n - 3\ell - 3t)^4} (1 + O(1/(n - 4\ell - 3t))).$$

Proof. We define another two switching operations on $\mathcal{S}_6 \times \mathcal{S}_6$ as shown in Figure 2.

t_2 -switching: Let xyz be a triangle that is contained in both G_1 and G_2 . Take two distinct triangles from G_1 (G_2) which do not contain any edge in $G_1 \cap G_2$ and label the end vertices as $x_1y_1z_1$ and $x_2y_2z_2$ ($x'_1y'_1z'_1$ and $x'_2y'_2z'_2$) respectively. Replace the six triangles under consideration by $aa_1a_2 \in G_1$ and $aa'_1a'_2 \in G_2$, where $a \in \{x, y, z\}$. This switching is applicable only if all these fifteen vertices a, a_i, a'_i for $a \in \{x, y, z\}$ and $i = 1, 2$ are distinct.

inverse t_2 -switching: Recall from the definition of inverse t_1 -switching that a vertex x is pure if both triangles containing x in G_1 and G_2 do not contain any edge in $G_1 \cap G_2$. Choose three pure vertices $a, a \in \{x, y, z\}$ and label the neighbours of a in G_1 (G_2) by a_1 and a_2 (a'_1 and a'_2). The inverse t_2 -switching replaces the six triangles under consideration by $xyz, x_iy_iz_i \in G_1$ for $i = 1, 2$ and $xyz, x'_iy'_iz'_i \in G_2$ for $i = 1, 2$. This switching is applicable only if all these fifteen vertices a, a_i, a'_i for $a \in \{x, y, z\}$ and $i = 1, 2$ are distinct.

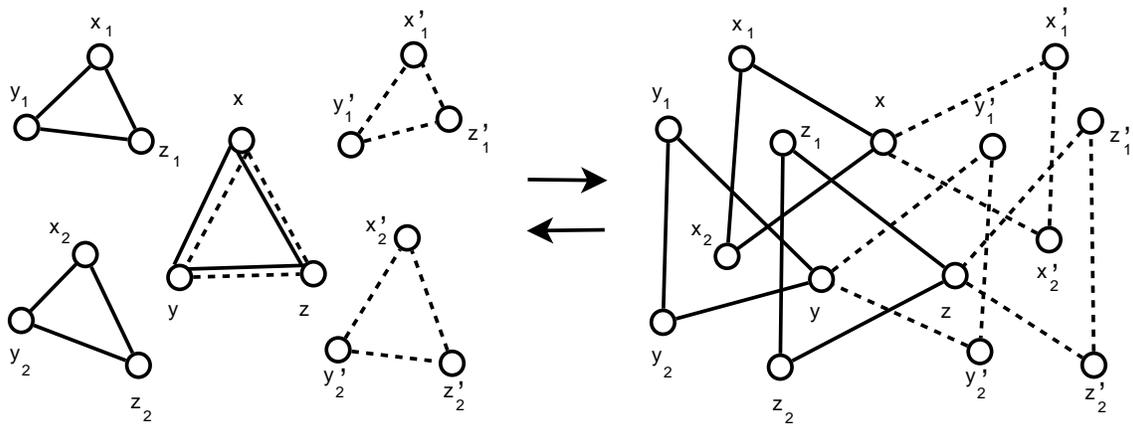


Figure 2: t_2 -switching and its inverse

For any $g \in F_{\ell,3t}$ and $g' \in F_{\ell,3t-3}$, define $N(g)$ and $N'(g')$ the same way as in the proof of Lemma 13. Following an analogous argument of Lemma 13, it is not hard to show that

$$\begin{aligned} 6t \cdot 6^2 \binom{n/3 - (\ell + t) - 6}{2}^2 &\leq N(g) \leq 6t \cdot 6^2 \binom{n/3 - (\ell + t)}{2}^2 \\ (4(n - 4(\ell - 1) - 3t - 10))^3 &\leq N'(g') \leq (4(n - 4(\ell - 1) - 3t))^3. \end{aligned}$$

Thus,

$$\frac{32(n - 4\ell - 3t - 6)^3}{3(n - 3\ell - 3t)^4} \leq \frac{f_{\ell,3t}}{f_{\ell,3(t-1)}} \leq \frac{32(n - 4\ell - 3t + 4)^3}{3(n - 3\ell - 3t - 21)^4}. \blacksquare$$

Corollary 15 For all $j = o(n)$,

$$\frac{f_{j-3k-3,3k+3}}{f_{j-3k,3k}} \sim \frac{4[j - 3k - 1]_3}{3n}.$$

Proof. This follows by Lemmas 13 and 14 and

$$\frac{f_{j-3k-3,3k+3}}{f_{j-3k,3k}} = \frac{f_{j-3k-3,3k+3}}{f_{j-3k-3,3k}} \prod_{i=0}^2 \frac{f_{j-3k-i-1,3k}}{f_{j-3k-i,3k}}. \blacksquare$$

Lemma 16 Assume $\liminf_{n \rightarrow \infty} p(n) > 0$. Let $\gamma(n) = n / \ln \ln n$. Then

$$\sum_{j \geq \gamma(n)} f_j = o(|\mathcal{S}_6| \mu_n).$$

Proof. Let $G \in \mathcal{S}_6$ and let $\kappa_j(G)$ be the number of graphs in \mathcal{S}_6 which shares at least j edges with G . We estimate an upper bound of $\kappa_j(G)$. Let $j = \ell + 3t$ and we consider the number of graphs G' in \mathcal{S}_6 that shares at least ℓ and $3t$ edges of type 1 and 2 respectively with G . Then there are $\binom{n/3}{t}$ ways to choose the t triangles contained both in G and G' . Then there are $\binom{n/3-t}{\ell} 3^\ell$ ways to choose the ℓ triangles in G that contain the ℓ edges of type 1 and to locate these ℓ edges. Given these ℓ edges in G' , there are at most $[n - 3t - 2\ell]_\ell$ ways to choose another ℓ vertices to form the ℓ triangles in G' . Then there are at most

$$\frac{(n - 3t - 3\ell)!}{6^{n/3-t-\ell} (n/3 - t - \ell)!} \leq 9^n n^{2(n/3-t-\ell)}$$

ways to partition the remaining $n - 3t - 3\ell$ vertices into vertex disjoint triangles in G' . Hence

$$\kappa_j(G) \leq \sum_{\ell} \binom{n/3}{t} \binom{n/3-t}{\ell} 3^\ell [n - 3t - 2\ell]_\ell 9^n n^{2(n/3-t-\ell)} \leq n \cdot \max_{\ell} \{n^t n^{2\ell} \ell^{-\ell} 9^n n^{2(n/3-t-\ell)}\},$$

where $t = (j - \ell)/3$. Thus,

$$\ln(\kappa_j(G)) \leq \max_{\ell} \{(2n/3 - t) \ln n - \ell \ln(\ell)\} + O(n).$$

We consider only $j \geq \gamma(n)$. So the maximum is achieved at $\ell = n^{1/3}$. Thus

$$\ln(\kappa_j(G)) \leq \frac{2n}{3} \ln n - \frac{j}{3} \ln n + O(n),$$

We also have

$$\ln \mu_n = n \ln p + \frac{2n}{3} \ln n + O(n).$$

So

$$\ln(\kappa_j(G)) - \ln \mu_n \leq -\frac{j}{3} \ln n - n \ln p + O(n) \rightarrow -\infty,$$

as $n \rightarrow \infty$ since $\liminf_{n \rightarrow \infty} p(n) > 0$, which completes the proof of the lemma. \blacksquare

Proof of Theorem 11. For any $j \geq 0$,

$$r_j = \frac{\sum_{k=0}^{\lfloor j/3 \rfloor} f_{j-3k,3k}}{\sum_{k=0}^{\lfloor (j-1)/3 \rfloor} f_{j-1-3k,3k}}. \quad (6.2)$$

By Corollary 15, for all $j = o(n^{1/3})$, $r_j \sim f_{j,0}/f_{j-1,0}$. By Lemma 13, this ratio is asymptotic to $2/j$. This verifies Theorem 1 (a). Let $\gamma(n) = n/\ln \ln n$. Lemma 16 verifies condition (c). The proof is completed by verifying condition (b). Since $r_j \sim 2/j$ for all $j = o(n^{1/3})$, we only need to show that for all $n^{1/3}/\ln n \leq j \leq \gamma(n)$, $r_j \leq m/2N$. It follows directly from the following two facts.

(a) Let $\widehat{k} = \min\{k : j - 3k \leq \ln n\}$. By Corollary 15,

$$\sum_{k=0}^{\lfloor j/3 \rfloor} f_{j-3k,3k} \sim \sum_{k=0}^{\widehat{k}} f_{j-3k,3k}, \quad \sum_{k=0}^{\lfloor (j-1)/3 \rfloor} f_{j-1-3k,3k} \sim \sum_{k=0}^{\widehat{k}} f_{j-1-3k,3k}.$$

(b) By Lemma 13, for all $0 \leq k \leq \widehat{k}$, $f_{j-3k,3k}/f_{j-1-3k,3k} = o(1)$. \blacksquare

7 Proofs of Theorems 1 and 3

Before approaching Theorems 1 and 3, we first prove a technical lemma.

Lemma 17 *Let $N = \binom{n}{2}$ and let $p = m(n)/N$, where $0 < m(n) < N$. Then for any integer $\ell = \ell(n) \geq 0$ such that $\limsup_{n \rightarrow \infty} \ell(n)/m(n) < 1$,*

$$\binom{N-\ell}{m-\ell} / \binom{N}{m} = p^\ell \exp\left(-\frac{1-p}{pN} \frac{\ell^2 - \ell}{2} + O(\ell^3/m^2)\right).$$

Moreover, if $\ell = \Omega(\sqrt{m})$, then

$$\binom{N-\ell}{m-\ell} / \binom{N}{m} = p^\ell \exp\left(-\frac{1-p}{pN} \frac{\ell^2}{2} + O(\ell^3/m^2)\right).$$

Proof.

$$\begin{aligned}
\binom{N-\ell}{m-\ell} / \binom{N}{m} &= \frac{[m]_\ell}{[N]_\ell} = \prod_{i=0}^{\ell-1} \frac{m-i}{N-i} \\
&= \prod_{i=0}^{\ell-1} \frac{m}{N} \exp\left(-\frac{i}{m} + \frac{i}{N} + O(i^2/m^2)\right) \quad (\text{since } \limsup_{n \rightarrow \infty} \ell(n)/m(n) < 1) \\
&= p^\ell \exp\left(-\frac{1-p}{pN} \frac{\ell^2 - \ell}{2} + O(\ell^3/m^2)\right).
\end{aligned}$$

If we have further that $\ell = \Omega(\sqrt{m})$, then $\ell/pN = O(\ell^3/m^2)$. ■

Proof of Theorem 1. In this proof, the probability space refers to the random graph $\mathcal{G}(n, m)$ only. Let $s = |\mathcal{S}|$. By (2.1) and (2.2),

$$\mathbf{E}X_n = s(m/N)^h \exp\left(-\frac{N-m}{mN} \frac{h^2}{2} + O(h^3/m^2)\right).$$

We also have

$$\mathbf{E}X_n^2 = \sum_{j=0}^h f_j \binom{N-(2h-j)}{m-(2h-j)} / \binom{N}{m}.$$

Let $g(j) = f_j \binom{N-(2h-j)}{m-(2h-j)} / \binom{N}{m}$. By condition (a), for every $K > 0$ and any $1 \leq j \leq Kh^2/m$,

$$\frac{g(j)}{g(j-1)} = r_j \cdot \frac{N}{m} (1 + O(h/m)) = \frac{h^2}{mj} (1 + O(h/m) + o(m/h^2)) = \frac{h^2}{mj} (1 + o(m/h^2)), \quad (7.1)$$

where the last equality holds because $h^3 = o(m^2)$. By condition (c) and the fact that for any integer $0 \leq j \leq h$, $\binom{N-(2h-j)}{m-(2h-j)} \leq \binom{N-h}{m-h}$, we also have that

$$\sum_{j > \gamma(n)} g(j) \leq t(n) \binom{N-h}{m-h} / \binom{N}{m} = t(n) \mu_n / s.$$

Then for all sufficiently large $K > 0$,

$$\begin{aligned}
\mathbf{E}X_n^2 &= \sum_{j=0}^h g(j) = \sum_{j=0}^{Kh^2/m} g(j) + O(g(Kh^2/m)) + O(t(n) \mu_n / s) \\
&= (1 + O(K^{-1})) \sum_{j=0}^{Kh^2/m} g(j) + O(t(n) \mu_n / s), \quad (7.2)
\end{aligned}$$

where the second equality holds because of condition (b) and the last equality holds by (7.1). Next, we estimate $\sum_{j=0}^{Kh^2/m} g(j)$. By (7.1) and Lemma 17,

$$\begin{aligned} \sum_{j=0}^{Kh^2/m} g(j) &= f_0 \frac{\binom{N-2h}{m}}{\binom{N}{m}} \sum_{j=0}^{Kh^2/m} \frac{(h^2/m)^j}{j!} (1 + o(jm/h^2)) \\ &= f_0 \cdot (m/N)^{2h} \exp\left(-\frac{N-m}{mN} \frac{(2h)^2}{2} + O(h^3/m^2)\right) \left(\exp(h^2/m + o(1)) + \Gamma(K)\right), \\ &= f_0 \cdot (m/N)^{2h} \exp\left(-\frac{N-m}{mN} 2h^2\right) \exp(h^2/m) \left(1 + o(1) + O(\Gamma(K) \exp(-h^2/m))\right), \end{aligned} \quad (7.3)$$

where

$$\Gamma(K) = O\left(\frac{(h^2/m)^{Kh^2/m}}{(Kh^2/m)!}\right) = O\left(\left(\frac{(eh^2/m)}{(Kh^2/m)}\right)^{Kh^2/m}\right),$$

which goes to 0 as $K \rightarrow \infty$, since $h^2/m = \Omega(1)$. By (7.2) and (7.3), for every $\epsilon > 0$, there is a sufficiently large K , such that

$$\mathbf{E}X_n^2 = (1 + O(\epsilon)) f_0 \cdot (m/N)^{2h} \exp\left(-\frac{N-m}{mN} 2h^2\right) \exp(h^2/m) + O(t(n)\mu_n/s). \quad (7.4)$$

We also have

$$s^2 = \sum_{j=0}^h f_j = f_0 \sum_{j=0}^h \prod_{i=1}^j r_i.$$

With the same reasoning as before, it is enough to sum over the first Kh^2/N terms, leaving an arbitrarily small tail plus an error term $O(t(n))$. This yields

$$s^2 = (1 + O(\epsilon)) f_0 \exp(h^2/N) + O(t(n)).$$

Since $t(n) = o(\mu_n s) = o(s^2)$ by condition (c), we obtain

$$f_0 = (1 + O(\epsilon)) s^2 \exp(-h^2/N).$$

Combining with (7.4) and again by condition (c), we obtain

$$\begin{aligned} \mathbf{E}X_n^2 &= (1 + O(\epsilon)) s^2 (m/N)^{2h} \exp\left(-\frac{N-m}{mN} 2h^2\right) \exp(h^2/m - h^2/N) + O(t(n)\mu_n/s) \\ &= (1 + O(\epsilon)) s^2 (m/N)^{2h} \exp\left(-\frac{N-m}{mN} h^2\right) + o(\mu_n^2) = (1 + O(\epsilon)) (\mathbf{E}X_n)^2. \end{aligned}$$

As this holds for every $\epsilon > 0$, we have $\mathbf{E}X_n^2 = (1 + o(1)) (\mathbf{E}X_n)^2$. Then for every $\epsilon > 0$,

$$\mathbf{P}(|X_n/\mathbf{E}X_n - 1| > \epsilon) \rightarrow 0, \quad \text{as } n \rightarrow \infty,$$

by Chebyshev's inequality. \blacksquare

Proof of Theorem 3. Let Y_n denote the number of edges in $\mathcal{G}(n, p)$, then $Y_n \sim \text{Bin}(N, p)$. Hence we have

$$Y_n - pN = O_p(\sqrt{p(1-p)N}), \quad (7.5)$$

where $f(n) = O_p(g(n))$ for some $g(n) \geq 0$ means $\mathbf{P}(|f(n)| > Kg(n)) \rightarrow 0$ as $K \rightarrow \infty$ and $n \rightarrow \infty$. Similarly we use the notation $f(n) = o_p(g(n))$ meaning that for every $\epsilon > 0$, $\mathbf{P}(|f(n)| > \epsilon g(n)) \rightarrow 0$ as $n \rightarrow \infty$. Since $X_n/\mathbf{E}_{\mathcal{G}(n,m)X_n} \xrightarrow{p} 1$ in $\mathcal{G}(n, m)$ for all $m = pN + O(\sqrt{p(1-p)N})$ by assumption and $\ln(\mathbf{E}_{\mathcal{G}(n,m)X_n}) = \ln|\mathcal{S}| + h \ln(m/N) + (N-m)h^2/2mN + o(1)$ by (2.2), by conditioning on Y_n , we have

$$\ln X_n - \ln|\mathcal{S}| - h \ln(Y_n/N) + \frac{1 - Y_n/N}{Y_n} \frac{h^2}{2} \xrightarrow{p} 0. \quad (7.6)$$

By (7.5),

$$\frac{1 - Y_n/N}{Y_n} \frac{h^2}{2} = \frac{h^2(1-p)}{2Np} \left(1 + O_p \left(\sqrt{\frac{p}{(1-p)N}} + \sqrt{\frac{1-p}{pN}} \right) \right) = \frac{h^2(1-p)}{2Np} + o_p(1), \quad (7.7)$$

where the equality above holds because $h^3 = o(p^2 n^4)$. We also have

$$\ln(Y_n/N) = \ln p(1 + Y_n^* \sqrt{(1-p)/pN}) = \ln p + \sqrt{(1-p)/pN} Y_n^* + O_p((1-p)/pN), \quad (7.8)$$

where

$$Y_n^* = \frac{Y_n - pN}{\sqrt{p(1-p)N}}$$

is the normalised variable of Y_n . Recall that $\lambda_n = |\mathcal{S}|p^h$ from (2.1) and $\mathbf{E}X_n = \lambda_n$. Combining with (7.6)–(7.8), we have

$$\ln(X_n/\lambda_n) + \frac{\beta_n^2}{2} = \beta_n Y_n^* + o_p(1). \quad (7.9)$$

Since $\beta_n = \Omega(1)$, (7.9) immediately yields

$$\frac{\ln(e^{\beta_n^2/2} X_n/\lambda_n)}{\beta_n} = Y_n^* + o_p(1).$$

Since $Y_n^* \xrightarrow{d} \mathcal{N}(0, 1)$, the theorem follows. ■

8 Concluding remarks

It was proved in [4] that $m \gg n^{3/2}$ is required for the concentration of X_n in $\mathcal{G}(n, m)$, where X_n denotes the number of Hamilton cycles or perfect matchings or spanning trees, as the variable will become asymptotically log-normally distributed when $m = \Theta(n^{3/2})$. However, we do not think this condition is sufficient in the case of triangle-factors. It is surprising that the critical point of m when X_n changes from small deviation ($\mathbf{E}X_n^2 \sim (\mathbf{E}X_n)^2$) to large deviation ($\limsup_{n \rightarrow \infty} \mathbf{E}X_n^2/(\mathbf{E}X_n)^2 > 1$) in $\mathcal{G}(n, m)$ seems to be different for Hamilton cycles and for triangle-factors. We guess $m = n^{5/3}$ might be the critical point for the latter case.

As explained in Section 4, the most interesting set \mathcal{S} to be studied is perhaps the one containing graphs isomorphic to an unlabelled graph H_n on n vertices. Unfortunately, it is not easy to define the sequence $(H_n)_{n \geq 1}$ in general and for a general H_n , computing r_j might be hard. It will be interesting to discover more classes of such graph sequences (H_n) and see whether the corresponding random variables X_n follow the log-normal paradigm.

References

- [1] A. Frieze and S. Suen, Counting the number of Hamilton cycles in random digraphs, *Random Structures Algorithms* 3 (1992), no. 3, 235–241.
- [2] P. Gao, Distribution of the number of spanning regular subgraphs in random graphs. *Random Structures Algorithms*, (2012), in press, DOI: 10.1002/rsa.20418.
- [3] S. Janson, Random regular graphs: asymptotic distributions and contiguity, *Combin. Probab. Comput.* 4 (1995), no. 4, 369–405.
- [4] S. Janson, The numbers of spanning trees, Hamilton cycles and perfect matchings in a random graph, *Combin. Probab. Comput.* 3 (1994), 97–126.
- [5] B.D. McKay, Asymptotics for symmetric 0-1 matrices with prescribed row sums, *Ars Combinatoria* 19A (1985), 15–25.
- [6] A. Ruciński, Subgraphs of random graphs: a general approach, *Random graphs '83* (Poznan, 1983), pp. 221–229, North-Holland Math. Stud. 118, North-Holland, Amsterdam, 1985.
- [7] A. Ruciński, When are small subgraphs of a random graph normally distributed? *Probab. Theory Related Fields*, 78 (1988), no. 1, 1–10.
- [8] R. W. Robinson and N.C. Wormald, Almost all cubic graphs are Hamiltonian, *Random Structures Algorithms* 3 (1992), no. 2, 117–125.
- [9] R.W. Robinson and N. C. Wormald, Almost all regular graphs are Hamiltonian, *Random Structures Algorithms* 5 (1994), no. 2, 363–374.
- [10] E. M. Wright, For how many edges is a graph almost certainly Hamiltonian? *J. London Math. Soc.* (2) 8 (1974), 44–48.
- [11] E. M. Wright, For how many edges is a digraph almost certainly Hamiltonian? *Proc. Amer. Math. Soc.* 41 (1973), 384–388.