3-star factors in random d-regular graphs

Hilda Assiyatun*
Department of Mathematics
Institut Teknologi Bandung
Bandung, Indonesia
hilda@math.itb.ac.id

Nicholas Wormald†
Department of Combinatorics and Optimization
University of Waterloo
Waterloo ON, Canada
nwormald@uwaterloo.ca

Abstract

The small subgraph conditioning method first appeared when Robinson and the second author showed the almost sure hamiltonicity of random d-regular graphs. Since then it has been used to study the almost sure existence of, and the asymptotic distribution of, regular spanning subgraphs of various types in random d-regular graphs and hypergraphs. In this paper, we use the method to prove the almost sure existence of 3-star factors in random d-regular graphs. This is essentially the first application of the method to non-regular subgraphs in such graphs.

1 Introduction

Was shown by Robinson and the second author that for fixed d a random d-regular graph contains a Hamilton cycle with probability tending to 1 as the number n of vertices tends to infinity (provided the necessary conditions of $d \geq 3$ and dn even are satisfied). A key ingredient in the proof in [12] and [13] was understanding the distribution of the number of perfect matchings in such graphs when n is even. The method used is called the *small subgraph conditioning method* in [14]. It has been successfully used to determine the existence with high probability, and the asymptotic distribution, of the number of k-regular spanning subgraphs in random d-regular graphs, for $k = 1, 2$ (see [7] and [10]). It has also been applied to the number of long cycles in random d-regular graphs by Garmo [5], which seems to be the only application to non-spanning subgraphs. The method is given in an accessible form in the main theorems of Molloy et al. [9] and Janson [7]. In the latter, the method and its proof are slightly streamlined and the implications for the asymptotic distribution of the number of subgraphs in question are given explicitly. The aim of the present paper is to investigate some other fundamental but non-regular spanning subgraphs.

A *star* is a tree with at most one vertex whose degree is greater than 1, and a *k-star* is a star with k leaves. A *k-star factor* in a graph is a spanning subgraph whose components are k-stars. In

*This research formed part of this author's PhD Thesis at the University of Melbourne, 2002

†The research was carried out while both authors were in the Department of Mathematics and Statistics, University of Melbourne.

these terms, we can rephrase the perfect matching result mentioned above as the statement that d-regular graphs with high probability have a 1-star factor (for n even).

We use the notation $\mathbf{P}$ (probability), $\mathbf{E}$ (expectation) and $\mathbf{Var}$ (variance). We say an event Y_n occurs a.a.s. (asymptotically almost surely) if $\mathbf{P}Y_n \to 1$ as n goes to infinity. The asymptotically almost sure hamiltonicity of random d-regular graphs ($d \geq 3$) also implies the existence a.a.s. of a 2-star factor in such graphs when $n \equiv 0 \bmod 3$. One can easily see this by going along the Hamilton cycle, colouring two successive edges by red and the next edge by blue and repeating this colouring until all edges are coloured. The red edges constitute a 2-star factor. In [1] Assiyatun determined the asymptotic distribution of the number of 2-star factors in random d-regular graphs, for $d \geq 3$. A natural question which arises from these results is the existence a.a.s. of a k-star factor for $k \geq 3$. As a random 3-regular graph a.a.s. has no 3-star factors (see Corollary 2), in this paper we will study the existence a.a.s. of 3-star factors in random d-regular graphs, for fixed $d \geq 4$. It would be of interest to know about k-star factors for $k > 3$, but this would appear to require considerably more complicated analysis.

The most difficult part of the small subgraph conditioning method is invariably the computation of $\mathbf{E}Y^2$, where Y denotes the number of subgraphs whose existence or distribution is being studied in the random regular graphs. However it turns out that this is too complicated for general d. So we start by proving the existence a.a.s. of a 3-star factor in random 4-regular graphs (provided the number of vertices satisfies the obvious necessary condition). Using the contiguity of models of random regular graphs ([7], [14]) we then obtain the existence a.a.s. of a 3-star factor in random d-regular graphs, for fixed $d \geq 4$. The main results obtained in this paper are presented in the following theorem and corollary.

Let $\mathcal{G}_{n,d}$ denote the probability space of d-regular graphs on n vertices.

Theorem 1 *Restrict n to $0 \bmod 4$. Then $G \in \mathcal{G}_{n,4}$ a.a.s. has a 3-star factor. Furthermore, letting Y_4 denote the number of 3-star factors in $G \in \mathcal{G}_{n,4}$,*

$$\frac{Y_4}{\mathbf{E}Y_4} \xrightarrow{d} W = \Pi_{k=3}^{\infty} (1+\delta_k)^{Z_k} e^{-\lambda_k \delta_k} \text{ as } n \to \infty,$$

where Z_k are independent Poisson variables with $\mathbf{E}Z_k = \lambda_k$ for $k \geq 3$, $\lambda_k = 3^k/2k$ and $\delta_k = (-1/5 + 2i/5)^k + (-1/5 - 2i/5)^k$.

Note that $\delta_k = 2\left(\sqrt{5}/5\right)^k \cos(k\theta) > -1$ for all $k \geq 1$, where $\theta = \arctan 2$.

Corollary 1 *Restricting to $n \equiv 0 \bmod 4$ and for fixed $d \geq 4$, $G \in \mathcal{G}_{n,d}$ a.a.s. has a 3-star factor.*

The proof of Theorem 1 has a further consequence. Form a random 4-regular graph by starting with a 3-star factor, and then choosing a random completion to a 4-regular graph. Then the resulting random graph is contiguous to $\mathcal{G}_{n,4}$. This means that the two random graphs have the same asymptotically almost sure properties. (See proof of Corollary 1 for a precise definition; the proof of this claim uses [14, Theorem 4.1] together with the calculations in Sections 3 and 4.)

Instead of working directly with $\mathcal{G}_{n,d}$, we will use the *pairing model* which was first given by Bollobás (see [3]) and implicitly by Bender and Canfield [2]. This model can be described as follows. Let $V = \bigcup_{i=1}^{n} V_i$ be a fixed set of dn points, where $|V_i| = d$, for every i. A perfect matching of points of V into $dn/2$ pairs is called a *pairing*. A pairing P corresponds to a d-regular pseudograph $G(P)$ in which each V_i is regarded as a vertex and each pair is an edge. We use $\mathcal{P}_{n,d}$ to denote the probability space of all pairings. As shown in [3], for d fixed, the probability that the pseudograph has no loops or multiple edges (i.e. is simple) is asymptotically bounded below by a positive constant. Moreover, each simple graph arises with the same probability as $G(P)$ for $P \in \mathcal{P}_{n,d}$. Hence the following is true.

Lemma 1 *A property of graphs that is holds a.a.s. for the random pseudographs arising from $\mathcal{P}_{n,d}$ will also hold a.a.s. in $\mathcal{G}_{n,d}$.*

This paper consists of five sections. In Section 2 we analyse the expectation of the number of d-star factors and of the number of 3-star factors in the random d-regular pseudographs coming from $\mathcal{P}_{n,d}$. In Section 3 we deal with the variance of the number of 3-star-factors in these pseudographs, while in Section 4 we evaluate its expectation conditioned on the short cycle distribution. The proofs of the main results are presented in the last section.

For two sequences a_n and b_n, we denote $a_n \sim b_n$ if the ratio $\frac{a_n}{b_n}$ tends to 1 as n goes to infinity. We denote the falling factorial $n(n-1)\cdots(n-m+1)$ by $[n]_m$.

2 The expectation of the number of star factors

Throughout this paper we define

$$N(2m) = \frac{(2m)!}{m!2^m},$$

which is the number of perfect matchings of $2m$ points.

As mentioned in the introduction we now show that a d-regular random graph a.a.s. has no d-star factors. For this we need the following lemma. Note that counting subgraphs of the pseudograph coming from $\mathcal{P}_{n,d}$ is equivalent to counting the corresponding sets of pairs in the pairing, and for such counting purposes, parallel edges are distinguishable from each other (especially as they come from distinct pairs in the pairing).

Lemma 2 *For fixed $d \geq 3$, let $n \equiv 0 \bmod (d+1)$ and let S be the number of d-star factors in the random d-regular pseudograph coming from $\mathcal{P}_{n,d}$. Then*

$$\mathbf{E}S \sim \sqrt{d+1}\left(\left(\frac{d-1}{d+1}\right)^{\frac{d(d-1)}{2(d+1)}}(d+1)^{\frac{1}{d+1}}\right)^n.$$

Proof. Let G be a d-star factor on n vertices. The number of automorphisms on G is

$$\left(\frac{n}{d+1}\right)!(d!)^{n/(d+1)}.$$

As we work in the pairing model, for a given d-star in the pseudograph, the points can be chosen in $d!d^d$ ways. Therefore, the number of ways to choose a set of pairs in the pairing corresponding to a d-star factor is

$$\frac{n!}{\left(\frac{n}{d+1}\right)!(d!)^{n/(d+1)}}\left(d!d^d\right)^{n/(d+1)} = \frac{n!}{\left(\frac{n}{d+1}\right)!}d^{dn/(d+1)}. \tag{1}$$

Since every centre of d-stars has degree d and the other vertices are of degree 1, the number of ways to complete a d-star factor to a pairing in $\mathcal{P}_{n,d}$ is

$$N\left(\frac{d(d-1)n}{d+1}\right). \tag{2}$$

Multiplying (1) and (2) together, and then dividing by the total number of pairings $N(dn)$, we have

$$\mathbf{E}S = \frac{n!\left(\frac{d(d-1)n}{d+1}\right)!(dn/2)!d^{dn/(d+1)}2^{dn/2}}{\left(\frac{n}{d+1}\right)!\left(\frac{d(d-1)n}{2(d+1)}\right)!(dn)!2^{\frac{d(d-1)n}{2(d+1)}}}.$$

Using Stirling's formula we now obtain the expected value of S as claimed in the lemma. ∎

Corollary 2 *For fixed $d \geq 3$, a random graph $G \in G_{n,d}$ a.a.s. does not have a d-star factor.*

Proof. In the expectation of S in Lemma 2, the expression raised to the power n has negative derivative for $d \geq 3$ and has value $2^{-1/4}$ at $d = 3$. So it is always less then $2^{-1/4} < 1$. Since $\mathbf{P}(S > 0) \leq \mathbf{E}S$, we have $\mathbf{P}(S > 0) \to 0$. Thus, for $d \geq 3$, $P \in P_{n,d}$ a.a.s. does not have a d-star factor. This is consequently a.a.s. true conditioned on no loops or multiple edges, by Lemma 1. The claim for $G \in G_{n,d}$ follows immediately. ∎

Let $n \equiv 0 \bmod 4$ and define Y_d' to be the number of 3-star factors in a random d-regular pseudograph coming from $\mathcal{P}_{n,d}$, for fixed $d \geq 4$.

Theorem 2

$$\mathbf{E}Y_d' \sim 2\left(d(d-3/2)^{d/2-3/4}\left(\frac{2}{d^d}\right)^{1/2}\left(\frac{(d-1)(d-2)}{3!}\right)^{1/4}\right)^n.$$

Proof. Following the proof of Lemma 2, we have that

$$\begin{aligned}\mathbf{E}Y_d' &= \frac{n!}{(n/4)!}\left(d\left(\frac{(d-1)(d-2)}{3!}\right)^{1/4}\right)^n \times \frac{N(n(d-3/2))}{N(nd)} \\ &= \frac{n!(n(d-3/2))!(nd/2)!}{(n/4)!(n(d/2-3/4))!(nd)!}\left(2^{3/4}d\left(\frac{(d-1)(d-2)}{3!}\right)^{1/4}\right)^n.\end{aligned} \tag{3}$$

Applying Stirling's formula, we obtain the expectation of Y_d' as claimed. ∎

It is easy to verify that the quantity raised to the power n in the expression for $\mathbf{E}Y_d'$ in Theorem 2 is strictly greater than 1 for $d \geq 4$, though we do not rely on this except for the case $d = 4$.

3 The variance of Y_4'

In the rest of this paper let $n \equiv 0 \bmod 4$ and recall that Y_4' is the number of 3-star factors in a random 4-regular pseudograph coming from $\mathcal{P}_{n,4}$.

Following the method used in [11] (see also [4] and [10] for similar arguments) we obtain the following theorem.

Theorem 3

$$\mathbf{E}Y_4' \sim 2\left(\frac{5^{5/4}}{2^{11/4}}\right)^n \quad \textit{and} \quad \mathbf{Var}Y_4' \sim (\mathbf{E}Y_4')^2\left(\frac{25}{4\sqrt{13}}-1\right).$$

Proof. The first part of the theorem is obtained from Theorem 2 for $d = 4$. To show the second part of the theorem we count the ways to lay down an ordered pair of 3-star factors in $P \in \mathcal{P}_{n,4}$. In general, a set of pairs in P inducing a subgraph of a given type (e.g. a star or a star factor) will be called by the same name in the pairing. Let S_i be a 3-star factor of P, for $i = 1, 2$. Let $T = S_1 \cap S_2$ and suppose T consists of x' 1-stars, y' 2-stars and z' 3-stars (see Figure 1). We refer to the set of i-stars in T as T_i, for $i = 1, 2, 3$.

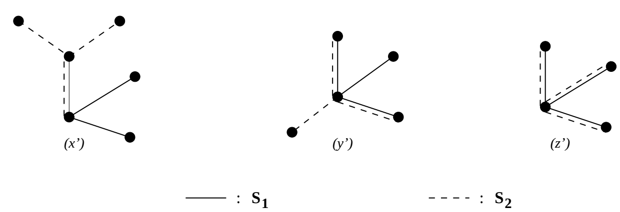

Figure 1: Two intersecting 3-star factors

Given S_1, the number of possibilities of the intersection T is given by

$$\binom{n/4}{x'}3^{x'}\binom{n/4-x'}{y'}3^{y'}\binom{n/4-x'-y'}{z'} = \frac{(n/4)!3^{x'+y'}}{(n/4-x'-y'-z')!x'!y'!z'!}. \tag{4}$$

We observe here that there are $n/4 - x' - y' - z'$ 3-stars in both 3-star factors that do not share any edge. We call these edge-disjoint 3-stars *isolated* 3-stars.

Now we have to complete S_2 by creating the isolated 3-stars and completing T_2 and T_1 into 3-stars. The centres of the isolated 3-stars in S_2 cannot be chosen from the vertex set of T nor the centres of the isolated 3-stars in S_1. There are $2x'+3y'+4z'+n/4-x'-y'-z' = n/4+x'+2y'+3z'$ such vertices. Thus, the number of ways to choose these centres is

$$\binom{3n/4-x'-2y'-3z'}{n/4-x'-y'-z'}. \tag{5}$$

The number of ways to choose the leaves of the isolated 3-stars in S_2 is

$$\prod_{k=0}^{n/4-x'-y'-z'-1}\binom{3n/4-x'-2y'-3z'-3k}{3}$$

$$= \frac{(3n/4 - x' - 2y' - 3z')!}{(2x' + y')!(3!)^{n/4 - x' - y' - z'}}. \tag{6}$$

Note here that there are $2x' + y'$ vertices left to assign for the completion of T_2. Then the number of ways to choose y' leaves to complete y' 2-stars in T_2 is

$$\frac{(2x' + y')!}{(2x')!}. \tag{7}$$

The number of ways to choose the leaves to complete T_1 is

$$\prod_{k=0}^{x'-1} \binom{2x' - 2k}{2} = \frac{(2x')!}{2^{x'}}. \tag{8}$$

So far we have determined the graph corresponding to S_2 but not chosen the pairs of points corresponding to its edges. The number of choices for these is

$$(3!)^{x'}(3!)^{n/4 - x' - y' - z'} 3^{n/2 - y' - 2z'}. \tag{9}$$

Having S_1 and S_2, we observe that in T_3 there are $3z'$ vertices of degree 1 and z' vertices of degree 3. In T_2 there are $2y'$ vertices of degree 1 and y' vertices of degree 4. In T_1 there are $2x'$ vertices of degree 3. Observing the isolated 3-stars in S_1 and S_2, we note that there are $n/2 - 2x' - 2y' - 2z'$ vertices of degree 4. The remaining vertices (there are $n/2 - y' - 2z'$ of these) are of degree 2. Thus, the number of free points in V is

$$9z' + z' + 6y' + 2x' + n - 2y' - 4z' = n + 2x' + 4y' + 6z'.$$

Therefore, the number of ways to complete the pairing P is

$$N(n + 2x' + 4y' + 6z'). \tag{10}$$

Hence, multiplying equations (4–10) by the number of ways to choose S_1 as in (3) with $d = 4$ and then dividing by $N(4n)$, we have

$$\begin{aligned} \mathbf{E}Y_4'(Y_4' - 1) \;\; = \;\; & \frac{n!(2n)!2^{3n/2}3^{n/2}4^n}{(4n)!} \\ & \sum_{R'} \frac{(3n/4 - x' - 2y' - 3z')!^2 (n + 2x' + 4y' + 6z')! 3^{2x' - 2z'} 2^{-x' - 2y' - 3z'}}{(n/4 - x' - y' - z')!^2 (n/2 - y' - 2z')!(n/2 + x' + 2y' + 3z')! x'! y'! z'!} \end{aligned}$$

where $R' = \{(x', y', z') : x', y', z' \geq 0,\ x' + y' + z' \leq n/4\}$.

Set $x = \frac{x'}{n}, y = \frac{y'}{n}$, and $z = \frac{z'}{n}$. We will now assume that all arguments in the factorial above go to infinity with n. This is justified by noting that Stirling's formula is correct to within a constant factor even if the argument does not go to infinity, and so the formulae in the following argument similarly have such accuracy in all cases. It will then be clear that the cases where not all arguments go to infinity are negligible. So Stirling's formula gives

$$\begin{aligned}\mathbf{E}Y_4'(Y_4'-1) &\sim \sqrt{\pi n}\left(\frac{3^{1/2}}{2^{5/2}}\right)^n\left(\frac{n}{e}\right)^{-n} \\ &\quad\times\frac{1}{4\pi^2n^2}\left(\frac{n}{e}\right)^n\sum_R \alpha(x,y,z)\left(F(x,y,z)\right)^n \\ &\sim \frac{1}{4(\pi n)^{3/2}}\left(\frac{3^{1/2}}{2^{5/2}}\right)^n\sum_R \alpha(x,y,z)\left(F(x,y,z)\right)^n\end{aligned} \tag{11}$$

where

$$R=\{(x,y,z): x,y,z\geq 0,\ x+y+z\leq 1/4\},$$

$$F(x,y,z)=\frac{f(3/4-x-2y-3z)^2f(1+2x+4y+6z)3^{2x-2z}2^{-x-2y-3z}}{f(1/4-x-y-z)^2f(1/2-y-2z)f(1/2+x+2y+3z)f(x)f(y)f(z)}$$

with $f(x)=x^x$ and

$$\alpha(x,y,z)=\left(\frac{2(3/4-x-2y-3z)^2}{(1/4-x-y-z)^2(1/2-y-2z)xyz}\right)^{1/2}.$$

Since by convention $f(0)=1$, it can be seen that F is continuous in R. We now wish to find the main contribution of the sum, which comes from the maximum of F in R. The following three lemmas prove that the maximum of F is attained at $\mathbf{x}_{\max}=(9/64,\ 3/64,\ 1/64)$, with $F(\mathbf{x}_{\max})=\frac{5^{5/2}}{2^3 3^{1/2}}$.

Lemma 3 *Let F and R be as in (11). Then $\mathbf{x}_{\max}=(9/64,\ 3/64\ 1/64)$ with $F(\mathbf{x}_{\max})=\frac{5^{5/2}}{2^3 3^{1/2}}$ is the only local maximum point of F in the interior of R.*

Proof. First we look for all critical points of F in the interior of R. We set the partial derivatives of $\log F$, with respect to x,y and z, equal to 0, resulting in three equations:

$$\begin{aligned} 18(1/2+x+2y+3z)(1/4-x-y-z)^2-x(3/4-x-2y-3z)^2 &= 0 \quad (12)\\ 4(1/2+x+2y+3z)^2(1/4-x-y-z)^2(1/2-y-2z)-y(3/4-x-2y-3z)^4 &= 0 \quad (13)\\ 8(1/2+x+2y+3z)^3(1/4-x-y-z)^2(1/2-y-2z)^2-9z(3/4-x-2y-3z)^6 &= 0. \quad (14)\end{aligned}$$

We can simplify the above system by substituting the first equation into the second and the third. This results in two new equations:

$$\begin{aligned}(1/2-y-2z)x^2-3^4(1/4-x-y-z)^2y &= 0 \quad (15)\\ (1/2-y-2z)^2x^3-3^8(1/4-x-y-z)^4z &= 0. \quad (16)\end{aligned}$$

By substituting (15) into (16), we obtain

$$xz-y^2=0. \tag{17}$$

$\mathbf{x_i}$	$F(\mathbf{x_i})$	Hessian $F(\mathbf{x_i})$
$\mathbf{x}_{\max} = (0.140625, 0.046875, 0.015625)$	4.034358	negative definite
$\mathbf{x_1} \approx (0.001209, 0.016742, 0.231904)$	3.626396	indefinite

Table 1: The stationary points of F

Eliminate z by substituting $z = y^2/x$ from (17) into (12) and (15) and then take the resultant of the new equations with respect to y (using Maple or similar packages). This results in an equation in x:

$$Cx^{23}(64x-9)(12x-1)^2P(x) = 0$$

where C is a constant and

$$P(x) = 6665732096x^4 + 3603096576x^3 - 187899264x^2 + 2031237x - 2187. \tag{18}$$

It is easy to show that $x = 9/64$ results in $y = 3/64$ only, and $x = 1/12$ does not yield any feasible value for y.

Next we consider the solutions coming from the roots of $P(x)$ in (18). Using Maple we obtain three distinct positive roots of P, namely $x_1 \approx 0.001209$, $x_2 \approx 0.013576$ and $x_3 \approx 0.033933$. However, x_2 and x_3 do not give feasible values for y, while x_1 gives a unique $y_1 \approx 0.016742$.

From (17), each pair (x, y) results in a unique solution z. The nature of these critical points are investigated by determining the Hessian of F on the corresponding points. The result is depicted in Table 1. For convenience, we convert $\mathbf{x}_{\max}$ and its function value to rational. Table 1 shows that $\mathbf{x}_{\max} = (9/64,\ 3/64,\ 1/64)$ is the only local maximum point of F in the interior of R. The assertion follows. ∎

To study the behaviour of F on the boundary of R we generalise the approach used by Garmo in the proof of [6, Lemma 12] (see [5, Appendix A]). First let $\mathbf{x} = (x_1, x_2, \ldots, x_r)$ and $\mathbf{u_i} = (u_{1,i}, u_{2,i}, \ldots, u_{r,i})$ for fixed $r \geq 2$. Naturally, the log function is defined on the set of non-negative real numbers, with, by convention, $0 \cdot \log 0 = 0$.

Lemma 4 *Let R be a closed set in $\mathbb{R}^r$ and let ∂R be the boundary of R. Assume that every point in ∂R is the endpoint of an interval in $R \backslash \partial R$. Let $f_i(\mathbf{x}) = b_i + \mathbf{u_i}\mathbf{x}^T$ for $i = 1, \ldots, m$, where b_i and $\mathbf{u_i}$ are constant, such that $f_i(\mathbf{x}) > 0$ for all i and all $\mathbf{x} \in R \backslash \partial R$. Define F to be a function on R such that*

$$F(\mathbf{x}) = g_0(\mathbf{x}) + \sum_{i=1}^{m} a_i g_i(\mathbf{x}) = g_0(\mathbf{x}) + \sum_{i=1}^{m} a_i f_i(\mathbf{x}) \log f_i(\mathbf{x})$$

with $a_i < 0$ for $i \leq m_0 \leq m$. Suppose that for every $\mathbf{x} \in R$ the directional derivative of g_0 at $\mathbf{x}$ in any direction is bounded. Let $\mathbf{x_0} \in \partial R$ such that $f_i(\mathbf{x_0}) = 0$ for at least one $i \leq m_0$ and $f_i(\mathbf{x_0}) > 0$ for all $m_0 < i \leq m$. Then $\mathbf{x_0}$ is not a local maximum of F on R.

Proof. For a function g, the directional derivative of g at $\mathbf{x_0}$ in the direction of a unit vector $\mathbf{v}$ is

$$\left.\frac{\partial}{\partial t} g(\mathbf{x_0} + t\mathbf{v})\right|_{t=0}$$

(where only the right-hand partial derivative need to be taken, i.e. $t > 0$.)
For $i = 1, \ldots, m$ we have

$$\frac{\partial}{\partial t} g_i(\mathbf{x_0} + t\mathbf{v}) = \mathbf{u_i}\mathbf{v}^T \log\left(b_i + \mathbf{u_i}\left(\mathbf{x_0}^T + t\mathbf{v}^T\right)\right) + \mathbf{u_i}\mathbf{v}^T.$$

We observe that for $\mathbf{x_0} \in R$, if $f_i(\mathbf{x_0}) = 0$ then

$$\frac{\partial}{\partial t} g_i(\mathbf{x_0} + t\mathbf{v}) = \mathbf{u_i}\mathbf{v}^T \log\left(t\mathbf{u_i}\mathbf{v}^T\right) + \mathbf{u_i}\mathbf{v}^T.$$

Now suppose $\mathbf{x_0} \in \partial R$ satisfies $f_i(\mathbf{x_0}) = 0$ for $i \in I_0 \subset \{1, 2, \ldots, m_0\}$ and not any other i. Letting $\mathbf{p}$ be a point on an interval in the interior of R with endpoint $\mathbf{x_0}$, it follows that

$$\mathbf{u_i}\left(\mathbf{p} - \mathbf{x_0}\right)^T = \mathbf{u_i}\mathbf{p}^T - \mathbf{u_i}\mathbf{x_0}^T = \left(b_i + \mathbf{u_i}\mathbf{p}^T\right) - \left(b_i + \mathbf{u_i}\mathbf{x_0}^T\right) = b_i + \mathbf{u_i}\mathbf{p}^T > 0$$

for $i \in I_0$. Hence, choosing $\mathbf{v_0}$ to be the normalisation of $\mathbf{p} - \mathbf{x_0}$, i.e. $\mathbf{v_0} = \frac{\mathbf{p}-\mathbf{x_0}}{||\mathbf{p}-\mathbf{x_0}||}$, we have

$$\lim_{t\to 0} \frac{\partial}{\partial t} g_i(\mathbf{x_0} + t\mathbf{v_0}) = \mathbf{u_i}\mathbf{v_0}^T \lim_{t\to 0} \log\left(t\mathbf{u_i}\mathbf{v_0}^T\right) + \mathbf{u_i}\mathbf{v_0}^T = -\infty$$

for $i \in I_0$, whilst

$$\left.\frac{\partial}{\partial t} g_j(\mathbf{x_0} + t\mathbf{v_0})\right|_{t=0} = \mathbf{u_j}\mathbf{v_0}^T \log\left(b_j + \mathbf{u_j}\mathbf{x_0}^T\right) + \mathbf{u_j}\mathbf{v_0}^T$$

remains bounded for $j \notin I_0$. As $a_i < 0$ for $i \in I_0$ and the contribution from g_0 is bounded, the directional derivative of F at $\mathbf{x_0}$ is positive infinite. Hence $\mathbf{x_0}$ is not a local maximum of F on R.
∎

Now we are ready for the following lemma.

Lemma 5 *Let F and R be as in (11). Then the maximum of F on R does not occur on ∂R.*

Proof. Here we define $\mathbf{x} = (x, y, z)$ and $\mathbf{v} = (v_1, v_1, v_3)$. Following the notation in Lemma 4, we can write $\log F$ as

$$\log F(\mathbf{x}) = g_0(\mathbf{x}) + \sum_{i=1}^{6} a_i f_i(\mathbf{x}) \log f_i(\mathbf{x})$$

where

$$\begin{aligned} g_0(\mathbf{x}) &= (1/2 + x + 2y + 3z)\log(1/2 + x + 2y + 3z) \\ &\quad +(2x - 2z)\log 3 + (1 + x + 2y + 3z)\log 2, \end{aligned}$$

$a_1 = a_2 = a_3 = a_6 = -1,\ a_4 = -2,\ a_5 = 2;\ b_1 = b_2 = b_3 = 0,\ b_4 = 1/4,\ b_5 = 3/4,\ b_6 = 1/2$ and

$$\begin{aligned} \mathbf{u_1} &= (1,\ 0,\ 0) \\ \mathbf{u_2} &= (0,\ 1,\ 0) \\ \mathbf{u_3} &= (0,\ 0,\ 1) \\ \mathbf{u_4} &= (-1,\ -1,\ -1) \\ \mathbf{u_5} &= (-1,\ -2,\ -3) \\ \mathbf{u_6} &= (0,\ -1,\ -2). \end{aligned}$$

For g_0 we have

$$\begin{aligned}\frac{\partial}{\partial t} g_0(\mathbf{x}+t\mathbf{v})\Big|_{t=0} &= (v_1+2v_2+3v_3)\log(1/2+x+2y+3z)\\ &\quad +(v_1+2v_2+3v_3)\log 2+2\,(v_1-v_3)\log 3+v_1+2v_2+3v_3\end{aligned}$$

which is bounded for all $\mathbf{x}\in R$ since all arguments of the log function are bounded away from 0.

It may be verified that ∂R consists of four faces, six edges and four corners. However, in view of Lemma 4, we only need to consider the part of ∂R that corresponds to $a_i > 0$, in this case when $i = 5$. To determine this part we must solve the system

$$\begin{aligned} f_i(\mathbf{x}) &\geq 0,\ i=1,2,3,4,6,\\ f_5(\mathbf{x}) &= 0,\end{aligned}$$

which is equivalent to

$$\begin{aligned} x &\geq 0\\ y &\geq 0\\ z &\geq 0\\ 1/4-x-y-z &\geq 0\\ 1/2-y-2z &\geq 0\\ 3/4-x-2y-3z &= 0.\end{aligned}$$

It is easy to show that the only solution to the system is $\mathbf{c_1} = (0,0,1/4)$. Consequently, for $\mathbf{x_0}\in \partial R\backslash\{\mathbf{c_1}\}$, $\log F$ and $\mathbf{x_0}$ satisfy the hypotheses of Lemma 4. Hence F does not have any maximum on $\partial R\backslash\{\mathbf{c_1}\}$. Moreover $F\left(\mathbf{c_1}\right) = \frac{5^{5/4}}{2^{1/4}3^{1/2}}$ is strictly less than $F\left(\mathbf{x}_{\max}\right)$. ∎

Now that we have determined that $\mathbf{x}_{\max} = (9/64,\ 3/64,\ 1/64)$ is the maximum of F on R, in the following lemma we will show that the sum in equation (11) can be approximated within a small region around the maximum.

Lemma 6 *Let $B = B(\mathbf{x_{max}},\delta)$ be a ball centred at $\mathbf{x_{max}}$ with diameter δ, where $\mathbf{x_{max}} = (\frac{9}{64},\frac{3}{64},\frac{1}{64})$ and $\delta = n^{-2/5}$. Then with F and R as in (11),*

$$\sum_R \alpha(\mathbf{x})F^n(\mathbf{x}) \sim \sum_B \alpha(\mathbf{x})F^n(\mathbf{x}).$$

Proof. Write

$$\sum_R \alpha(\mathbf{x})F^n(\mathbf{x}) = \sum_B \alpha(\mathbf{x})F^n(\mathbf{x}) + \sum_{R\backslash B} \alpha(\mathbf{x})F^n(\mathbf{x}).$$

It will be shown that $\sum_{R\backslash B}\alpha(\mathbf{x})F^n(\mathbf{x}) = o(\alpha(\mathbf{x_{max}})F^n(\mathbf{x_{max}}))$.

For $\mathbf{x}\in B$, the Taylor expansion of F at $\mathbf{x_{max}}$ is

$$F^n(\mathbf{x}) = F^n(\mathbf{x_{max}})\times\exp\left(-n\left(as_1^2+bs_2^2+cs_3^2+ds_1s_2+es_1s_3+fs_2s_3\right)+O\big(n^{-1/5}\big)\right)$$

where $s_1 = x - 9/64, \quad s_2 = y - 3/64, \quad s_3 = z - 1/64$,
$a = 4976/225$, $b = 14912/675$, $c = 22352/675$, $d = 2368/75$, $e = 1952/75$, $f = 9536/675$.
For $\mathbf{x}^* \in \partial B$, where ∂B is the boundary of B, we note that the exponential factor is

$$O\left(e^{-n^{1/5}}\right) = o(1)\,.$$

Therefore $\alpha(\mathbf{x}^*)F^n(\mathbf{x}^*) \sim \alpha(\mathbf{x_{max}})F^n(\mathbf{x_{max}})o(1) = o(\alpha(\mathbf{x_{max}})F^n(\mathbf{x_{max}}))$ for $\mathbf{x}^* \in \partial B$.
Since F attains its maximum uniquely at $\mathbf{x_{max}}$ then for $\mathbf{x} \in R\backslash B$

$$\alpha(\mathbf{x})F^n(\mathbf{x}) = O\left(\max_{\mathbf{x}^* \in \partial B} \alpha(\mathbf{x}^*)F^n(\mathbf{x}^*)\right).$$

Thus $\sum_{R\backslash B} \alpha(\mathbf{x})F^n(\mathbf{x}) = o(\alpha(\mathbf{x_{max}})F^n(\mathbf{x_{max}}))$. ∎

The remaining work is to determine $\sum_B \alpha(\mathbf{x})F^n(\mathbf{x})$. Since the summation concentrates near the maximum, each term $\alpha(\mathbf{x})$ can be taken as $\alpha(\mathbf{x_{max}})$ with $\alpha(\mathbf{x}_{\max}) = \frac{2^{13}5}{3^3}\sqrt{2}$. Referring to the Taylor expansion of F as in the proof of Lemma 6 we have

$$\sum_B \alpha(\mathbf{x})F^n(\mathbf{x}) \sim \alpha(\mathbf{x}_{\max})F^n(\mathbf{x}_{\max}) \sum_B \exp\left(-n\left(as_1^2 + bs_2^2 + cs_3^2 + ds_1s_2 + es_1s_3 + fs_2s_3\right)\right).$$

The summation is a Riemann sum for the triple integral.

$$n^{3/2}\int_{-n^{1/10}}^{n^{1/10}}\int_{-n^{1/10}}^{n^{1/10}}\int_{-n^{1/10}}^{n^{1/10}} \exp\left(-\left(at_1^2 + bt_2^2 + ct_3^2 + dt_1t_2 + et_1t_3 + ft_2t_3\right)\right) dt_1dt_2dt_3,$$

where

$$t_1 = \frac{(x - 9/64)n}{\sqrt{n}} \quad t_2 = \frac{(y - 3/64)n}{\sqrt{n}} \quad t_3 = \frac{(z - 1/64)n}{\sqrt{n}}.$$

As $n \to \infty$, the range of integration can be extended to $\pm\infty$ without altering the main asymptotic term. Thus it is asymptotic to

$$n^{3/2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{n\infty} \exp\left(-\left(at_1^2 + bt_2^2 + ct_3^2 + dt_1t_2 + et_1t_3 + ft_2t_3\right)\right) dt_1dt_2dt_3.$$

The evaluation of the triple integral results in $(\pi)^{3/2}\frac{3^35}{2^{12}}\sqrt{\frac{2}{13}}$. By multiplying these together, (11) becomes

$$\begin{aligned}\mathbf{E}Y_4'(Y_4'-1) &\sim \frac{1}{4(\pi n)^{3/2}}\left(\frac{3^{1/2}}{2^{5/2}}\right)^n \times \frac{2^{13}5}{3^3}\sqrt{2} \times \left(\frac{5^{5/2}}{2^3 3^{1/2}}\right)^n \\ &\quad \times(\pi n)^{3/2}\frac{3^35}{2^{12}}\sqrt{\frac{2}{13}} \\ &\sim \frac{25}{\sqrt{13}}\left(\frac{5^{5/2}}{2^{11/2}}\right)^n. \end{aligned} \tag{19}$$

Note that $\mathbf{E}Y_4' \to \infty$ implies $\mathbf{E}Y_4'(Y_4'-1) \sim \mathbf{E}Y'^2_4$. Thus, by (19), we obtain that

$$\mathbf{E}Y'^2_4 \sim \frac{25}{4\sqrt{13}}\left(\mathbf{E}Y_4'\right)^2.$$

Since $\mathbf{Var}Y_4' = \mathbf{E}Y'^2_4 - (\mathbf{E}Y_4')^2$ the above gives the required result. ∎

4 The expectation conditioned on cycle distribution

Lemma 7 *Let $n \equiv 0 \bmod 4$ and let X_k denote the number of cycles of length k in $G(P)$ for $P \in \mathcal{P}_{n,4}$. Then for any finite sequence $j_1, \ldots, j_m$ of nonnegative integers,*

$$\frac{\mathbf{E}\left(Y_4' \left[X_1\right]_{j_1} \cdots \left[X_m\right]_{j_m}\right)}{\mathbf{E}Y_4'} \to \prod_{k=1}^{m} \left(\lambda_k \left(1+\delta_k\right)\right)^{j_k} \quad \text{as } n \to \infty,$$

with $\lambda_k = 3^k/2k$ and $\delta_k = (-1/5 + 2i/5)^k + (-1/5 - 2i/5)^k$.

Proof. To prove the lemma we first establish

$$\frac{\mathbf{E}\left(Y_4' X_k\right)}{\mathbf{E}Y_4'} \sim \frac{3^k + (-3/5 + 6i/5)^k + (-3/5 - 6i/5)^k}{2k}. \tag{20}$$

The number of ways to choose a cycle of length k in the pairing, with a distinguished point in a pair, is

$$\frac{n!(12)^k}{(n-k)!}. \tag{21}$$

This induces an orientation and also a distinguished edge called a *root edge* in the cycle.

Let C denote the set of pairs that corresponds to an oriented and rooted k-cycle, and similarly define S to be the set of pairs corresponding to a 3-star factor. Fix C and suppose $C \cap S$ consists of s_1 1-stars, s_2 2-stars and $k - 2s_1 - 3s_2 = s_0$ 0-stars lying at leaves of the 3-star factor (by 0-stars we mean isolated vertices). Note that since $d = 4$, there can be no 0-stars of the intersection lying at the centres of stars in the 3-star factor. The edges of C can then be classified into 3 types. The first are the edges not lying in the 3-star-factor. We denote this type of edges by 0. The second type are the s_1 1-stars and the first edges of the s_2 2-stars whilst the last type are the second edges of the s_2 2-stars. We denote them by 1 and 2 respectively. If we walk along C from the root edge, then we obtain a sequence $S_0 \in \{0, 1, 2\}^k$.

For fixed C and S_0, the number of ways to choose the centres of the remaining $n/4 - s_1 - s_2$ 3-stars, together with the points used is

$$\binom{n-k}{n/4 - s_1 - s_2} 4!^{n/4 - s_1 - s_2}, \tag{22}$$

while the number of ways to choose the points in the centres of $s_1 + s_2$ 3-stars is

$$2^{s_1+s_2}. \tag{23}$$

The number of leaves remaining for the 3-star factor is

$$3(n/4 - s_1 - s_2) + 2s_1 + s_2 = 3n/4 - s_1 - 2s_2.$$

Hence, the number of ways to select these leaves, including the points used, is

$$\frac{(3n/4 - s_1 - 2s_2)!}{3!^{n/4 - s_1 - s_2} 2^{s_1}} 2^{s_1} 4^{3n/4 - k + s_1 + s_2} 2^{k - 2s_1 - 3s_2}. \tag{24}$$

The number of ways to complete the pairing given C and S_0 is

$$N(5n/2 - 2k + 2s_1 + 4s_2). \tag{25}$$

Multiply equations (22)–(25), sum over all possible S_0 and then multiply by (21). This results in the number of pairings containing a 3-star factor and an oriented and rooted cycle

$$\sum_{S_0} \frac{n!(3n/4 - s_1 - 2s_2)!4^n 6^k}{(n/4 - s_1 - s_2)!(3n/4 - k + s_1 + s_2)!2^{s_1} 4^{s_2}} N(5n/2 - 2k + 2s_1 + 4s_2).$$

Dividing this by the number of pairings with a 3-star factor, which is

$$\frac{n!}{(n/4)!} 4^n N(5n/2),$$

and then evaluating asymptotically we obtain

$$\left(\frac{9}{5}\right)^k \sum_{S_0} \left(\frac{5}{9}\right)^{s_1} \left(\frac{25}{27}\right)^{s_2}. \tag{26}$$

To determine the summation we follow an approach used in [7]. We can view 0, 1 and 2 as three states in a Markov Chain where the final state is equal to the initial state. We observe here that

(i) 1 followed by 0 means we pass a 1-star and this contributes a factor 5/9.

(ii) 1 edge followed by 2 edge means we pass a 2-star and this contributes a factor 25/27.

Thus, for the 'transition matrix' given by

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 5/9 & 0 & 25/27 \\ 1 & 0 & 0 \end{pmatrix}$$

we have

$$\sum_{S_0} \left(\frac{5}{9}\right)^{s_1} \left(\frac{25}{27}\right)^{s_2} = \operatorname{Tr}\left(A^k\right).$$

Since the eigenvalues of A are $\gamma_1 = 5/3$, $\gamma_2 = -1/3 + 2i/3$ and $\gamma_3 = -1/3 - 2i/3$ then

$$\operatorname{Tr}\left(A^k\right) = \left(\frac{5}{3}\right)^k + \left(\frac{-1+2i}{3}\right)^k + \left(\frac{-1-2i}{3}\right)^k.$$

Therefore (26) now becomes

$$3^k + (-3/5 + 6i/5)^k + (-3/5 - 6i/5)^k.$$

Finally, dividing by $2k$ to remove the orientation and rooting of the cycle we obtain (20). A similar argument also works for higher moments. It is quite straightforward, so we omit details. The proof is complete. ∎

5 Proofs of Theorem 1 and Corollary 1

We now have enough ammunition to verify Corollary 1 using [9, Theorem 1]. But for Theorem 1 we require the extension, given by Janson [7, Theorem 1], to what the small subgraph conditioning method is really saying about the distribution of the main random variable under consideration (see also [8, Section 9], in particular Theorem 9.12.)

Proof of Theorem 1: First we prove the following theorem.

Theorem 4 *Let $n \equiv 0 \bmod 4$. Then for $P \in \mathcal{P}_{n,4}$, $G(P)$ a.a.s. has a 3-star factor. Moreover,*

$$\frac{Y_4'}{\mathbf{E}Y_4'} \xrightarrow{d} W = \Pi_{k=1}^{\infty} (1+\delta_k)^{Z_k} e^{-\lambda_k \delta_k} \text{ as } n \to \infty,$$

where Z_k are independent Poisson variables with $\mathbf{E}Z_k = \lambda_k$ for $k \geq 0$, where $\lambda_k = 3^k/2k$ and $\delta_k = (-1/5 + 2i/5)^k + (-1/5 - 2i/5)^k$.

Proof. We use[14, Theorem 4.1], which is basically a reformulation of [7, Theorem 1]. We only need to show that the random variable Y_4' satisfies the conditions (a) to (d) in the theorem. Since X_k is the number of short cycles of length k in a pseudograph coming from $\mathcal{P}_{n,4}$ then (a) is satisfied with $\lambda_k = 3^k/2k$, by Bollobás's result on short cycles in $\mathcal{P}_{n,d}$ [3], while the other parts are fulfilled by Lemma 7 and Theorem 3. The proof is complete. ∎

Now Theorem 1 comes directly from Theorem 4, reasoning as in the proof of Corollary 2. From argument in [7, page 375] or [8, Remark 9.25], we also obtain

$$\frac{\mathbf{E}Y_4}{\mathbf{E}Y_4'} \sim \exp\left(-\lambda_1\delta_1 - \lambda_2\delta_2\right) = \exp\left(\frac{57}{50}\right) \text{ and}$$

$$\frac{\mathbf{E}{Y_4}^2}{(\mathbf{E}Y_4)^2} \sim \exp\left(-\lambda_1\delta_1^2 - \lambda_2\delta_7^2\right) \frac{\mathbf{E}{Y_4'}^2}{(\mathbf{E}Y_2')^2} = \left(\frac{25}{4\sqrt{13}}\right) \exp\left(\frac{-231}{625}\right). \quad ∎$$

Proof of Corollary 1: The result for $d > 4$ now follows from the argument used in [13] to extend the almost sure hamiltonicity of random 3-regular graphs to $d \geq 4$. The best way to present this is to include some definitions and a theorem about contiguity that can be found in [7] and [14].

Let $(\mathcal{G}_n)_{n\geq 1}$ and $\left(\widehat{\mathcal{G}}_n\right)_{n\geq 1}$ be two sequences of probability spaces such $\mathcal{G}_n$ and $\widehat{\mathcal{G}}_n$ differ only in the probabilities. These sequences are said *contiguous* if a sequence of events A_n is a.a.s. in $\mathcal{G}_n$ if and only if it is a.a.s. in $\widehat{\mathcal{G}}_n$. Contiguity is denoted by $\mathcal{G}_n \approx \widehat{\mathcal{G}}_n$.

Let $\mathcal{G}$ and $\widehat{\mathcal{G}}$ be two probability spaces of random graphs on the same vertex set. The *sum* of $\mathcal{G}$ and $\widehat{\mathcal{G}}$, denoted by $\mathcal{G} \oplus \widehat{\mathcal{G}}$, is defined as the space whose elements are defined by random graphs $\mathcal{G} \bigcup \widehat{\mathcal{G}}$ where $G \in \mathcal{G}$ and $\widehat{G} \in \widehat{\mathcal{G}}$ are independent conditioned on being edge-disjoint. Janson [7, Theorem 10] (see also [14, Corollary 4.17]) observed that the argument in [13] gives

$$(\mathcal{G}_{n,d-4} \oplus \mathcal{G}_{n,4}) \approx \mathcal{G}_{n,d}, \quad \text{for } d \geq 5 \text{ and even } n. \tag{27}$$

A graph $G \in \mathcal{G}_{n,d-4} \oplus \mathcal{G}_{n,4}$ by Theorem 1 a.a.s has a 3-star factor contained in its $\mathcal{G}_{n,4}$ subgraph. The existence a.a.s. of 3-star factors in $G \in \mathcal{G}_{n,d}$ follows immediately from (27). ∎

Acknowledgement The authors are thankful to the referees for a number of comments that helped to improve the presentation of the manuscript.

References

[1] H. Assiyatun. *Large Subgraphs of Regular Graphs.* PhD thesis, The University of Melbourne, Australia, 2002.

[2] E.A. Bender and E.R. Canfield, The asymptotic number of labeled graphs with given degree sequences, *J. Combin. Theory, Ser. A* **24** (1978), 296–307.

[3] B. Bollobás, *Random graphs*, Academic Press, London, 1985.

[4] A. Frieze, M. Jerrum, M. Molloy, R. Robinson, and N.C. Wormald. Generating and counting Hamilton cycles in random regular graphs. *Journal of Algorithms*, 21:176–198, 1996.

[5] H. Garmo. *Random Railways and Cycles in Random Regular Graphs.* PhD thesis, Uppsala University, Sweden, 1998.

[6] H. Garmo. *The asymptotic distribution of long cycles in random regular graphs. Random Structures and Algorithms*, 15:43–92, 1999.

[7] S. Janson. Random regular graphs: Asymptotic distribution and contiguity. *Combinatorics, Probability and Computing*, 4:369–405, 1995.

[8] S. Janson, T. Łuczak and A. Ruciński, **Random graphs**, Wiley, New York, 2000.

[9] M. Molloy, H. Robalewska, R.W. Robinson, and N.C. Wormald. 1-factorisations of random regular graphs. *Random Structures & Algorithms*, 10:305–321, 1997.

[10] H.D. Robalewska. 2-factors in random regular graphs. *Journal of Graph Theory*, 23(3):215–224, 1996.

[11] R.W. Robinson and N.C. Wormald. Existence of long cycles in random cubic graphs. In D.M. Jackson and S.A. Vanstone, editors, *Enumeration and Design*, pages 251–270. Academic Press, Toronto, 1984.

[12] R.W. Robinson and N.C. Wormald. Almost all cubic graphs are hamiltonian. *Random Structures & Algorithms*, 3:117–125, 1992.

[13] R.W. Robinson and N.C. Wormald. Almost all regular graphs are hamiltonian. *Random Structures & Algorithms*, 5:363–374, 1994.

[14] N.C. Wormald. Models of random regular graphs. In J.D. Lamb and D.A. Preece, editors, *Survey in Combinatorics*, volume 276 of *London Mathematical Society Lecture Notes Series.* Cambridge University Press, Cambridge, 1999.