

Random 4-regular graphs have 3-star decompositions asymptotically almost surely

Michelle Delcourt *

Luke Postle †

March 12, 2018

Abstract

Barát and Thomassen conjectured in 2006 that the edges of every planar 4-regular 4-edge-connected graph can be decomposed into copies of the star with 3 leaves. Shortly afterward, Lai constructed a counterexample to this conjecture. Using the small subgraph conditioning method of Robinson and Wormald, we prove that a random 4-regular graph has an S_3 -decomposition asymptotically almost surely, provided the number of vertices is divisible by 3.

Keywords: Configuration Model, Random Regular Graphs, Small Subgraph Conditioning Method, Star Decompositions.

1 Introduction

A question that has garnered much study is whether the edges of a graph G can be decomposed into copies of a small fixed subgraph, say F . Of course, some natural divisibility conditions arise for such a decomposition, namely that $e(F)$ must divide $e(G)$. Kotzig observed [9] that if G is connected and $e(G)$ is even, then G decomposes into copies of S_2 , the star with 2 leaves. What happens for larger F ; in particular, are there natural conditions when F is isomorphic to the S_3 , the star with 3 leaves? Not much was known about this problem until Thomassen’s breakthrough results [14] on the weak 3-flow conjecture. In particular, we note the following theorem which follows from a more general theorem of Lovász, Thomassen, Wu, and Zhang [11].

Theorem 1.1. *If $F \simeq S_k$, the star with k leaves, and G is a d -edge-connected graph such that k divides $e(G)$ and $2 \leq k \leq \lceil d/2 \rceil$, then the edge set of G decomposes into copies of F .*

In fact, Theorem 1.1 is tight for $k \geq 3$. To see this, first note that if $k > d$, then K_k is a d -edge-connected graph with no S_k decomposition. For $k \leq d$ with $k \geq 3$ and $k > \lceil d/2 \rceil$, consider k copies of K_d with edges added so that the resulting graph G is d -regular and d -edge-connected. If there existed an S_k -decomposition of G (a decomposition of the edges of G into copies of S_k), then because $k > d/2$, such a decomposition would naturally partition the vertices into $\frac{d}{2k}v(G) = \frac{d^2}{2}$ centers of the stars and $\frac{2k-d}{2k}v(G) = \frac{d(2k-d)}{2}$ non-centers. However, the non-centers must form

*Corresponding Author, Department of Mathematics, University of Illinois, Urbana, Illinois 61801, USA delcour2@illinois.edu. Research supported by NSF Graduate Research Fellowship DGE 1144245.

†Combinatorics and Optimization Department, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada lpotle@uwaterloo.ca. Partially supported by NSERC under Discovery Grant No. 2014-06162.

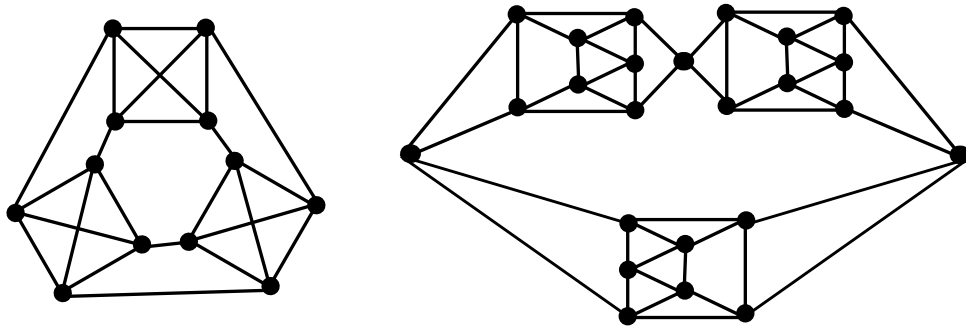


Figure 1: On the left is a non-planar 4-regular 4-edge-connected graph with no S_3 -decomposition. On the right is Lai's planar construction.

an independent set, and thus, there are at most k of them, the desired contradiction (because $k < \frac{d(2k-d)}{2}$ when $2k - d \geq 2$).

Thus, when F is isomorphic to S_3 , Theorem 1.1 implies that a d -regular d -edge-connected graph G has an F -decomposition if $d \geq 5$ and 3 divides $e(G)$. For $d = 3$, it is easy to observe that a 3-regular graph has an S_3 -decomposition if and only if it is bipartite. As for the case when $d = 4$, the construction in Figure 1 on the left provides a non-planar example of a 4-regular 4-edge-connected graph G where 3 divides $e(G)$ but G does not have an S_3 -decomposition. This led Barát and Thomassen [2], who knew of this example, to conjecture in 2006 that every planar 4-regular 4-edge-connected graph G where 3 divides $e(G)$ has an S_3 -decomposition. Unfortunately in the following year, Lai presented an infinite family of clever counterexamples (replicated in Figure 1 on the right) to this nice conjecture [10].

Given that a typical d -regular graph is d -edge-connected, a natural setting in which to study these questions is that of random regular graphs. We utilize the *configuration model* (also known as the *pairing model*) introduced by Bollobás [4]. Let $d \geq 1$ and dn be even; we take a total of dn points and partition them into n cells each consisting of exactly d points. Any perfect matching of $\frac{dn}{2}$ pairs of points is said to be a *configuration*, also known as a *pairing*. Each configuration corresponds to a multigraph (possibly with loops) where the cells are vertices and the pairs are edges. We denote the uniform probability space of configurations by $\mathcal{P}_{n,d}$. In the configuration model, we choose an element of $\mathcal{P}_{n,d}$ uniformly at random and discard the result if the corresponding d -regular multigraph has loops or parallel edges. This was shown to be equivalent to choosing a d -regular (simple) graph on n vertices uniformly at random (c.f. Wormald's survey paper [17] for more details).

Observe that in any simple 4-regular graph G , an orientation of the edges of G in which every in-degree is either 4 or 1 (alternatively every out-degree is either 0 or 3) is equivalent to an S_3 -decomposition, that is a decomposition of the edges of G into copies of S_3 ; namely, the vertices with out-degree 3 are the centers of the stars formed by their out-edges. In light of this, we consider orientations of the edges of a configuration where the out-degree of every cell is 0 or 3, where the *out-degree* of a cell is defined to be the number of points in the cell that are the tail of some edge in the orientation. We call such an orientation a $(3,0)$ -orientation.

The main result of this paper is as follows. Note that all asymptotics in this article are as n tends to infinity along positive integers divisible by 3.

Theorem 1.2. *A configuration in $\mathcal{P}_{n,4}$ has a $(3,0)$ -orientation asymptotically almost surely, provided that n is divisible by 3.*

Any 4-regular (simple) graph G on n vertices corresponds to exactly $(4!)^n = 24^n$ configurations in $\mathcal{P}_{n,4}$. Because each such graph corresponds to the same number of configurations, it follows that G is a (uniformly) random 4-regular (simple) graph in the configuration model. The probability that a configuration in $\mathcal{P}_{n,4}$ is simple tends to a positive constant as n tends to infinity (c.f. Wormald’s survey paper [17] for more details). Thus, we have the following as a corollary.

Corollary 1.3. *The edges of a random 4-regular (simple) graph on n vertices can be decomposed into copies of S_3 asymptotically almost surely, provided that n is divisible by 3.*

Our proof uses the small subgraph conditioning method of Robinson and Wormald [13]. We outline the proof of our main result in Section 2 before proving the remaining required individual lemmas in Sections 3, 4, and 5. However, first we note the connections between this problem and other interesting problems (as well as clarify some notation).

1.1 Extended History

There are many connections between orientations and decompositions; of particular interest is the following, known as the circular flow conjecture.

Conjecture 1.4 (Jaeger 1984 [6]). *Let $k \geq 3$ be odd. Every $(2k - 2)$ -edge-connected graph G has a mod k -orientation, that is, an orientation of $E(G)$ such that for every vertex the difference between its out-degree and in-degree is $0 \bmod k$.*

Jaeger proved that his conjecture reduces to the special case of odd regular graphs as follows.

Conjecture 1.5 (Jaeger 1988 [7]). *Let $k \geq 3$ be odd. Every $(2k - 2)$ -edge-connected, $(2k - 1)$ -regular graph has a mod k -orientation, that is, an orientation of $E(G)$ in which every in-degree is either $(3k - 1)/2$ or $(k - 1)/2$.*

Note that when $k = 3$, Conjecture 1.4 is actually equivalent to Tutte’s nowhere-zero 3-flow conjecture [16], one of the most famous open problems in structural graph theory. When $k = 5$, Jaeger’s conjecture implies the equally famous Tutte’s nowhere-zero 5-flow conjecture [15]. Thomassen [14] proved Conjecture 1.4 for multigraphs when the edge-connectivity is at least $2k^2 + k$. Lovász, Thomassen, Wu, and Zhang [11] later improved this and proved Conjecture 1.4 for graphs with edge-connectivity at least $3k - 3$. Despite these massive breakthroughs, proving Jaeger’s conjecture still seems intractable. Yet, as noted before, a typical $(2k - 1)$ -regular graph is $(2k - 1)$ -edge-connected, and therefore, a natural idea is to study Conjecture 1.5 in the setting of random $(2k - 1)$ -regular graphs. Using spectral techniques, Jaeger’s conjecture was confirmed to hold for random regular graphs provided that k is large enough as follows. The proof however makes use of the Expander Mixing Lemma and does not apply when k is too small.

Theorem 1.6 (Alon and Prałat 2011 [1]). *For large k , Jaeger’s conjecture holds asymptotically almost surely for random $(2k - 1)$ -regular graphs.*

Recently, utilizing the small subgraph conditioning method of Robinson and Wormald [13], Prałat and Wormald [12] were able to confirm Jaeger’s conjecture (Conjecture 1.5) for the case when $k = 3$, namely they proved the following theorem.

Theorem 1.7 (Prałat and Wormald 2015+ [12]). *Tutte’s nowhere-zero 3-flow conjecture holds asymptotically almost surely for random 5-regular graphs.*

Given these results, we were inspired to consider decompositions of random regular graphs, in particular whether the Barát-Thomassen conjecture might hold in the random case. Given Theorem 1.1, it is also natural to wonder more generally whether random d -regular graphs have S_k decompositions for some $k > \lceil d/2 \rceil$. We believe our methods could be applied to these questions.

As for other subgraphs F , Barát and Thomassen [2] conjectured in 2006 that for every tree T , there exists c_T such that every c_T -edge-connected graph has a decomposition of its edges into copies of $F \simeq T$. Theorem 1.1 confirmed this when T is a star and indeed gives the best possible value of c_T in that case. More recently, Bensmail, Harutyunyan, Le, Merker, and Thomassé [3] proved the conjecture for all trees T . However, determining what the best possible edge-connectivity is or, in the case of random regular graphs, what the best possible degree is, are still open problems.

1.2 Notation

Throughout this paper, if G is a multigraph, then we let $V(G)$ and $E(G)$ denote the vertex and edge sets of G respectively whereas $v(G)$ (or n) and $e(G)$ denote the number of vertices and edges in G . We say that an event $\mathcal{A} = \mathcal{A}(n)$ holds asymptotically almost surely (a.a.s.) if $\mathbb{P}[\mathcal{A}(n)] \rightarrow 1$ as $n \rightarrow \infty$ with the obvious necessary parity restrictions on n . For example in the case of finding S_3 -decompositions of 4-regular graphs, the necessary condition is that 3 divides n (and hence the number of edges is also divisible by 3). We denote the set $\{1, \dots, n\}$ by $[n]$ and the falling factorial $\frac{n!}{(n-j)!}$ by $(n)_j$.

2 Outline of the Proof of the Main Result

Let $Y = Y(n)$ denote the number of $(3, 0)$ -orientations of a random element of $\mathcal{P}_{n,4}$. In Section 3, we approximate $\mathbb{E}[Y]$ using Stirling’s approximation as follows.

Lemma 2.1.

$$\mathbb{E}[Y] \sim \frac{3}{\sqrt{2}} \left(\frac{27}{16} \right)^{n/3} > 0.$$

In order to show that configurations admit at least one $(3, 0)$ -orientation, we need to show that a.a.s. $Y > 0$. It is natural then to try to use the second moment method (coming from Chebyshev’s inequality) which says that if Y is a non-negative random variable and $\frac{\mathbb{E}[Y^2]}{\mathbb{E}[Y]^2} \rightarrow 0$ as $n \rightarrow \infty$, then a.a.s. $Y > 0$. To that end, we approximate $\mathbb{E}[Y^2]$ in Section 4 using optimization, Taylor expansions and multivariable integration to obtain the following.

Lemma 2.2.

$$\mathbb{E}[Y^2] \sim \frac{2\pi n}{9} \cdot \frac{81}{4\pi n} \sqrt{\frac{3}{2}} \cdot \left(\frac{27}{16} \right)^{2n/3} = \sqrt{\frac{3}{2}} \cdot \frac{9}{2} \left(\frac{27}{16} \right)^{2n/3}.$$

Unfortunately, the second moment method does not apply because $\mathbb{E}[Y]^2$ and $\mathbb{E}[Y^2]$ are of the same order, as the following corollary notes.

Corollary 2.3.

$$\frac{\mathbb{E}[Y^2]}{\mathbb{E}[Y]^2} \sim \sqrt{\frac{3}{2}} > 0.$$

Here the distribution of Y is affected by small but not too common (expected number is bounded) subgraphs of the random 4-regular graph, namely short cycles. In such situations we can attempt to apply the small subgraph conditioning method. When this method works, by conditioning on the small subgraph counts, we are able to control the variance of Y and in so doing show that $Y > 0$ asymptotically almost surely.

To understand how this works, consider partitioning the set of all 4-regular graphs on n vertices (with n divisible by 3) by the number of triangles. Within each partition class, the expected number of $(3,0)$ -orientations may be smaller than $\mathbb{E}[Y]$, though by at most a constant factor. Meanwhile the variance inside each class is smaller than the variance of Y . Applying the second moment method to the classes individually yields an increase in the probability that $Y > 0$, yet it still does not show that this probability tends to 1 asymptotically. So we further partition the classes by the number of 4-cycles, then by the number of 5-cycles, and so on. Surprisingly, by conditioning on the numbers of all cycles, it is possible to reduce the variance of Y to any desired fraction of $\mathbb{E}[Y]^2$. Intuitively, this seems plausible as graphs that have the same number of triangles, 4-cycles, etc. tend to have a similar structure and so admit less variance in the number of S_3 -decompositions. Thankfully we do not actually perform such an analysis, relying on the method of Robinson and Wormald [13]; for a proof see Janson's [8] paper.

Theorem 2.4 (Wormald 1999 [17]). *Let $\lambda_j > 0$ and $\delta_j > -1$ be real numbers, for all $j \geq 1$. Suppose for each n there are non-negative random variables $X_j = X_j(n)$, $j \geq 1$, and $Y = Y(n)$ (defined on the same probability space) such that X_j is integer valued and $\mathbb{E}[Y] > 0$ (for n sufficiently large). Furthermore, suppose that*

(1.) *for each $j \geq 1$, X_1, X_2, \dots, X_j are asymptotically independent Poisson random variables with*

$$\mathbb{E}[X_i] \sim \lambda_i, \text{ for all } i \in [j];$$

(2.)

$$\frac{\mathbb{E}[Y(X_1)_{\ell_1} \dots (X_j)_{\ell_j}]}{\mathbb{E}[Y]} \sim \prod_{i=1}^j (\lambda_i (1 + \delta_i))^{\ell_i}$$

for any fixed ℓ_1, \dots, ℓ_j where $(X)_\ell$ is the falling factorial;

(3.)

$$\sum_{i \geq 1} \lambda_i \delta_i^2 < \infty; \text{ and } \frac{\mathbb{E}[Y(n)^2]}{\mathbb{E}[Y(n)]^2} \leq \exp \left(\sum_{i \geq 1} \lambda_i \delta_i^2 \right) + o(1) \text{ as } n \rightarrow \infty.$$

Then, asymptotically almost surely $Y > 0$.

As in most applications of this method in the literature, we let X_j denote the number of cycles of length j in the multigraph corresponding to a random element of $\mathcal{P}_{n,4}$. Here, for $j \geq 1$, X_1, X_2, \dots, X_j are asymptotically independent Poisson random variables and

$$\mathbb{E}[X_j] \sim \lambda_j := \frac{3^j}{2 \cdot j}.$$

This is an immediate consequence of the following theorem of Bollobás.

Theorem 2.5 (Bollobás 1980 [4]). *For d fixed, let X_j denote the number of cycles of length j in the random multigraph resulting from a configuration in $\mathcal{P}_{n,d}$. For $j \geq 1$, X_1, \dots, X_j are asymptotically independent Poisson random variables with means $\lambda_j = \frac{(d-1)^j}{2 \cdot j}$.*

In Section 5, we compute $\mathbb{E}[YX_j]$ as follows by extending orientations of small cycles to orientations of the entire graph.

Lemma 2.6. *For all $j \geq 1$,*

$$\frac{\mathbb{E}[YX_j]}{\mathbb{E}[Y]} \sim \frac{3^j}{2 \cdot j} \left(1 + \left(-\frac{1}{3} \right)^j \right) = \lambda_j \left(1 + \left(-\frac{1}{3} \right)^j \right).$$

An easy observation from the first examinations of random graphs is that, for any fixed subgraph H with more edges than vertices, a multigraph corresponding to a random element of $\mathcal{P}_{n,4}$ asymptotically almost surely contains no subgraph isomorphic to H . Informally speaking, we would not expect to have two cycles sharing edges (or for that matter vertices). Therefore, we concentrate on disjoint cycles and roughly think of them as being independent. These observations combined with Lemma 2.6 imply the following more general form of Lemma 2.6, which computes the joint factorial moments.

Corollary 2.7. *For all $j \geq 1$, if (ℓ_1, \dots, ℓ_j) is a sequence of non-negative integers, then*

$$\frac{\mathbb{E}[Y(X_1)_{\ell_1} \dots (X_j)_{\ell_j}]}{\mathbb{E}[Y]} \sim \prod_{i=1}^j \left(\frac{3^i}{2 \cdot i} \left(1 + \left(-\frac{1}{3} \right)^i \right) \right)^{\ell_i} = \prod_{i=1}^j \left(\lambda_i \left(1 + \left(-\frac{1}{3} \right)^i \right) \right)^{\ell_i}.$$

From Lemma 2.6, $\frac{\mathbb{E}[YX_j]}{\mathbb{E}[Y]} \sim \lambda_j \left(1 + \left(-\frac{1}{3} \right)^j \right)$; thus, we set $\delta_j := -\left(\frac{1}{3} \right)^j > -1$ and verify the following.

Lemma 2.8.

$$\sum_{i \geq 1} \lambda_i \delta_i^2 < \infty \text{ and } \exp \left(\sum_{i \geq 1} \lambda_i \delta_i^2 \right) = \sqrt{\frac{3}{2}} \sim \frac{\mathbb{E}[Y^2]}{\mathbb{E}[Y]^2}.$$

Proof. Recall that $\lambda_i = \frac{3^i}{2 \cdot i}$. Using that $\sum_{i \geq 1} \frac{x^i}{i} = -\ln(1-x)$ for all $-1 < x < 1$, we obtain that

$$\sum_{i \geq 1} \lambda_i \delta_i^2 = \sum_{i \geq 1} \frac{3^i}{2 \cdot i} \cdot \left(-\frac{1}{3} \right)^{2i} = \frac{1}{2} \sum_{i \geq 1} \frac{\left(\frac{1}{3} \right)^i}{i} = \frac{1}{2} (-\ln(2/3)) < \infty.$$

Thus,

$$\exp \left(\sum_{i \geq 1} \lambda_i \delta_i^2 \right) = \exp \left(\frac{1}{2} (-\ln(2/3)) \right) = \sqrt{\frac{3}{2}}.$$

□

Modulo proofs of Lemma 2.1 (Section 3), Lemma 2.2 (Section 4), and Lemma 2.6 (Section 5), Theorem 2.4 now implies our main result as follows.

Proof of Main Result (Theorem 1.3). Let 3 divide n and $Y = Y(n)$ denote the number of $(3, 0)$ -orientations of a random element of $\mathcal{P}_{n,4}$. Let X_j denote the number of cycles of length j in a random element of $\mathcal{P}_{n,4}$. We apply Theorem 2.4 to Y and X_j . Note that (1.) holds by Theorem 2.5, (2.) holds by Corollary 2.7, and (3.) holds by Lemma 2.8. Thus $Y > 0$ asymptotically almost surely, as desired. \square

3 Expected Number of Decompositions

We let $Y = Y(n)$ denote the number of $(3, 0)$ -orientations of a random element of $\mathcal{P}_{n,4}$. We will make use of the following definition. Given n cells each consisting of 4 points, a *signature* is a set of $2n/3$ points no two belonging to the same cell. We call these points the *special points* of the signature. We refer to a cell as a *center* if it contains a special point and as a *leaf* otherwise. We say a point is an *in-point* if it is special or in a leaf of the signature and say it is an *out-point* otherwise.

We say a configuration in $\mathcal{P}_{n,4}$ *extends* a signature if the configuration forms a perfect matching between the in-points and the out-points of the signature. We note that a $(3, 0)$ -orientation extends exactly one signature. In this signature, the *centers* correspond to the $2n/3$ cells of out-degree 3 in the orientation (here the special point in each center is the head of the only incoming edge) and the *leaves* correspond to the remaining $n/3$ cells of out-degree 0 in the orientation.

To prove Lemma 2.1 though, we switch the order of counting and instead count the number of configurations that extend a given signature. We are now prepared to prove Lemma 2.1 as follows.

Proof of Lemma 2.1. There are a total of $\binom{n}{2n/3} 4^{2n/3}$ signatures for n cells of 4 points. Recall that a configuration extends a given signature if and only if the configuration matches the in-points of the signature with its out-points. Thus, there are $(\frac{4n}{2})! = (2n)!$ configurations that extend a given signature to a $(3, 0)$ -orientation. Using Stirling's approximation $s! \sim \sqrt{2\pi s} \left(\frac{s}{e}\right)^s$, we see that

$$\mathbb{E}[Y] = \frac{\binom{n}{2n/3} 4^{2n/3} (2n)!}{M(4n)} = 4^{5n/3} \frac{\binom{n}{2n/3}}{\binom{4n}{2n}} \sim \frac{3}{\sqrt{2}} \left(\frac{3^3}{2^4}\right)^{n/3} = \frac{3}{\sqrt{2}} \left(\frac{27}{16}\right)^{n/3} > 0,$$

where

$$M(4n) = \frac{(4n)!}{\left(\frac{4n}{2}\right)! \cdot 2^{(4n)/2}} = \frac{(4n)!}{(2n)! \cdot 2^{2n}}$$

is the number of perfect matchings of $4n$ points. \square

4 The Second Moment Method

In order to calculate $\mathbb{E}[Y^2]$ for Lemma 2.2, we should count the number of pairs of signatures that a given configuration extends. As in the proof of Lemma 2.1, we invert this count by fixing a pair of signatures σ_1 and σ_2 and then calculating how many configurations that they both jointly extend. To facilitate this count, we consider how the two signatures overlap. One might think that

there would be some necessary restriction on how the signatures overlap in order to guarantee the existence of even one configuration that they jointly extend, but strangely this is not the case.

Lemma 4.1. *For each A and B , there are*

$$\binom{n}{A, B, \frac{2n}{3} - A - B, \frac{2n}{3} - A - B, A + B - \frac{n}{3}} \cdot 4^{\frac{4n}{3} - A - B} \cdot 3^B$$

pairs of signatures such that the number of cells that are centers of both signatures with the same special point is A and the number of cells that are centers of both signatures with different special points is B .

Furthermore, for each such pair of signatures, there are

$$(3A + 2B)! \cdot (2n - 3A - 2B)!$$

configurations extending both signatures.

Proof of Lemma 4.1. Let A denote the number of cells that are centers in both σ_1 and σ_2 and have the same special point. Let B denote the number of cells that are centers in both σ_1 and σ_2 and have different special points. Note that $A + B$ is maximized when all centers in σ_1 are centers in σ_2 as well; thus, $A + B \leq \frac{2n}{3}$.

We see that we may write $\mathbb{E}[Y^2]$ in terms of n , A , and B as follows. Let C_1 denote the set of centers of σ_1 and C_2 the set of centers of σ_2 . Note that $|C_1 \cap C_2| = A + B$. Hence $|C_1 \setminus C_2| = \frac{2n}{3} - A - B = |C_2 \setminus C_1|$. The set of remaining cells is precisely the set of cells that are leaves in both signatures. There are a total of $n - |C_1 \cup C_2| = A + B - \frac{n}{3}$ of them. Since this is non-negative, we have that $A + B \geq \frac{n}{3}$. Hence, we see that for each possible value of A and B , there are

$$\binom{n}{A, B, \frac{2n}{3} - A - B, \frac{2n}{3} - A - B, A + B - \frac{n}{3}}$$

ways to partition the cells into these types.

There are $4^{2(\frac{2n}{3} - A - B)}$ ways to pick points that are in a leaf in one signature but a special point in the other signature. There are 4^A choices of special points from the centers in both σ_1 and σ_2 that have the same special point. Likewise, there are $(4 \cdot 3)^B$ ways to choose special points for the centers of both σ_1 and σ_2 with different special points. This proves the first assertion.

For the second assertion, note that a configuration extends both σ_1 and σ_2 if and only if the points that are in-points of both σ_1 and σ_2 are matched to points that are out-points of both σ_1 and σ_2 , and the points that are in-points of σ_1 and out-points of σ_2 get matched to points that are out-points of σ_1 but in-points of σ_2 and vice versa. There are $3A + 2B$ points that are out-points of both signatures since $3A$ of them are contained in centers with the same special point and $2B$ of them are contained in centers with different special points. Since there are $2n$ in-points and $2n$ out-points in each signature, it follows that there are $2n - 3A - 2B$ points that are an out-point in σ_1 and an in-point in σ_2 . Similarly there are $2n - 3A - 2B$ points that are an in-point in σ_1 and an outpoint in σ_2 . Hence, there are $3A + 2B$ points that are in-points in both signatures. Thus there are $(3A + 2B)!$ to match the the points that are out-points of both signatures to the points that are in-points of both signatures and $(2n - 3A - 2B)!$ to match the remaining points. The second assertion now follows. \square

Corollary 4.2.

$$\mathbb{E}[Y^2] = \sum_{A,B} \frac{(2n)! \cdot n! \cdot 4^{\frac{7n}{3}} \cdot 3^B \cdot (3A+2B)! \cdot (2n-3A-2B)!}{(4n)! \cdot 4^{A+B} \cdot A! \cdot B! \cdot ((\frac{2n}{3} - A - B)!)^2 \cdot (A+B-\frac{n}{3})!}$$

where A and B are non-negative integers such that $\frac{n}{3} \leq A+B \leq \frac{2n}{3}$.

Proof of Corollary 4.2. The computation goes as follows. We range over all possibilities of how two signatures may overlap, i.e. we range over A and B . Using Lemma 4.1 and tidying the formula gives that

$$\begin{aligned} \mathbb{E}[Y^2] &= \frac{1}{M(4n)} \sum_{\text{pairs of signatures}} \# \text{configurations extending both} \\ &= \frac{2^{2n}(2n)!}{(4n)!} \sum_{A,B} \binom{n}{A, B, \frac{2n}{3}-A-B, \frac{2n}{3}-A-B, A+B-\frac{n}{3}} \\ &\quad \cdot 4^{2(\frac{2n}{3}-A-B)} \cdot 4^A \cdot (4 \cdot 3)^B \cdot (3A+2B)! \cdot (2n-3A-2B)! \\ &= \sum_{A,B} \frac{(2n)! \cdot n! \cdot 4^{\frac{7n}{3}} \cdot 3^B \cdot (3A+2B)! \cdot (2n-3A-2B)!}{(4n)! \cdot 4^{A+B} \cdot A! \cdot B! \cdot ((\frac{2n}{3} - A - B)!)^2 \cdot (A+B-\frac{n}{3})!} \end{aligned}$$

□

It is useful to normalize A and B by letting $a = A/n$ and $b = B/n$. We let L denote the region

$$L := \left\{ \left(\frac{A}{n}, \frac{B}{n} \right) \in \mathbb{R}^2 : A, B \in \mathbb{Z} \cap [0, 2n/3] \text{ and } \frac{n}{3} \leq A+B \leq \frac{2n}{3} \right\}.$$

Thus the sum in Corollary 4.2 ranges over L . We will also need to consider points of L but without the restriction of A and B being integral. Thus we let R denote the region

$$R := \left\{ (a, b) \in \mathbb{R}^2 : 0 \leq a, b \leq \frac{2}{3} \text{ and } \frac{1}{3} \leq a+b \leq \frac{2}{3} \right\}.$$

To prove Lemma 2.2, we will apply Stirling's formula to the formula in Corollary 4.2. Doing so will yield an exponential part and a polynomial part.

To that end, we introduce the two following functions. First for the exponential part, let us define

$$\begin{aligned} f(a, b) &:= b(\ln 3 - \ln 4) + (2 - 3a - 2b) \ln(2 - 3a - 2b) + (3a + 2b) \ln(3a + 2b) - a(\ln a + \ln 4) \\ &\quad - b \ln b - 2 \left(\frac{2}{3} - a - b \right) \ln \left(\frac{2}{3} - a - b \right) - \left(a + b - \frac{1}{3} \right) \ln \left(a + b - \frac{1}{3} \right) - \frac{2}{3} \ln 4, \end{aligned}$$

and therefore,

$$e^{f(a,b)} = \frac{3^b \cdot (2 - 3a - 2b)^{(2-3a-2b)} \cdot (3a + 2b)^{(3a+2b)}}{4^{\frac{2}{3}+a+b} \cdot a^a \cdot b^b \cdot \left(\frac{2}{3} - a - b \right)^{2(\frac{2}{3}-a-b)} \cdot \left(a + b - \frac{1}{3} \right)^{(a+b-\frac{1}{3})}}.$$

Now for the polynomial part, let us define

$$g(a, b) := \frac{1}{2\pi} \cdot \sqrt{\frac{(3a+2b) \cdot (2-3a-2b)}{2 \cdot (a + \frac{1}{6n}) \cdot (b + \frac{1}{6n}) \cdot (\frac{2}{3} - a - b + \frac{1}{6n})^2 \cdot (a + b - \frac{1}{3} + \frac{1}{6n})}}.$$

We are now ready to apply Stirling's formula to Corollary 2.2, where recall that the formula is $s! = (1 + O(\frac{1}{s})) \sqrt{2\pi s} (\frac{s}{e})^s$. In fact, we apply a variant Stirling's formula known as Gosper's formula which is $s! = (1 + O(\frac{1}{s})) \sqrt{\pi(2s + \frac{1}{3})} (\frac{s}{e})^s$. We do this because Stirling's formula approximates $0!$ as 0 instead of 1 , which is unwieldy for division.

Lemma 4.3.

$$\mathbb{E}[Y^2] = \sum_{(a,b) \in L} S(a, b) \cdot \frac{g(a, b)}{n} \cdot e^{f(a, b)n}$$

where $S(a, b)$ is the error term arising from the applications of Gosper's formula.

Proof of Lemma 4.3. Using Gosper's formula $s! = (1 + O(\frac{1}{s})) \sqrt{2\pi(s + \frac{1}{6})} (\frac{s}{e})^s$, we have the following (where $S(\frac{A}{n}, \frac{B}{n})$ denotes the error factor arising from using Gosper approximations in the calculation below).

Thus,

$$\begin{aligned} \mathbb{E}[Y^2] &= \sum_{(\frac{A}{n}, \frac{B}{n}) \in L} \frac{(2n)! \cdot n! \cdot 4^{\frac{7n}{3}} \cdot 3^B \cdot (3A+2B)! \cdot (2n-3A-2B)!}{(4n)! \cdot 4^{A+B} \cdot A! \cdot B! \cdot (\frac{2n}{3} - A - B)! \cdot (\frac{2n}{3} - A - B)! \cdot (A+B - \frac{n}{3})!} \\ &= \sum_{(\frac{A}{n}, \frac{B}{n}) \in L} S\left(\frac{A}{n}, \frac{B}{n}\right) \cdot \sqrt{\frac{2^4 \cdot \pi^4 \cdot (2n + \frac{1}{6}) \cdot (n + \frac{1}{6}) \cdot (3A+2B + \frac{1}{6}) \cdot (2n-3A-2B + \frac{1}{6})}{2^6 \cdot \pi^6 \cdot (4n + \frac{1}{6}) \cdot (A + \frac{1}{6}) \cdot (B + \frac{1}{6}) \cdot (\frac{2n}{3} - A - B + \frac{1}{6})^2 \cdot (A+B - \frac{n}{3} + \frac{1}{6})}} \cdot e^{f(\frac{A}{n}, \frac{B}{n})n} \\ &\sim \sum_{(a,b) \in L} S(a, b) \cdot \frac{1}{2\pi n} \cdot \sqrt{\frac{(3a+2b) \cdot (2-3a-2b)}{2 \cdot (a + \frac{1}{6n}) \cdot (b + \frac{1}{6n}) \cdot (\frac{2}{3} - a - b + \frac{1}{6n})^2 \cdot (a + b - \frac{1}{3} + \frac{1}{6n})}} \cdot e^{f(a, b)n} \\ &= \sum_{(a,b) \in L} S(a, b) \cdot \frac{g(a, b)}{n} \cdot e^{f(a, b)n}. \end{aligned}$$

□

4.1 Multivariate Calculus

In order to approximate $\mathbb{E}[Y^2]$, we first need to determine the global maximum of f on the region L and since we use continuous techniques, we will instead find the global maximum of f on R . To approximate the function f , we then will take the Taylor expansion of f around the point attaining the global maximum, since the maximum value (as we will show) is unique. We extend the definition of f continuously to the boundary to the boundary of R by defining $x \ln x := 0$ when $x = 0$. We prove the following.

Lemma 4.4. *The global maximum of f on the region R is $2 \ln(3) - \frac{4}{3} \ln(4)$. This value is uniquely achieved at $P_0 = (a_0, b_0) = (\frac{1}{9}, \frac{1}{3})$. Furthermore, the Hessian matrix, $D^2 f(P_0)$, has determinant 81 and is negative definite.*

Proof of Lemma 4.4. First we investigate the interior of R . To examine the stationary points of f , we start by computing the first partials of f :

$$\frac{\partial f}{\partial a} = -3 \ln(2 - 3a - 2b) + 3 \ln(3a + 2b) - \ln a - \ln 4 + 2 \ln \left(\frac{2}{3} - a - b \right) - \ln \left(a + b - \frac{1}{3} \right)$$

$$\text{and } \frac{\partial f}{\partial b} = \ln 3 - \ln 4 - 2 \ln(2 - 3a - 2b) + 2 \ln(3a + 2b) - \ln b + 2 \ln \left(\frac{2}{3} - a - b \right) - \ln \left(a + b - \frac{1}{3} \right).$$

By setting $\frac{\partial f}{\partial a} = 0$, exponentiating both sides, and rearranging, we obtain

$$(3a + 2b)^3 \left(\frac{2}{3} - a - b \right)^2 = 4a(2 - 3a - 2b)^3 \left(a + b - \frac{1}{3} \right). \quad (1)$$

By setting $\frac{\partial f}{\partial b} = 0$, exponentiating both sides, and rearranging we obtain $3a(2 - 3a - 2b) = b(3a + 2b)$; this simplifies to

$$6a - 9a^2 = 9ab + 2b^2. \quad (2)$$

Using the quadratic equation to solve (2) yields

$$b = -\frac{9a}{4} \pm \frac{1}{4} \sqrt{9a^2 + 48a}.$$

Since the negative solution yields no point in R , let us define $b^*(a) = -\frac{9a}{4} + \frac{1}{4} \sqrt{9a^2 + 48a}$. Let $h(a) = (3a + 2b^*(a))^3 \left(\frac{2}{3} - a - b^*(a) \right)^2 - 4a(2 - 3a - 2b^*(a))^3 \left(a + b^*(a) - \frac{1}{3} \right)$. Now one can check that the only zeros of $h(a)$ with $0 \leq a \leq \frac{2}{3}$ are $a = 0$, $a = \frac{1}{9}$ and $a = \frac{2}{3}$. However, $b^*(0) = 0$ and hence $a = 0$ yields a common solution $(0, 0)$ of (1) and (2) that is not in R . Thus the only common solutions of (1) and (2) that lie in R are $(\frac{1}{9}, \frac{1}{3})$ and $(\frac{2}{3}, 0)$. We compute that $f(\frac{1}{9}, \frac{1}{3}) = 2 \ln(3) - \frac{4}{3} \ln(4) \approx 0.348832$ and $f(\frac{2}{3}, 0) = \ln(3) - \frac{2}{3} \ln(4) = \frac{1}{2} \cdot f(\frac{1}{9}, \frac{1}{3}) \approx 0.174416$.

Now we examine the boundary of R . As for the corner points, $f(\frac{2}{3}, 0) \approx 0.174416$ as above, $f(\frac{1}{3}, 0) = \ln(3) - \ln(4) \approx -0.287768$, $f(0, \frac{2}{3}) = -\frac{1}{3} \ln(3) \approx -0.366204$ and $f(0, \frac{1}{3}) = \frac{2}{3} \ln(\frac{4}{3}) \approx 0.191788$.

So we consider the interiors of the segments of the boundary. First consider the segment $a = 0$ and $\frac{1}{3} \leq b \leq \frac{2}{3}$. Note that $\frac{d^2}{db^2} f(0, b) = \frac{2}{1-b} + \frac{1}{b} - \frac{2}{\frac{2}{3}-b} - \frac{1}{b-\frac{1}{3}}$, which for $\frac{1}{3} < b < \frac{2}{3}$ is at most $3 + 3 - 9 < 0$. Hence $f(0, b)$ is concave when $\frac{1}{3} < b < \frac{2}{3}$ and so there is one local maximum in the interior of the segment. This point can be found by setting $\frac{d}{db} f(0, b) = 0$, which occurs at $b = b_1 \approx 0.393226$ at which point $f(0, b_1) \approx 0.253396$.

Next consider the segment $a = \frac{1}{3} - b$ and $0 \leq b \leq \frac{1}{3}$. $\frac{d^2}{db^2} f(\frac{1}{3} - b, b) = \frac{1}{1+b} + \frac{1}{1-b} - \frac{1}{b} - \frac{1}{\frac{1}{3}-b}$, which for $0 < b < \frac{1}{3}$ is at most $1 + \frac{3}{2} - 6 < 0$. Hence $f(\frac{1}{3} - b, b)$ is concave when $0 < b < \frac{1}{3}$ and so there is one local maximum in the interior of the segment. This point can be found by setting $\frac{d}{db} f(\frac{1}{3} - b, b) = 0$, which occurs at $b = b_2 \approx 0.280776$ at which point $f(\frac{1}{3} - b_2, b_2) \approx 0.245950$.

Next consider the segment $a = \frac{2}{3} - b$ and $0 \leq b \leq \frac{2}{3}$. One can check that $f(\frac{2}{3} - b, b)$ is decreasing when $0 < b < \frac{2}{3}$ since $\frac{d}{db} f(\frac{2}{3} - b, b) = \ln \left(\frac{2-3b}{2-b} \right) < 0$ for $0 < b < \frac{2}{3}$. Hence $f(\frac{2}{3} - b, b)$ is maximized for $0 \leq b \leq \frac{2}{3}$ when $b = 0$; as above $f(\frac{2}{3}, 0) \approx 0.174416$.

Next consider the segment $b = 0$ and $\frac{1}{3} \leq a \leq \frac{2}{3}$. One can check that $f(a, 0)$ is increasing when $\frac{1}{3} < a < \frac{2}{3}$ since $\frac{d}{da}f(a, 0) = \ln\left(\frac{9a^2}{4(2-3a)(3a-1)}\right) > 0$ for $\frac{1}{3} < a < \frac{2}{3}$. Hence $f(a, 0)$ is maximized for $\frac{1}{3} \leq a \leq \frac{2}{3}$ when $a = \frac{2}{3}$; as above, $f\left(\frac{2}{3}, 0\right) \approx 0.174416$.

Therefore, the unique global maximum occurs at $P_0 = (a_0, b_0) = \left(\frac{1}{9}, \frac{1}{3}\right)$. This proves the first assertion. Note that this corresponds to setting $3a + 2b = 2 - 3a - 2b$ and therefore $3A + 2B = 2n - 3A - 2B = n$. In other words, the number of points that are in-points of one signature and out-points of the other is equal to the number of points that are either in-points of both signatures or out-points of both signatures.

To compute the Hessian, first we take second partials of f and evaluate at P_0 :

$$\begin{aligned}\frac{\partial^2 f}{\partial a^2} &= \frac{9}{2-3a-2b} + \frac{9}{3a+2b} - \frac{1}{a} - \frac{2}{\frac{2}{3}-a-b} - \frac{1}{a+b-\frac{1}{3}}, \\ \frac{\partial^2 f}{\partial a \partial b} &= \frac{\partial^2 f}{\partial b \partial a} = \frac{6}{2-3a-2b} + \frac{6}{3a+2b} - \frac{2}{\frac{2}{3}-a-b} - \frac{1}{a+b-\frac{1}{3}}, \\ \text{and } \frac{\partial^2 f}{\partial b^2} &= \frac{4}{2-3a-2b} + \frac{4}{3a+2b} - \frac{1}{b} - \frac{2}{\frac{2}{3}-a-b} - \frac{1}{a+b-\frac{1}{3}}.\end{aligned}$$

Note that

$$\frac{\partial^2 f}{\partial a^2}\Big|_{P_0} = -9, \quad \frac{\partial^2 f}{\partial a \partial b}\Big|_{P_0} = \frac{\partial^2 f}{\partial b \partial a}\Big|_{P_0} = -6, \quad \text{and} \quad \frac{\partial^2 f}{\partial b^2}\Big|_{P_0} = -13.$$

Thus, Hessian matrix of f evaluated at P_0 is

$$H := D^2f(P_0) = \begin{pmatrix} -9 & -6 \\ -6 & -13 \end{pmatrix}.$$

The determinant of H is 81 with eigenvalues $-11 + 2\sqrt{10}$ and $-11 - 2\sqrt{10}$; thus, H is negative definite (this also implies that P_0 must be a local maximum). \square

4.2 Integrating

Our next lemma, Lemma 4.5, approximates the sum in Lemma 4.3 in terms of $\det D^2f(P_0)$, $g(P_0)$ and $e^{f(P_0)n}$. This is done using a Taylor expansion around P_0 combined with multivariable Gaussian integrals to calculate $\mathbb{E}[Y^2]$ more precisely as follows. The theory used in this proof has already been codified into a black box theorem by Greenhill, Janson and Ruciński [5], specifically Theorem 2.3 and its special case Theorem 6.3.

Indeed, given that by Lemma 4.4, f has a unique maximum at some interior point of R and the Hessian $D^2f(P_0)$ is strictly negative definite, our Lemma 4.5 follows easily from Theorem 6.3 of [5] by letting $\mathcal{L} = \mathbb{Z}^2$, $r = 2$, K be our set R defined above, $\phi = f$, $\psi = g$ and $\ell_n = 0$ for each positive integer n . Nevertheless, for the sake of the reader, we include a proof of Lemma 4.5 for completeness.

Lemma 4.5.

$$\mathbb{E}[Y^2] \sim \frac{2\pi}{\sqrt{|\det D^2f(P_0)|}} \cdot g(P_0) \cdot e^{f(P_0)n}.$$

Proof of Lemma 4.5. By Lemma 4.3,

$$\mathbb{E}[Y^2] = \sum_{(a,b) \in L} S(a,b) \cdot \frac{g(a,b)}{n} \cdot e^{f(a,b)n},$$

where $S(a,b)$ is the error term arising from the applications of Gosper's formula.

As before, we denote the Hessian matrix of f evaluated at P_0 by $H := D^2 f(P_0)$. We denote the gradient vector of f evaluated at P_0 by $D := Df(P_0) = [0, 0]$. We integrate near the maximum using a second-order Taylor series expansion. Let $[P - P_0]$ denote a row vector with components $[(a - a_0), (b - b_0)]$, and let $[P - P_0]^T$ be the transpose, a column vector. This gives that $f(a, b) = f(P)$ near P_0 is

$$\begin{aligned} f(P) &= f(P_0) + D[P - P_0]^T + \frac{1}{2}[P - P_0]H[P - P_0]^T + O(\|P - P_0\|^3) \\ &= f(P_0) + \frac{1}{2}[P - P_0]H[P - P_0]^T + O(\|P - P_0\|^3). \end{aligned}$$

By Taylor's Theorem, we note that the error is valid provided that $\|P - P_0\| = o(1)$.

Let $R' = \{P \in R : \|P - P_0\| = o(n^{-1})\}$ and $L' = L \cap R'$. Note that for all $P \in R'$, $S(P)g(P) \sim g(P_0)$ because $S(P) \sim 1$ and $\|P - P_0\|^3 = o(n^{-3})$. Thus,

$$\sum_{P \in L'} S(P) \cdot \frac{g(P)}{n} \cdot e^{f(P)n} \sim \frac{g(P_0)}{n} \cdot \sum_{P \in L'} e^{f(P)n} \sim \frac{g(P_0)}{n} \cdot e^{f(P_0)n} \cdot \sum_{P \in L'} \exp\left(\frac{1}{2}[P - P_0]H[P - P_0]^T n\right).$$

We note then that if we divide the sum by a factor of n^2 , then this becomes a Riemann sum over R' . That Riemann sum in turn approximates an integral as $n \rightarrow \infty$. Hence

$$\sum_{P \in L'} \exp\left(\frac{1}{2}[P - P_0]H[P - P_0]^T n\right) \sim n^2 \cdot \int \int_{P \in R'} \exp\left(\frac{1}{2}[P - P_0]H[P - P_0]^T n\right) dP.$$

Now we change variables by letting $x = (a - a_0)\sqrt{n}$ and $y = (b - b_0)\sqrt{n}$ for $P = (a, b)$. Note that the region of x, y corresponding to R' is the whole real plane. Thus, this change of variable transforms the integral into

$$\int \int_{P \in R'} \exp\left(\frac{1}{2}[P - P_0]H[P - P_0]^T n\right) dP \sim \frac{1}{n} \cdot \int \int_{\mathbb{R}^2} \exp\left([x, y] \frac{H}{2} [x, y]^T\right) dx dy,$$

where H is the Hessian matrix of f evaluated at P_0 . Diagonalizing and using the Gaussian integral, that is $\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$, we see that the integral evaluates to

$$\frac{1}{n} \cdot \sqrt{\frac{\pi^2}{|\det \frac{H}{2}|}} = \frac{1}{n} \cdot \sqrt{\frac{4 \cdot \pi^2}{|\det H|}} = \frac{2\pi}{n\sqrt{|\det H|}}.$$

Therefore,

$$\sum_{P \in L'} S(P) \cdot \frac{g(P)}{n} \cdot e^{f(P)n} \sim \frac{2\pi}{\sqrt{|\det H|}} \cdot g(P_0) \cdot e^{f(P_0)n}.$$

Because H is negative definite, the value of f on the boundary of R' is $f(P_0) - \Omega(n^{-2})$. However, f is independent of n and P_0 is a global maximum. Thus,

$$\max_{P \in R \setminus R'} f(P) = f(P_0) - \Omega(n^{-2}).$$

Observe that

$$\begin{aligned} \mathbb{E}[Y^2] &= \sum_{(a,b) \in L} S(a,b) \cdot \frac{g(a,b)}{n} \cdot e^{f(a,b)n} \\ &= \sum_{(a,b) \in L \setminus L'} S(a,b) \cdot \frac{g(a,b)}{n} \cdot e^{f(a,b)n} + \sum_{(a,b) \in L'} S(a,b) \cdot \frac{g(a,b)}{n} \cdot e^{f(a,b)n} \\ &\sim \sum_{(a,b) \in R \setminus R'} S(a,b) \cdot \frac{g(a,b)}{n} \cdot e^{f(a,b)n} + \frac{2\pi}{\sqrt{|\det H|}} \cdot g(P_0) \cdot e^{f(P_0)n}. \end{aligned}$$

Now consider $P = (a,b) \in L \setminus L'$. Since $L \setminus L' \subseteq R \setminus R'$, we have that $e^{f(a,b)n} = e^{f(P_0)n} \cdot \exp(-\Omega(n^{1/2}))$. Yet $S(a,b) = O(1)$ and $g(a,b) = O(n^{5/2})$ as each of the terms in the denominator of g are $O(n)$. Thus for each $(a,b) \in L \setminus L'$, we see that

$$S(a,b) \cdot \frac{g(a,b)}{n} \cdot e^{f(a,b)n} = e^{f(P_0)n} \cdot \exp(-\Omega(n^{1/2})).$$

Note that as there are only a polynomial number of points in $L \setminus L'$, namely at most n^2 , the sum over points in $L \setminus L'$ is also $e^{f(P_0)n} \cdot \exp(-\Omega(n^{1/2}))$. Therefore, $\mathbb{E}[Y^2] \sim \frac{2\pi}{\sqrt{|\det H|}} \cdot g(P_0) \cdot e^{f(P_0)n}$, as desired. \square

We are ready to prove Lemma 2.2.

Proof of Lemma 2.2. By Lemma 4.5,

$$\mathbb{E}[Y^2] \sim \frac{2\pi}{\sqrt{|\det D^2 f(P_0)|}} \cdot g(P_0) \cdot e^{f(P_0)n}.$$

Note that $\det D^2 f(P_0) = 81$ by Lemma 4.4. Moreover,

$$g(P_0) = \frac{1}{2\pi} \cdot \sqrt{\frac{1 \cdot 1}{2 \cdot \frac{1}{3} \cdot \frac{1}{9} \cdot \left(\frac{2}{9}\right)^2 \cdot \frac{1}{9}}} = \frac{1}{2\pi} \cdot \sqrt{\frac{3 \cdot 9^4}{8}} = \frac{81}{4\pi} \cdot \sqrt{\frac{3}{2}}.$$

Finally, $f(P_0) = 2 \ln(3) - \frac{4}{3} \ln(4)$. Hence

$$\mathbb{E}[Y^2] \sim \frac{2\pi}{9} \cdot g(P_0) \cdot e^{f(P_0)n} = \frac{2\pi}{9} \cdot \frac{81}{4\pi} \cdot \sqrt{\frac{3}{2}} \cdot \left(\frac{27}{16}\right)^{2n/3} = \sqrt{\frac{3}{2}} \cdot \frac{9}{2} \left(\frac{27}{16}\right)^{2n/3}.$$

Therefore, $\frac{\mathbb{E}[Y^2]}{\mathbb{E}[Y]^2} \sim \sqrt{\frac{3}{2}}$. \square

5 Joint Factorial Moments

In this section, we prove Lemma 2.6. By definition,

$$\mathbb{E}[YX_j] = \frac{1}{M(4n)} \sum_{j\text{-cycle } C} (\# \text{ orientations of cycle } C) \cdot (\# \text{ extensions of orientations of } C).$$

Note that this is equivalent to

$$\mathbb{E}[YX_j] = \frac{1}{M(4n)} \sum_{\text{oriented } j\text{-cycle } C} \# \text{ extensions of orientations of } C.$$

By counting how many configurations extend such oriented cycles, we will prove the following.

Lemma 5.1. *The number of oriented cycles with s sinks and s sources is*

$$\frac{(n)_j}{j} \binom{j}{2s} (4 \cdot 3)^j,$$

while the number of extensions to $(3, 0)$ -orientations for any oriented cycle is

$$\binom{n-j}{\frac{2n}{3}-j+s} 4^{\frac{2n}{3}-j+s} \cdot 2^s (2n-j)!$$

Therefore, we have that

$$\mathbb{E}[YX_j] = \frac{(n)_j}{M(4n) \cdot j} 4^{2n/3} 3^j (2n-j)! \sum_{s=0}^{\lfloor j/2 \rfloor} \binom{j}{2s} \binom{n-j}{\frac{2n}{3}-j+s} 2^{3s}.$$

Proof. Any oriented cycle must have the same number of sources (vertices with out-degree equal to 2) and sinks (vertices with out-degree equal to 0); an oriented cycle of length j can have s sources, s sinks and $j - 2s$ other vertices for some $0 \leq s \leq \lfloor \frac{j}{2} \rfloor$. The number of oriented cycles of length j with exactly s sources and s sinks is

$$\frac{(n)_j}{j} \binom{j}{2s} (4 \cdot 3)^j.$$

To see this, choose a set of j vertices ($\binom{n}{j}$ ways). The number of cyclic permutations of j entries is $\frac{(j-1)!}{2}$, where we divide by 2 for reversing the cycle and we can pick sources and sinks in $2 \binom{j}{2s}$ ways (the sources and sinks must alternate around the cycle), since once the sources and sinks have been chosen, then the orientation of the whole cycle is determined. Finally, every vertex in the cycle needs to pick two of its 4 points in ordered fashion from the configuration for endpoints of edges in the cycle.

Let C be a cycle of length j ; vertices that are sinks in C cannot have out-degree 3 and so are not centers. All other vertices of C must be centers. Thus the number of leaves in C is s and the number of centers in C is $j - s$. The number of extensions is then given by first completing the signature.

To this end, we choose $\frac{2n}{3} - (j - s)$ of the $n - j$ vertices outside of C to be centers; this can be done in $\binom{n-j}{\frac{2n}{3}-j+s}$ ways. Then we choose a special point for each such center; this can be done in $4^{\frac{2n}{3}-j+s}$ ways. For each source, for the non-cycle edges we must orient one edge out and one edge; this can be done in 2^s ways. The points that are ends of edges of the cycle are already matched. To extend this to a $(3, 0)$ -orientation, we match the remaining in-points and out-points, of which there are $2n - j$ of each; hence this can be done in $(2n - j)!$ ways. Combining all of these together, we find that the number of extensions to $(3, 0)$ -orientations for any oriented cycle is

$$\binom{n-j}{\frac{2n}{3}-j+s} 4^{\frac{2n}{3}-j+s} \cdot 2^s \left(\frac{4n-2j}{2} \right)! = \binom{n-j}{\frac{2n}{3}-j+s} 4^{\frac{2n}{3}-j+s} \cdot 2^s (2n-j)!$$

Therefore the whole expression is

$$\begin{aligned} \mathbb{E}[YX_j] &= \frac{1}{M(4n)} \sum_{s=0}^{\lfloor j/2 \rfloor} \frac{(n)_j}{j} \binom{j}{2s} (4 \cdot 3)^j \binom{n-j}{\frac{2n}{3}-j+s} 4^{\frac{2n}{3}-j+s} \cdot 2^s (2n-j)! \\ &= \frac{(n)_j}{M(4n) \cdot j} 4^{2n/3} 3^j (2n-j)! \sum_{s=0}^{\lfloor j/2 \rfloor} \binom{j}{2s} \binom{n-j}{\frac{2n}{3}-j+s} 2^{3s}. \end{aligned}$$

□

Recall that

$$\mathbb{E}[Y] = \frac{\binom{n}{2n/3} 4^{2n/3} (2n)!}{M(4n)}.$$

We are now ready to prove Lemma 2.6 as follows.

Proof of Lemma 2.6. In the computation of $\frac{\mathbb{E}[YX_j]}{\mathbb{E}[Y]}$, we use the following approximation where y is a constant and x goes to infinity,

$$\frac{x!}{(x-y)!} \sim \frac{\sqrt{2\pi x} \left(\frac{x}{e}\right)^x}{\sqrt{2\pi(x-y)} \left(\frac{x-y}{e}\right)^{x-y}} \sim \left(\frac{x}{e}\right)^y \cdot \left(\frac{x}{x-y}\right)^{x-y} = \left(\frac{x}{e}\right)^y \cdot \left(1 + \frac{y}{x-y}\right)^{x-y} \sim x^y.$$

Thus,

$$\begin{aligned} \frac{\mathbb{E}[YX_j]}{\mathbb{E}[Y]} &= \frac{n!}{j \cdot (n-j)!} 3^j \sum_{s=0}^{\lfloor j/2 \rfloor} \binom{j}{2s} \frac{\binom{n-j}{\frac{2n}{3}-j+s}}{\binom{n}{2n/3}} \frac{(2n-j)!}{(2n)!} 2^{3s} \\ &= \frac{3^j}{j} \sum_{s=0}^{\lfloor j/2 \rfloor} \binom{j}{2s} \frac{(\frac{2n}{3})!}{(\frac{2n}{3}-j+s)!} \frac{(\frac{n}{3})!}{(\frac{n}{3}-s)!} \frac{(2n-j)!}{(2n)!} 2^{3s} \\ &\sim \frac{3^j}{j} \sum_{s=0}^{\lfloor j/2 \rfloor} \binom{j}{2s} \left(\frac{2n}{3}\right)^{j-s} \left(\frac{n}{3}\right)^s \frac{2^{3s}}{(2n)^j} = \frac{1}{j} \sum_{s=0}^{\lfloor j/2 \rfloor} \binom{j}{2s} 2^{2s}. \end{aligned}$$

Note that $\binom{j}{2s}$ is the coefficient of x^{2s} in $q(x) := (1+x)^j$, so

$$\begin{aligned}\frac{\mathbb{E}[YX_j]}{\mathbb{E}[Y]} &\sim \frac{1}{j} \sum_{s=0}^{\lfloor j/2 \rfloor} \binom{j}{2s} 2^{2s} = \frac{1}{j} \cdot \frac{(q(2) + q(-2))}{2} = \frac{1}{2 \cdot j} (3^j + (-1)^j) \\ &= \frac{3^j}{2 \cdot j} \left(1 + \left(-\frac{1}{3} \right)^j \right) = \lambda_j \left(1 + \left(-\frac{1}{3} \right)^j \right).\end{aligned}$$

□

6 Acknowledgments

The authors would like to thank Bernard Lidický for useful discussion and helpful comments. We would also like to thank the anonymous referees for helping to improve the clarity of our exposition.

References

- [1] N. Alon and P. Pralat, Modular orientations of random and quasi-random regular graphs, *Combin. Probab. Comput.*, **20** (2011), 321–329.
- [2] J. Barát and C. Thomassen, Claw-decompositions and Tutte-orientations, *J. Graph Theory*, **52** (2006), 135–146.
- [3] J. Bensmail, A. Harutyunyan, T.-N. Le, M. Merker, and S. Thomassé, A proof of the Barát-Thomassen conjecture, *J. Combin. Theory Ser. B*, **124** (2017), 39–55.
- [4] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, *European J. Combin.*, **1** (1980), 311–316.
- [5] C. Greenhill, S. Janson, and A. Ruciński, On the number of perfect matchings in random lifts, *Combin. Probab. Comput.*, 19(5-6), 2010, 791–817.
- [6] F. Jaeger, On circular flows in graphs, *Finite and infinite sets*, Vol. I, II (Eger, 1981), vol. 37 of *Colloq. Math. Soc. János Bolyai*, North-Holland, Amsterdam, 1984, 391–402.
- [7] F. Jaeger, Nowhere-zero flow problems, *Selected topics in graph theory*, **3**, Academic Press, San Diego, CA, 1988, 71–95.
- [8] S. Janson, Random regular graphs: asymptotic distributions and contiguity, *Combin. Probab. Comput.*, **4** (1995), 369–405. 16
- [9] A. Kotzig, From the theory of finite regular graphs of degree three and four, *Casopis Pěst. Mat.*, **82** (1957), 76–92 (in Slovak).
- [10] H.-J. Lai, Mod $(2p + 1)$ -orientations and $K_{1,2p+1}$ -decompositions, *SIAM J. Discrete Math.*, **21** (2007), 844–850.
- [11] L.M. Lovász, C. Thomassen, Y. Wu, and C.-Q. Zhang, Nowhere-zero 3-flows and modulo k -orientations, *J. Combin. Theory Ser. B*, **103** (2013), 587–598.

- [12] P. Prałat and N. Wormald, Almost all 5-regular graphs have a 3-flow, arXiv:1503.03572.
- [13] R.W. Robinson and N.C. Wormald, Almost all cubic graphs are Hamiltonian, *Random Structures Algorithms*, **3** (1992), 117–125.
- [14] C. Thomassen, The weak 3-flow conjecture and the weak circular flow conjecture, *J. Combin. Theory Ser. B*, **102** (2012), 521–529.
- [15] W.T. Tutte, A contribution to the theory of chromatic polynomials, *Canadian J. Math.*, **6** (1954), 80–91.
- [16] W.T. Tutte, On the algebraic theory of graph colorings, *J. Combinatorial Theory*, **1** (1966), 15–50.
- [17] N.C. Wormald, Models of random regular graphs, *Surveys in combinatorics*, 1999 (Canterbury), vol. 267 of London Math. Soc. Lecture Note Ser., Cambridge Univ. Press, Cambridge, 1999, 239–298.