On the chromatic number of random d-regular graphs

Graeme Kemkes*
Department of Mathematics
University of California, San Diego
La Jolla, CA 92093-0112, US
gdkemkes@uwaterloo.ca

Xavier Pérez-Giménez†
Department of Combinatorics and Optimization
University of Waterloo
Waterloo ON, Canada
xperez@uwaterloo.ca

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

Abstract

In this work we show that, for any fixed d, random d-regular graphs asymptotically almost surely can be coloured with k colours, where k is the smallest integer satisfying $d < 2(k-1)\log(k-1)$. From previous lower bounds due to Molloy and Reed, this establishes the chromatic number to be asymptotically almost surely $k-1$ or k. If moreover $d > (2k-3)\log(k-1)$, then the value $k-1$ is discarded and thus the chromatic number is exactly determined. Hence we improve a recently announced result by Achlioptas and Moore in which the chromatic number was allowed to take the value $k+1$. Our proof applies the small subgraph conditioning method to the number of equitable k-colourings, where a colouring is *equitable* if the number of vertices of each colour is equal.

1 Introduction

The chromatic number χ of random graphs is a topic that has attracted considerable interest since the breakthrough achieved by Shamir and Spencer [19], which marked one of the first applications of martingales in combinatorics. For the classical Erdős–Rényi model $\mathcal{G}(n,p)$, a

*Partially supported by NSERC CGS and PDF awards. Some of this research appeared in this author's PhD thesis at the University of Waterloo.

†Partially supported by the Province of Ontario under the Post-Doctoral Fellowship (PDF) Program.

‡Supported by the Canada Research Chairs Program and NSERC.

celebrated result by Bollobás [7] later extended by Łuczak [15] showed that if $pn \to \infty$ then asymptotically almost surely (a.a.s.)

$$\chi(\mathcal{G}(n,p)) \sim \frac{n\log(1/(1-p))}{2\log(np)}.$$

Here and in similar statements, an event occurs a.a.s. if its probability tends to 1 as n tends to infinity. For $p = c/n$, Achlioptas and Naor [3] proved that the chromatic number of $\mathcal{G}(n,p)$ is a.a.s. k or $k+1$ where k is the smallest positive integer with $2k\log k > c$. Moreover, they discarded the case k for roughly half of the values of c. (Here and throughout this paper all logarithms are natural.) In the same direction, Coja-Oghlan, Panagiotou and Steger [8] showed that a.a.s. $\chi(\mathcal{G}(n,p)) \in \{k, k+1, k+2\}$ for $p < n^{-3/4-\epsilon}$ where k is the smallest positive integer satisfying $2k\log k > p(n-1)$. Meanwhile, some other results gave concentration of the chromatic number without determining the values so precisely: Łuczak [16] proved that $\chi(\mathcal{G}(n,p))$ is a.a.s. two point concentrated if $p < n^{-5/6-\epsilon}$, and later Alon and Krivelevich [4] extended this to $p < n^{-1/2-\epsilon}$.

More recently, results have been published about the chromatic number for the model $\mathcal{G}_{n,d}$ of random d-regular graphs, which is the probability space on d-regular graphs with n vertices having uniform distribution. For basic results and notation on random regular graphs, see [22]. Hereinafter, dn is always assumed to be even for feasibility. For fixed d, Molloy and Reed [17] showed that if $q(1-1/q)^{d/2} < 1$ then $\chi(\mathcal{G}_{n,d}) > q$ a.a.s. Then, for $d < n^{1/3-\epsilon}$, Frieze and Łuczak [12] established that

$$\chi(\mathcal{G}_{n,d}) = \frac{d}{2\log d} + O\left(\frac{d\log\log d}{\log^2 d}\right),$$

and later Cooper, Frieze, Reed and Riordan [9] extended the same asymptotic formula to apply to $d \le n^{1-\epsilon}$. Similarly, the range $n^{6/7+\epsilon} \le d \le 0.9n$ was covered by Krivelevich, Sudakov, Vu and Wormald [14], who showed that $\chi(\mathcal{G}_{n,d}) \sim \frac{n}{2\log_b d}$ a.a.s. where $b = n/(n-d)$. Achlioptas and Moore [2] recently announced a significant new result for constant d. They stated that if k is the smallest integer satisfying $d < 2(k-1)\log(k-1)$ then a.a.s. $\chi(\mathcal{G}_{n,d})$ is $k-1$, k, or $k+1$. If, in addition, $d > (2k-3)\log(k-1)$, then a.a.s. $\chi(\mathcal{G}_{n,d})$ is k or $k+1$. They also established two point concentration of $\chi(\mathcal{G}_{n,d})$ for $d = d(n)$ bounded above by some small power of n, which was extended to $d = o(n^{1/5})$ by Ben-Shimon and Krivelevich [5].

In this paper we restrict the set of possible values for the chromatic number given by Achlioptas and Moore, and show that $\chi(\mathcal{G}_{n,d})$ a.a.s. cannot be $k+1$. Therefore this reduces the range of possibilities for $\chi(\mathcal{G}_{n,d})$ to only a.a.s. $k-1$ and k, in the first case, and establishes that $\chi(\mathcal{G}_{n,d}) = k$ a.a.s. in the second case. For example, it establishes the previously unknown result that a.a.s. $\chi(\mathcal{G}_{n,10}) = 5$ and $\chi(10^6) = 46523$. It also provides an alternate proof of the result of Shi and Wormald [21] that a.a.s. $\chi(\mathcal{G}_{n,6}) = 4$. Curiously, our result is not strong enough to prove that a.a.s. $\chi(\mathcal{G}_{n,4}) = 3$, which was established by Shi and Wormald [20]. We essentially need to show that $\mathcal{G}_{n,d}$ is a.a.s. k-colourable, since the above-mentioned lower bound of Molloy and Reed implies that $\mathcal{G}_{n,d}$ is a.a.s. not $(k-2)$-colourable, and for the second case, not $(k-1)$-colourable. Our basic approach for the upper bound is similar to that of Achlioptas and Moore, in that we analyse the second moment of the number Y

of equitable k-colourings of random regular graphs. In fact, Achlioptas and Moore found that the second central moment $\mathbf{E}(Y - \mathbf{E}Y)^2$ is (essentially, in a more or less equivalent model of random graphs) asymptotically a non-zero constant times the square of the first. Consequently, Chebyshev's inequality fails to show the result which we claim above. In cases like this, this failure of the second moment inequality to establish $Y > 0$ a.a.s. can, at least for random structures similar to $\mathcal{G}_{n,d}$, be overcome by using the small subgraph conditioning method of Robinson and the third author. (See [13, Chapter 9] and [22] for a full exposition of the method.) Using this, we show that $\mathcal{G}_{n,d}$ is a.a.s. k-colourable.

Theorem 1 *Given any integer $d \geq 3$, let k be the smallest integer such that $d < 2(k-1)\log(k-1)$. Then the chromatic number of $\mathcal{G}_{n,d}$ is a.a.s. $k-1$ or k. If furthermore $d > (2k-3)\log(k-1)$, then the chromatic number of $\mathcal{G}_{n,d}$ is a.a.s. k.*

In part simultaneously with this work, a similar approach was used by the authors together with Díaz, Kaporis and Kirousis [11] to show that, provided a certain *maximum* hypothesis about a specific function is true, a random 5-regular graph is a.a.s. 3-colourable. However, we were unable to verify the maximum hypothesis.

Actually, almost all previous applications of the small subgraph conditioning method were for a random variable that counted large subgraphs in the random graph. To apply the method in the present setting we need to calculate the first and second moments of the number of equitable k-colourings (n is then required to be divisible by k), as well as joint moments of the number of such colourings and the number of short cycles. These computations are done in the well-known *pairing* or *configuration model* $\mathcal{P}_{n,d}$ which was first introduced by Bollobás [6]. A *pairing* in $\mathcal{P}_{n,d}$ is a perfect matching on a set of dn points which are grouped into n cells of d points each. A random pairing naturally corresponds in an obvious way to a random d-regular multigraph (possibly containing loops or multiple edges), in which each cell becomes a vertex. Colourings of the multigraph then correspond to assignments of colours to the cells of the model. The reader should refer to [22] for aspects of the pairing model not explained here.

Proposition 2 *Fix integers $d, k \geq 3$. Let Y be the number of equitable k-colourings of a random d-regular multigraph $\mathcal{P}_{n,d}$ (where n is restricted to the set of multiples of k).*

(a) For $m \geq 1$, let X_m be the number of m-cycles in $\mathcal{P}_{n,d}$. Then

$$\mathbf{E}Y \sim k^{k/2}\left(\frac{k-1}{2\pi(k-2)}\right)^{(k-1)/2} n^{-(k-1)/2} k^n \left(1-\frac{1}{k}\right)^{dn/2}$$

and

$$\mathbf{E}(Y[X_1]_{p_1}\cdots[X_j]_{p_j}) \sim \prod_{m=1}^{j} \left(\lambda_m\left(1+\delta_m\right)\right)^{p_m} \mathbf{E}(Y) \tag{1.1}$$

where

$$\lambda_m = \frac{(d-1)^m}{2m} \qquad \text{and} \qquad \delta_m = \frac{(-1)^m}{(k-1)^{m-1}}.$$

(b) If furthermore $d < 2(k-1)\log(k-1)$ then

$$\mathbf{E}(Y^2) \sim \frac{k^k(k-1)^{k(k-1)}}{(k^2-2k-d+2)^{(k-1)^2/2}(2\pi(k-2))^{k-1}} n^{-(k-1)} k^{2n} \left(1-\frac{1}{k}\right)^{dn}.$$

We next compute

$$\sum_{m\geq 1} \lambda_m \delta_m^2 = (k-1)^2 \log\left(\frac{k-1}{\sqrt{k^2-2k-d+2}}\right), \tag{1.2}$$

and verify that for n divisible by k

$$\frac{\mathbf{E}(Y^2)}{(\mathbf{E}Y)^2} \sim \left(\frac{k-1}{\sqrt{k^2-2k-d+2}}\right)^{(k-1)^2} = \exp\left(\sum_{m\geq 1} \lambda_m \delta_m^2\right), \tag{1.3}$$

which is the last ingredient required for the application of the small subgraph conditioning method.

Proof of Theorem 1 (for n divisible by k). Assume throughout the proof that k divides n, and observe that all the conditions of Theorem 4.1 in [22] are verified by Proposition 2, (1.2) and (1.3). Thus we may apply the small subgraph conditioning method to conclude that $\mathbf{P}(Y > 0 \mid \mathcal{E}) \to 1$, where $\mathcal{E} = \bigwedge_{\delta_k=-1}\{X_k = 0\} = \{X_1 = 0\}$ is the event of having no loops. Because $\mathbf{P}(X_2 = 0)$ is bounded away from 0 for large n (see e.g. [22]), it follows that $Y > 0$ a.a.s. for the simple graphs in $\mathcal{G}_{n,d}$, thus proving the required upper bound on the chromatic number. Now observe that from our choice of k we have $d \geq 2(k-2)\log(k-2) > (2k-5)\log(k-2)$. Then the required lower bounds follow immediately from the fact that if $d > (2q-1)\log q$ then $\chi(\mathcal{G}_{n,d}) > q$ (applied to $q = k-2$ or $q = k-1$ for each case in the statement). This is just a slightly weaker formulation of the result given by Molloy and Reed [17]. (Their proof is reported in [21], Theorem 1.3.) ∎

The following two sections supply the proof of Proposition 2. Finally, in Section 4, we conclude the proof of Theorem 1 by extending the argument to general n.

2 Joint moments: proof of Proposition 2(a)

Let Y be the number of equitable k-colourings of a random d-regular multigraph $\mathcal{P}_{n,d}$. For $m \geq 1$, let X_m be the number of m-cycles in $\mathcal{P}_{n,d}$. We estimate the expected value of Y by enumerating all equitable k-colourings of all multigraphs in $\mathcal{P}_{n,d}$. There are $\binom{n}{n/k,n/k,\ldots,n/k}$ ways to choose the k colour classes. These choices are all equivalent so fix one. Suppose there are $b_{ij} = b_{ji}$ edges between colour class i and colour class j (for $1 \leq i,j \leq k$ and $i \neq j$). The colours of the neighbours of all of the points of colour class i can be then chosen in $(dn/k)!/\prod_{\substack{1\leq j\leq k\\ j\neq i}} b_{ij}!$ ways. After this determination is made, edges are constructed by putting a perfect matching between the corresponding points in each pair of classes, in

one of $\prod_{1\le i<j\le k} b_{ij}!$ ways. Thus we have

$$\begin{aligned}
|\mathcal{P}_{n,d}|\,\mathbf{E}(Y) &= \binom{n}{n/k, n/k, \ldots, n/k} \sum_{\{b_{ij}\}} \left(\prod_{i=1}^{k} \frac{(dn/k)!}{\prod_{\substack{1\le j\le k\\ j\ne i}} b_{ij}!} \right) \prod_{1\le i<j\le k} b_{ij}! \\
&= \binom{n}{n/k, n/k, \ldots, n/k} (dn/k)!^k \sum_{\{b_{ij}\}} \frac{1}{\prod_{1\le i<j\le k} b_{ij}!} \\
&= \binom{n}{n/k, n/k, \ldots, n/k} (dn/k)!^k \Big[\prod_{l=1}^{k} x_l^{dn/k} \Big] \prod_{1\le i<j\le k} \sum_{l\ge 0} \frac{(x_i x_j)^l}{l!} \\
&= \binom{n}{n/k, n/k, \ldots, n/k} (dn/k)!^k \Big[\prod_{l=1}^{k} x_l^{dn/k} \Big] \exp\Big(\sum_{1\le i<j\le k} x_i x_j \Big),
\end{aligned}$$

where square brackets denote the extraction of a coefficient from a generating function. A particular case of the following result gives us an accurate estimate of that coefficient. The proof is based on the saddlepoint method and is included later in this section.

Lemma 3 *Let $k, d, a_1, a_2, \ldots, a_k$ be fixed integers with $k \ge 3$, d positive, and $s = \sum_{j=1}^{k} a_j$ even. Let $C_{a_1,a_2,\ldots,a_k}$ denote the coefficient of $x_1^{dn/k+a_1} x_2^{dn/k+a_2} \cdots x_k^{dn/k+a_k}$ in the generating function $\exp(\sum_{1\le j<l\le k} x_j x_l)$. Then as $n \to \infty$ we have $C_{a_1,a_2,\ldots,a_k} \sim C(s)$, where*

$$C(s) = (2\pi)^{-k} \left(\frac{k(k-1)}{dn} \right)^{(dn+s)/2} 2e^{dn/2} (2\pi)^{k/2} \left(\frac{k(k-1)}{dn} \right)^{k/2} (2k-2)^{-1/2} (k-2)^{-(k-1)/2}.$$

Hence we deduce

$$|\mathcal{P}_{n,d}|\,\mathbf{E}(Y) \sim \binom{n}{n/k, n/k, \ldots, n/k} (dn/k)!^k C(0). \tag{2.1}$$

Combining this with the well-known formula for the number of pairs on dn points,

$$|\mathcal{P}_{n,d}| = (dn-1)!! = \frac{(dn)!}{(dn/2)! 2^{dn/2}} \sim \sqrt{2} \left(\frac{dn}{e} \right)^{dn/2}, \tag{2.2}$$

and after some basic manipulations using Stirling's formula we obtain the estimate for $\mathbf{E}(Y)$ stated in the proposition.

Next we estimate the expected value of YX_m where Y is the number of equitable k-colourings and X_m the number of length-m cycles. It is more convenient to count rooted oriented cycles, which introduces a factor of $2m$ into our calculations. It will be helpful to have the following definitions. For a rooted oriented cycle in a coloured graph, define its *colour type* to be the sequence T of colours on its vertices. For $j = 1, 2, \ldots, k$, let $\alpha_j(T)$ denote the number of vertices in T which have colour j. Note that the sum $\sum_j \alpha_j(T)$ is m.

To calculate the expected value of YX_m, we will count, for each equitable k-colouring and each rooted oriented m-cycle, the number of pairings which contain this cycle and respect this colouring.

As before, there are $\binom{n}{n/k,n/k,\ldots,n/k}$ ways to choose the equitable k-colouring. All are equivalent, so fix one. To enumerate the cycles and pairings which respect this colouring, we will sum over all colour types T. Once a colour type has been chosen, each vertex of the cycle can be placed in the pairing model by choosing a vertex of the correct colour and an ordered pair of points in that vertex to be used by the cycle. Hence, in total, there are asymptotically $(d(d-1)n/k)^m$ ways to place the rooted oriented cycle in the pairing model. We now have

$$\mathbf{E}(YX_m) \sim \frac{1}{2m}\binom{n}{n/k, n/k, \ldots, n/k}\left(\frac{d(d-1)n}{k}\right)^m \frac{1}{|\mathcal{P}_{n,d}|}\sum_T f(T),$$

where $f(T)$ is the number of pairings which respect a fixed equitable k-colouring and fixed rooted oriented cycle of colour type T. To count these pairings, suppose there are $b_{ij} = b_{ji}$ edges between colour class i and colour class j (for $1 \le i, j \le k$ and $i \ne j$), excluding the edges of the prescribed cycle. The colours of the neighbours of all of the unmatched points of colour class i can be then chosen in $(dn/k - 2\alpha_i(T))!/\prod_{\substack{1\le j\le k\\ i\ne j}} b_{ij}!$ ways. After this determination is made, edges are constructed by putting a perfect matching between the corresponding points in each pair of classes, in one of $\prod_{1\le i<j\le k} b_{ij}!$ ways. Thus we have

$$\begin{aligned}
f(T) &= \sum_{\{b_{ij}\}}\left(\prod_{i=1}^{k}\frac{(dn/k-2\alpha_i(T))!}{\prod_{\substack{1\le j\le k\\ i\ne j}} b_{ij}!}\right)\prod_{1\le i<j\le k} b_{ij}! \\
&= \sum_{\{b_{ij}\}}\frac{\prod_i (dn/k-2\alpha_i(T))!}{\prod_{i<j} b_{ij}!} \\
&\sim \frac{(dn/k)!^k}{(dn/k)^{2m}}\sum_{\{b_{ij}\}}\frac{1}{\prod_{1\le i<j\le k} b_{ij}!} \\
&\sim \frac{(dn/k)!^k}{(dn/k)^{2m}}\Big[\prod_{l=1}^{k} x_l^{dn/k-2\alpha_l(T)}\Big]\prod_{1\le i<j\le k}\sum_{l\ge 0}\frac{(x_ix_j)^l}{l!} \\
&\sim \frac{(dn/k)!^k}{(dn/k)^{2m}}\Big[\prod_{l=1}^{k} x_l^{dn/k-2\alpha_l(T)}\Big]\exp\left(\sum_{1\le i<j\le k} x_ix_j\right).
\end{aligned}$$

By Lemma 3 the asymptotic value of the coefficient in the last expression is $C(-2m)$, making the entire expression independent of T. Moreover, the number t_m of possible colour types for a rooted oriented cycle of length m satisfies the obvious recurrence $t_m + t_{m-1} = k(k-1)^{m-1}$ with $t_1 = 0$. So we have $t_m = (k-1)^m + (k-1)(-1)^m$ and therefore

$$\begin{aligned}
\mathbf{E}(YX_m) \sim \frac{1}{2m}\binom{n}{n/k, n/k, \ldots, n/k}\left(\frac{d(d-1)n}{k}\right)^m \frac{1}{|\mathcal{P}_{n,d}|}\frac{(dn/k)!^k}{(dn/k)^{2m}} \\
\times((k-1)^m + (k-1)(-1)^m)C(-2m).
\end{aligned}$$

Comparing this expression with (2.1) we see that

$$\begin{aligned}
\frac{\mathbf{E}(YX_m)}{\mathbf{E}(Y)} &\sim \frac{1}{2m}\left(\frac{d(d-1)n}{k}\right)^m ((k-1)^m + (k-1)(-1)^m)\frac{1}{(dn/k)^{2m}}C(-2m)/C(0) \\
&\sim \frac{1}{2m}\left(\frac{d(d-1)n}{k}\right)^m ((k-1)^m + (k-1)(-1)^m)\frac{1}{(dn/k)^{2m}}\left(\frac{k(k-1)}{dn}\right)^{-m} \\
&\sim \frac{(d-1)^m}{2m}\left(1 + \frac{(-1)^m}{(k-1)^{m-1}}\right). \\
&\sim \lambda_m(1+\delta_m).
\end{aligned}$$

The above argument is easily extended to work for higher moments, by counting the pairings that contain a given equitable k-colouring and set of oriented cycles of the appropriate lengths. The contribution from cases where the cycles intersect turns out to be negligible, for the following reasons. Suppose that the cycles form a subgraph H with ν vertices and μ edges, and the total length of cycles is ν_0. Then in the case of disjoint cycles, $\nu = \mu = \nu_0$. A factor of $\Theta(n^{\nu_0-\nu})$ is lost if there is a reduction in the number of vertices of H, compared with the disjoint case, because of the reduced number of ways of placing the cycles on the coloured vertices. Similarly, a factor $\Theta(n^{\nu_0-\mu})$ is gained in the function f for the reduction in the number of edges of H, because of the corresponding increase in the number of points to be paired up at the end. Thus, the contribution from such an arrangement of cycles to the quantity being estimated is of the order of $n^{\nu-\mu}$ times that of the contribution from disjoint cycles. In all non-disjoint cases, H has more edges than vertices, since its minimum degree is at least 2, and it has at least one vertex of degree at least 3. There are only finitely many isomorphism types of H to consider, so the contribution from the case of disjoint cycles is of the order of n times the rest. The significant terms in this case decompose into a product of the factors corresponding to the individual cycles, and we obtain

$$\mathbf{E}(Y[X_1]_{p_1}\cdots[X_j]_{p_j})/\mathbf{E}(Y) \sim \prod_{m=1}^{j}\left(\lambda_m\left(1+\delta_m\right)\right)^{p_m}$$

as claimed. ∎

It only remains to prove Lemma 3. Before doing so, we need the following result, which will be used several times in the paper.

Lemma 4 *Let k be a positive integer. Define the function $f : \mathbb{R}^k \to \mathbb{C}$ by*

$$f(\theta) = ia(n,\theta) - c_1 n\theta^\top B\theta$$

where i is the imaginary unit, a is a real function, B is a fixed k-by-k positive definite real matrix, and $c_1 > 0$ is a real constant. Let $\delta = c_2 n^{-1/2}\log n$ for some real constant $c_2 > 0$. Then, as $n \to \infty$,

$$\int_{[-\delta,\delta]^k} e^{f(\theta)}d\theta = \int_{[-\infty,\infty]^k} e^{f(\theta)}d\theta + O(e^{-c(\log n)^2})$$

for some constant $c > 0$.

Proof. Note that in order to bound the integral of $e^{f(\theta)}$ over $[-\infty,\infty]^k \setminus [-\delta,\delta]^k$ we only need to consider the real part of $f(\theta)$. Moreover, since B is positive definite we have $\theta^\top B\theta \geq \lambda|\theta|^2$ where $\lambda > 0$ is the smallest eigenvalue of B. The proof is elementary in view of these two observations. ∎

Proof of Lemma 3. We will use the saddlepoint method. First we use Cauchy's formula to express $C_{a_1,a_2,\ldots,a_k}$ as an integral over the product of circles $z_j = re^{i\theta_j}$, $-\pi \leq \theta_j \leq \pi$ for $j = 1,2,\ldots,k$, where $r = \sqrt{dn/k(k-1)}$.

$$\begin{aligned}
C_{a_1,a_2,\ldots,a_k} &= \frac{1}{(2\pi i)^k}\int_{|z_1|=r}\int_{|z_2|=r}\cdots\int_{|z_k|=r}\frac{\exp\left(\sum_{j<l} z_j z_l\right)}{z_1^{dn/k+a_1+1} z_2^{dn/k+a_2+1}\cdots z_k^{dn/k+a_k+1}}dz_1dz_2\cdots dz_k \\
&= \frac{1}{(2\pi)^k}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}\frac{\exp\left(\sum_{j<l}(re^{i\theta_j})(re^{i\theta_l})\right)}{(re^{i\theta_1})^{dn/k+a_1}(re^{i\theta_2})^{dn/k+a_2}\cdots(re^{i\theta_k})^{dn/k+a_k}}d\theta_1d\theta_2\cdots d\theta_k \\
&= \frac{1}{(2\pi)^k r^{dn+s}}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}\frac{\exp(r^2\sum_{j<l}e^{i(\theta_j+\theta_l)})}{\exp(i\sum_j(dn/k+a_j)\theta_j)}d\theta_1\cdots d\theta_k.
\end{aligned}$$

Let $g(\theta)$ denote the integrand in the last expression above. Letting $\mathbf{1}$ denote the vector of 1's, consider the image of $g(\theta)$ under the transformation $\theta \mapsto \theta + \pi\mathbf{1}$. It is clear that the numerator is fixed by this transformation. The denominator becomes

$$\begin{aligned}
\exp(i\sum_j(dn/k+a_j)(\theta_j+\pi)) &= \exp(i\sum_j(dn/k+a_j)\theta_j)\exp(i(dn+s)\pi) \\
&= \exp(i\sum_j(dn/k+a_j)\theta_j)
\end{aligned}$$

since dn (the sum of the vertex degrees) and s are both even. So $g(\theta)$ is fixed by this transformation. Letting $\delta = \log n/\sqrt{n}$, this means that the integrals of $g(\theta)$ over regions $\{\theta : |\theta_j| \leq \delta, j = 1,2,\ldots,k\}$ and $\{\theta : \pi-\delta \leq |\theta_j| \leq \pi, j = 1,2,\ldots,k\}$ are equal. We will prove that the integral I of $g(\theta)$ over each of these regions is asymptotically equal to

$$\begin{aligned}
I &= e^{dn/2}(2\pi)^{k/2}\left(\frac{k(k-1)}{dn}\right)^{k/2}(2k-2)^{-1/2}(k-2)^{-(k-1)/2} \\
&= K\exp\left(dn/2 - \frac{k}{2}\log n\right),
\end{aligned}$$

where K is a constant, and we will show that the integral over the remaining regions is asymptotically smaller. From these results the proposition follows.

To prove that the integral over vectors θ in the remaining regions is asymptotically smaller, there are two cases: either $|\theta_{j^*}| \leq \delta$ and $\pi-\delta \leq |\theta_{l^*}| \leq \pi$ for some distinct j^* and l^*, or $\delta \leq |\theta_{j^*}| \leq \pi-\delta$ for some j^*.

In the first case, suppose that $|\theta_{j^*}| \leq \delta$ and $\pi - \delta \leq |\theta_{l^*}| \leq \pi$ for some distinct j^* and l^*. Then $\pi - 2\delta \leq |\theta_{j^*} + \theta_{l^*}| \leq \pi + 2\delta$ and hence $\cos(\theta_{j*} + \theta_{l*}) \leq 0$. So

$$\begin{aligned}
|g(\theta)| &= \exp\left(r^2 \sum_{j<l} \cos(\theta_j + \theta_l)\right) \\
&\leq \exp\left(r^2(\binom{k}{2} - 1) + r^2 \cos(\theta_{j^*} + \theta_{l^*})\right) \\
&\leq \exp\left(r^2(\binom{k}{2} - 1)\right) \\
&= \exp\left(\frac{dn}{2} - \frac{dn}{k(k-1)}\right) \\
&= o(I).
\end{aligned}$$

In the second case we suppose that $\delta \leq |\theta_{j^*}| \leq \pi - \delta$ for some j^*. If there is a value of l^* for which $|\theta_{j^*} + \theta_{l*}| > \delta/2$ then $\delta/2 < |\theta_{j^*} + \theta_{l*}| < 2\pi - \delta/2$. This means

$$\cos(\theta_{j^*} + \theta_{l*}) < \cos(\delta/2) = 1 - \frac{\delta^2}{8} + O(\delta^4)$$

and hence

$$\begin{aligned}
|g(\theta)| &= \exp\left(r^2 \sum_{j<l} \cos(\theta_j + \theta_l)\right) \\
&\leq \exp\left(r^2(\binom{k}{2} - 1) + r^2 \cos(\theta_{j^*} + \theta_{l^*})\right) \\
&= \exp\left(r^2(\binom{k}{2} - 1) + r^2(1 - \frac{\delta^2}{8} + O(\delta^4))\right) \\
&= \exp\left(r^2\binom{k}{2} - r^2\frac{\delta^2}{8} + O(r^2\delta^4)\right) \\
&= \exp\left(\frac{dn}{2} - \frac{d(\log n)^2}{8k(k-1)} + o(1)\right) \\
&= o(I).
\end{aligned}$$

Otherwise, there is no such l^*. That is, for all l^* not equal to j^* we have $|\theta_{l^*} - (-\theta_{j^*})| \leq \delta/2$. This implies that all θ_l with $l \neq j^*$ have the same sign and satisfy $\delta/2 \leq |\theta_l| \leq \pi - \delta/2$. Since $k \geq 3$ we can choose two distinct such l, say l^* and l^{**}, and deduce

$$\delta \leq |\theta_{l^*} + \theta_{l^{**}}| \leq 2\pi - \delta.$$

Using the same argument as above, it follows that $|g(\theta)| = o(I)$.

This completes the proof that the integral of $g(\theta)$ over these regions is asymptotically negligible, as claimed.

It remains to show the integral of $g(\theta)$ over the region $\{\theta : |\theta_j| \leq \delta, j = 1, 2, \ldots, k\}$ is asymptotically equal to I. We begin by expanding

$$\begin{aligned}
\log g(\theta) &= r^2\binom{k}{2} + i(r^2(k-1) - dn/k + O(1))\sum_{j=1}^{k}\theta_j - \frac{1}{2}r^2\sum_{j<l}(\theta_j+\theta_l)^2 + O(r^2\sum_{j=1}^{k}|\theta_j|^3) \\
&= r^2\binom{k}{2} - \frac{1}{2}r^2\sum_{j<l}(\theta_j+\theta_l)^2 + o(1)
\end{aligned}$$

since $r^2 = dn/(k(k-1))$ and $|\theta_j| \leq \delta = \log n/\sqrt{n}$ for all j. The quadratic order term can be written as $-\frac{1}{2}r^2\sum_{j<l}(\theta_j+\theta_l)^2 = -\frac{1}{2}\theta^\top A\theta$. Here, $\theta^\top$ denotes the transpose of the column vector θ and A is the matrix $A = r^2(\mathbf{1}\mathbf{1}^\top + (k-2)I_k)$, where I_k is the k-by-k identity matrix. By Lemma 4 and since A is positive definite, we have for some constant $c > 0$

$$\begin{aligned}
&\int_{-\delta}^{\delta}\int_{-\delta}^{\delta}\cdots\int_{-\delta}^{\delta}\exp\left(-\frac{1}{2}r^2\sum_{j<l}(\theta_j+\theta_l)^2\right)d\theta_1 d\theta_2\cdots d\theta_k \\
= &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}\theta^\top A\theta\right)d\theta_1 d\theta_2\cdots d\theta_k + O(e^{-c\log^2 n}).
\end{aligned}$$

It is well-known (see Equation 4.6.3 in [10]) that such integrals have the value $(2\pi)^{k/2}(\det A)^{-1/2}$, giving us

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}r^2\sum_{j<l}(\theta_j+\theta_l)^2\right)d\theta_1 d\theta_2\cdots d\theta_k \\
= &(2\pi)^{k/2}\left(r^{2k}(2k-2)(k-2)^{k-1}\right)^{-1/2}.
\end{aligned}$$

We conclude

$$\begin{aligned}
\int_{-\delta}^{\delta}\int_{-\delta}^{\delta}\cdots\int_{-\delta}^{\delta}g(\theta)d\theta_1 d\theta_2\cdots d\theta_k &\sim e^{r^2\binom{k}{2}}(2\pi)^{k/2}\left(r^{2k}(2k-2)(k-2)^{k-1}\right)^{-1/2}. \\
&\sim e^{dn/2}(2\pi)^{k/2}\left(\frac{k(k-1)}{dn}\right)^{k/2}(2k-2)^{-1/2}(k-2)^{-(k-1)/2} \\
&= I
\end{aligned}$$

as claimed. ∎

3 Second moment: proof of Proposition 2(*b*)

Throughout this section, fix positive integers d and $k \geq 3$ satisfying $d < 2(k-1)\log(k-1)$. Assume 2 divides dn and k divides n. Let C_1 and C_2 be equitable k-colourings of a pairing $P \in \mathcal{P}_{n,d}$. The *colour count* of (C_1, C_2) is the k-by-k matrix $M = [m_{p,q}]$ where $m_{p,q}n/k$ is the number of cells coloured p in C_1 and coloured q in C_2. Let $\mathcal{M}$ be the set of k-by-k *doubly*

stochastic matrices (i.e. nonnegative matrices with each row sum and column sum equal to 1). Since the colourings are equitable, we must have $M \in \mathcal{M}$. Define $T(M)$ to be the set of triples (P, C_1, C_2) where $P \in \mathcal{P}_{n,d}$ and (C_1, C_2) is a pair of equitable k-colourings of P having colour count M. Then,

$$\mathbf{E}(Y^2) = \sum_{M \in \mathcal{M} \cap \frac{k}{n}\mathbb{Z}^{k^2}} \frac{|T(M)|}{|\mathcal{P}_{n,d}|}. \tag{3.1}$$

In order to estimate the sum in (3.1), we first obtain an exact expression for $|T(M)|$, where $M = [m_{p,q}]$ is any k-by-k doubly stochastic matrix whose entries are integer multiples of k/n.

For all $1 \le p, q \le k$ we must choose $m_{p,q}n/k$ cells to be assigned the colour p in the first colouring and q in the second colouring. We say that such a cell and its points have label (p, q). The number of ways of doing this is given by

the multinomial coefficient

$$\frac{n!}{\prod\limits_{1 \le p,q \le k} (m_{p,q}n/k)!}.$$

Now we must select the edges of the pairing in a way which is compatible with the two colourings. Suppose we know the number b_{pqrs} of edges from points labelled (p, q) to points labelled (r, s) for all $1 \le p, q, r, s \le k$ with $p \ne r$ and $q \ne s$. Then we choose, for each ordered pair of labels $((p, q), (r, s))$, which b_{pqrs} of the points labelled (p, q) will be paired with points labelled (r, s). The number of ways of doing this is

$$\prod_{1 \le p,q \le k} \frac{(dm_{p,q}n/k)!}{\prod\limits_{\substack{1 \le r,s \le k \\ r \ne p, s \ne q}} b_{pqrs}!}.$$

Finally, for each unordered pair of labels $\{(p, q), (r, s)\}$, we choose a bijection between the points labelled (p, q) and the points labelled (r, s) that were designated to be paired with each other. The number of ways of doing this is

$$\prod_{\substack{1 \le p,q,r,s \le k \\ p < r, q \ne s}} b_{pqrs}!.$$

Observe that the only restrictions on b_{pqrs} required in our counting are

$$b_{pqrs} = b_{rspq} \qquad \forall p, q, r, s \in \{1, \ldots, k\}, r \ne p, s \ne q, \tag{3.2}$$

$$\sum_{\substack{1 \le r,s \le k \\ r \ne p, s \ne q}} b_{pqrs} = dm_{p,q}n/k \qquad \forall p, q \in \{1, \ldots, k\}. \tag{3.3}$$

Thus, the total number of triples (P, C_1, C_2) in $T(M)$ is

$$|T(M)| = n! \left(\prod_{1 \le p,q \le k} \frac{(dm_{p,q}n/k)!}{(m_{p,q}n/k)!} \right) \sum_{\mathcal{B}(M,n)} \prod_{\substack{1 \le p,q,r,s \le k \\ p < r, q \ne s}} \frac{1}{b_{pqrs}!}, \tag{3.4}$$

where $\mathcal{B}(M,n)$ is the set of tuples of non-negative integers $(b_{pqrs})_{\substack{1\le p,q,r,s\le k\\ r\ne p, s\ne q}}$ satisfying (3.2) and (3.3). Note that (3.4) is the expression used in [2].

As we shall see later, the main weight of the sum in (3.1) corresponds to terms in which M is 'near' $(1/k)J_k$, where J_k denotes the k-by-k matrix of ones. To state this more precisely, for $\delta > 0$ and any positive integer p, we define $\mathsf{B}_p(\delta)$ to be the set of p-by-p matrices $M = [m_{ij}]$ for which $\max_{i,j} |m_{ij} - (1/k)| < \delta$. It will be shown that the essential contribution to (3.1) comes from terms such that $M \in \mathsf{B}_k(\epsilon \log n/n^{1/2})$, where $\epsilon > 0$ is a small constant that will be specified later. Thus we now proceed to bound $|T(M)|/|\mathcal{P}_{n,d}|$ for each $M \in \mathcal{M} \cap (k/n)\mathbb{Z}^{k^2}$, and then find more accurate asymptotic expressions for $M \in \mathsf{B}_k(\epsilon \log n/n^{1/2})$.

We begin by estimating the ratio of factorials in (3.4). Recall that one can write Stirling's formula as $x! = \xi(x)(x/e)^x$ where ξ is a function satisfying $\xi(x) \sim \sqrt{2\pi x}$ as $x \to \infty$ and $\xi(x) \ge 1$ for all $x \ge 0$. Thus,

$$\begin{aligned}
\prod_{1\le p,q\le k} \frac{(dm_{p,q}n/k)!}{(m_{p,q}n/k)!} &= \prod_{1\le p,q\le k} \frac{\xi(dm_{p,q}n/k)(dm_{p,q}n/(ke))^{dm_{p,q}n/k}}{\xi(m_{p,q}n/k)(m_{p,q}n/(ke))^{m_{p,q}n/k}} \\
&= \prod_{1\le p,q\le k} \frac{\xi(dm_{p,q}n/k)}{\xi(m_{p,q}n/k)} d^{dm_{p,q}n/k} \left(\frac{m_{p,q}n}{ke}\right)^{(d-1)m_{p,q}n/k} \\
&= d^{dn} \left(\frac{n}{ek}\right)^{(d-1)n} \prod_{1\le p,q\le k} \frac{\xi(dm_{p,q}n/k)}{\xi(m_{p,q}n/k)} {m_{p,q}}^{(d-1)m_{p,q}n/k}, \qquad (3.5)
\end{aligned}$$

where in the final step we used $\sum_{1\le p,q\le k} m_{p,q} = k$ which holds because M is doubly stochastic. (Throughout the article we use the convention $0^0 = 1$ and $0 \log 0 = 0$.) Moreover, since $\xi(m_{p,q}n/k) \ge 1$ and $\xi(dm_{p,q}n/k) = O(n^{1/2})$ for each $m_{p,q}$, we obtain the following bound, which does not depend on the particular M.

$$\prod_{1\le p,q\le k} \frac{(dm_{p,q}n/k)!}{(m_{p,q}n/k)!} = O\left(n^{k^2/2}\right) d^{dn} \left(\frac{n}{ek}\right)^{(d-1)n} \left(\prod_{1\le p,q\le k} {m_{p,q}}^{m_{p,q}}\right)^{(d-1)n/k}. \qquad (3.6)$$

Next we bound the inner sum in (3.4), by further applying Stirling's formula, and also using the obvious crude bound $|\mathcal{B}(M,n)| = O(n^{k^4})$ on the number of terms.

$$\begin{aligned}
\sum_{\mathcal{B}(M,n)} \prod_{\substack{1\le p,q,r,s\le k\\ p<r, q\ne s}} \frac{1}{b_{pqrs}!} &\sim \sum_{\mathcal{B}(M,n)} \prod_{\substack{1\le p,q,r,s\le k\\ p<r, q\ne s}} \frac{1}{\xi(b_{pqrs})(b_{pqrs}/e)^{b_{pqrs}}} \\
&\le \sum_{\mathcal{B}(M,n)} \prod_{\substack{1\le p,q,r,s\le k\\ p<r, q\ne s}} \frac{1}{(b_{pqrs}/e)^{b_{pqrs}}} \\
&= O\left(n^{k^4}\right) e^{dn/2} \max_{\mathcal{B}(M,n)} \left\{ \prod_{\substack{1\le p,q,r,s\le k\\ p<r, q\ne s}} \frac{1}{{b_{pqrs}}^{b_{pqrs}}} \right\}.
\end{aligned}$$

The following result allows us to derive a more explicit bound, conveniently expressed in terms of M and n. The proof follows the ideas in [2] and is given later in this section.

Lemma 5 *Let $M = [m_{pq}]$ be a fixed matrix in $\mathcal{M}$, and let $(b_{pqrs})_{\substack{1\le p,q,r,s\le k\\ p<r,q\neq s}}$ be any tuple of non-negative reals satisfying* (3.2) *and* (3.3)*. Then,*

$$\prod_{\substack{1\le p,q,r,s\le k\\ p<r,q\neq s}} \frac{1}{{b_{pqrs}}^{b_{pqrs}}} \le \left(\frac{1}{\prod_{1\le p,q\le k} {m_{p,q}}^{m_{p,q}}}\right)^{dn/k} \left(\frac{\sum_{\substack{1\le p,q,r,s\le k\\ p\neq r,q\neq s}} m_{p,q}m_{r,s}}{dn}\right)^{dn/2}.$$

Hence, we immediately deduce that

$$\sum_{\mathcal{B}(M,n)} \prod_{\substack{1\le p,q,r,s\le k\\ p<r,q\neq s}} \frac{1}{b_{pqrs}!} = O\left(n^{k^4}\right) \left(\frac{1}{\prod_{1\le p,q\le k} {m_{p,q}}^{m_{p,q}}}\right)^{dn/k} \left(\frac{\sum_{\substack{1\le p,q,r,s\le k\\ p\neq r,q\neq s}} m_{p,q}m_{r,s}}{dn/e}\right)^{dn/2}. \quad (3.7)$$

Finally, define for any $M \in \mathcal{M}$

$$\begin{aligned}\varphi(M) &= -\frac{1}{k}\sum_{1\le p,q\le k} m_{p,q}\log m_{p,q} + \frac{d}{2}\log\left(\frac{1}{k^2}\sum_{\substack{1\le p,q,r,s\le k\\ p\neq r,q\neq s}} m_{p,q}m_{r,s}\right)\\ &= -\frac{1}{k}\sum_{1\le p,q\le k} m_{p,q}\log m_{p,q} + \frac{d}{2}\log\left(1-\frac{2}{k}+\frac{1}{k^2}\sum_{1\le p,q\le k} {m_{p,q}}^2\right). \quad (3.8)\end{aligned}$$

(Recall the convention $0^0 = 1$ and $0\log 0 = 0$.) By combining (3.6), (3.7), the Stirling formula estimate $n! \sim \sqrt{2\pi n}(n/e)^n$ and (2.2), we have

$$\begin{aligned}\frac{|T(M)|}{|\mathcal{P}_{n,d}|} &= \frac{\sqrt{2\pi n}(n/e)^n}{\sqrt{2}(dn/e)^{dn/2}} O\left(n^{k^2/2}\right) d^{dn}\left(\frac{n}{ek}\right)^{(d-1)n}\left(\prod_{1\le p,q\le k} {m_{p,q}}^{m_{p,q}}\right)^{(d-1)n/k}\\ &\quad\times O\left(n^{k^4}\right)\left(\frac{1}{\prod_{1\le p,q\le k} {m_{p,q}}^{m_{p,q}}}\right)^{dn/k}\left(\frac{\sum_{\substack{1\le p,q,r,s\le k\\ p\neq r,q\neq s}} m_{p,q}m_{r,s}}{dn/e}\right)^{dn/2}\\ &= O\left(n^{k^4+k^2/2+1/2}\right)k^n\left(\prod_{1\le p,q\le k}\frac{1}{{m_{p,q}}^{m_{p,q}}}\right)^{n/k}\left(\frac{1}{k^2}\sum_{\substack{1\le p,q,r,s\le k\\ p\neq r,q\neq s}} m_{p,q}m_{r,s}\right)^{dn/2}\\ &\le \mathrm{poly}(n)k^n e^{n\varphi(M)}, \quad (3.9)\end{aligned}$$

for some polynomial poly(n) not depending on the particular M. This is a sufficient bound on $|T(M)|/|\mathcal{P}_{n,d}|$, since it will allow us to show that the sum in (3.1) receives a negligible contribution due to terms with M away from $(1/k)J_k$.

It remains to find a suitable asymptotic expression of $|T(M)|/|\mathcal{P}_{n,d}|$ for $M \in \mathsf{B}_k(\epsilon\log n/n^{1/2})$, by improving our previous estimates in (3.6) and (3.7). Before that, we state two technical algebraic results which will be needed in the asymptotic calculations. Hereinafter, $\mathbf{1}$ represents the k-dimensional vector of ones, while $\mathbf{1}^{(i)}$ represents the k-dimensional vector with entry 1 at position i and 0 elsewhere; I_k denotes the k-by-k identity matrix; $\mathbf{vec}A$ is the vector formed by stacking the columns of a matrix A to form a single column vector; and $A^{\otimes 2}$ is simply $A \otimes A$, with the standard notation $\otimes$ for the Kronecker product of matrices.

Lemma 6 *Consider the vectors*

$$\begin{aligned} f^{(p)} &= \tfrac{\sqrt{p}}{\sqrt{p+1}}\left(\tfrac{-1}{p}\textstyle\sum_{l=1}^{p}\mathbf{1}^{(l)} + \mathbf{1}^{(p+1)}\right), \quad 1 \le p \le k-1 \\ f^{(k)} &= \tfrac{1}{\sqrt{k}}\mathbf{1}. \end{aligned}$$

Define $f^{(p,q)} = f^{(p)} \otimes f^{(q)}$ *for* $1 \le p, q \le k$.

(a) An orthonormal basis of eigenvectors for the matrix $(J_k - I_k)^{\otimes 2} + (k-1)^2 I_{k^2}$ *is given by* $\{f^{(p,q)}\}_{p,q=1}^{k}$ *with corresponding eigenvalues*

$$\begin{aligned} \lambda_{p,q} &= k^2 - 2k + 2, && 1 \le p, q \le k-1 \\ \lambda_{p,k} &= (k-1)(k-2), && 1 \le p \le k-1 \\ \lambda_{k,q} &= (k-1)(k-2), && 1 \le q \le k-1 \\ \lambda_{k,k} &= 2(k-1)^2. \end{aligned}$$

The smallest of these eigenvalues is $(k-1)(k-2)$.

(b) Similarly, the eigenvectors of $(J_k + I_k)^{\otimes 2}$ *are also* $\{f^{(p,q)}\}_{p,q=1}^{k}$, *and the corresponding eigenvalues are 1 with multiplicity* $(k-1)^2$, $k+1$ *with multiplicity* $2(k-1)$, *and* $(k+1)^2$ *with multiplicity 1.*

Proof. Immediate by checking that the eigenvectors satisfy the required properties. ∎

Lemma 7 *Let* $A = [a_{i,j}]$ *be a* k*-by-*k *matrix whose rows and columns each have sum* 0. *Define* $\tilde{A}$ *to be the submatrix formed from* A *by deleting the last row and column. Let* $\{f^{(i,j)}\}_{i,j=1}^{k}$ *be the orthonormal basis defined in the statement of Lemma 6. Then,*

(a) $(\mathbf{vec}A)^\top f^{(i,k)} = (\mathbf{vec}A)^\top f^{(k,j)} = 0$ *for* $1 \le i, j \le k$,

(b) $\displaystyle\sum_{i=1}^{k-1}\sum_{j=1}^{k-1}\left((\mathbf{vec}A)^\top f^{(i,j)}\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{k} a_{i,j}^2$, *and*

(c) $\displaystyle(\mathbf{vec}\tilde{A})^\top (J_{k-1} + I_{k-1})^{\otimes 2}\mathbf{vec}\tilde{A} = \sum_{i=1}^{k}\sum_{j=1}^{k} a_{i,j}^2$.

The proof of Lemma 7 is given later in this section. We now proceed with the asymptotic calculations for $M \in \mathcal{M} \cap (k/n)\mathbb{Z}^{k^2} \cap \mathsf{B}_k(\epsilon \log n/n^{1/2})$. So define the matrix $A = A(M) = [a_{p,q}]$ by $A = M - (1/k)J_k$. Note that each row and column of A must have sum 0, and moreover $a_{p,q} < \epsilon \log n/n^{1/2}$ for each $p, q \in \{1, \ldots, k\}$. Let $\tilde{A}$ be the submatrix formed from A by deleting the last row and column. Under the new assumptions, we first derive a new asymptotic formula for the expression computed in (3.5). For $m_{p,q} = 1/k + a_{p,q}$ with $a_{p,q} = O(\log n/n^{1/2})$ we have

$$\frac{\xi(dm_{p,q}n/k)}{\xi(m_{p,q}n/k)} \sim \frac{\sqrt{dm_{p,q}n/k}}{\sqrt{m_{p,q}n/k}} \sim \sqrt{d}$$

for $1 \le p, q \le k$, and we expand

$$\begin{aligned}\sum_{1\le p,q\le k} m_{p,q}\log m_{p,q} &= \sum_{1\le p,q\le k}\left(\frac{1}{k}+a_{p,q}\right)\left(\log\frac{1}{k}+\log\left(1+ka_{p,q}\right)\right)\\ &= \sum_{1\le p,q\le k}\left(-\frac{1}{k}\log k+\frac{k}{2}a_{p,q}^2+O(a_{p,q}^3)\right)\\ &= k\left(-\log k+\frac{1}{2}(\mathbf{vec}\tilde{A})^\top(J_{k-1}+I_{k-1})^{\otimes 2}\mathbf{vec}\tilde{A}\right)+O(\log^3 n/n^{3/2})\end{aligned}$$

where we used Lemma 7(c) to rewrite $\sum_{p,q} a_{p,q}^2$ in the final step. Combining these estimates we can rewrite (3.5) as

$$\prod_{1\le p,q\le k}\frac{(dm_{p,q}n/k)!}{(m_{p,q}n/k)!}\sim d^{k^2/2}d^{dn}\left(\frac{n}{ek^2}\right)^{(d-1)n}\exp\left(n\frac{d-1}{2}(\mathbf{vec}\tilde{A})^\top(J_{k-1}+I_{k-1})^{\otimes 2}\mathbf{vec}\tilde{A}\right). \tag{3.10}$$

Next we rewrite the inner sum in (3.4) in terms of the natural generating function, letting square brackets denote the extraction of a coefficient

$$\begin{aligned}\sum_{\mathcal{B}(M,n)}\prod_{\substack{1\le p,q,r,s\le k\\ p<r,q\ne s}}\frac{1}{b_{pqrs}!} &= \Big[\prod_{1\le p,q\le k} {x_{p,q}}^{dm_{p,q}n/k}\Big]\prod_{\substack{1\le p,q,r,s\le k\\ p<r,q\ne s}}\sum_{i=0}^{\infty}\frac{(x_{p,q}x_{r,s})^i}{i!}\\ &= \Big[\prod_{1\le p,q\le k} {x_{p,q}}^{dm_{p,q}n/k}\Big]\exp\left(\frac{1}{2}\sum_{\substack{1\le p,q,r,s\le k\\ p\ne r,q\ne s}} x_{p,q}x_{r,s}\right).\end{aligned}$$

The following result provides an asymptotic characterisation of that coefficient. The proof uses the saddlepoint method, and can be found at the end of the section. Note that our still unspecified ϵ is determined by the statement.

Lemma 8 *There exists $\epsilon > 0$ such that for each $M \in \mathcal{M} \cap (k/n)\mathbb{Z}^{k^2} \cap \mathsf{B}_k(\epsilon \log n/n^{1/2})$ the coefficient C of $\left[\prod_{1\le p,q\le k} {x_{p,q}}^{dm_{p,q}n/k}\right]$ in the generating function $\exp\left(\frac{1}{2}\sum_{\substack{1\le p,q,r,s\le k\\ p\ne r,q\ne s}} x_{p,q}x_{r,s}\right)$ satisfies*

$$C\sim\frac{\gamma(k)}{\sqrt{\pi}}\left(\frac{1}{dn}\right)^{k^2/2}\left(\frac{k(k-1)}{\sqrt{dn/e}}\right)^{dn}\exp\left(\frac{-nd(k-1)^2(\mathbf{vec}\tilde{A})^\top(J_{k-1}+I_{k-1})^{\otimes 2}\mathbf{vec}\tilde{A}}{2(k^2-2k+2)}\right),$$

where $\gamma(k)$ is the constant $\gamma(k)=\dfrac{k^{k^2}(k-1)^{k(k-1)}}{(2\pi)^{k^2/2-1/2}(k^2-2k+2)^{(k-1)^2/2}(k-2)^{k-1}}$.

Hence, in view of (3.10) and Lemma 8, we obtain for $M \in \mathcal{M} \cap (k/n)\mathbb{Z}^{k^2} \cap \mathsf{B}_k(\epsilon \log n/n^{1/2})$ an asymptotic expression for (3.4) which improves the bound already stated in (3.9)

$$\frac{|T(M)|}{|\mathcal{P}_{n,d}|}\sim\gamma(k)\frac{(k-1)^{dn}}{n^{k^2/2-1/2}k^{(d-2)n}}\exp\left(-n\frac{k^2-2k-d+2}{2(k^2-2k+2)}(\mathbf{vec}\tilde{A})^\top(J_{k-1}+I_{k-1})^{\otimes 2}\mathbf{vec}\tilde{A}\right), \tag{3.11}$$

where $\gamma(k)$ is the constant defined in Lemma 8.

We are now in good shape to estimate the sum (3.1). We begin by computing the contribution of the terms near $(1/k)J_k$. For a $(k-1)$-by-$(k-1)$ matrix M define $\overline{M}$ to be the k-by-k matrix formed from M by adding a new row and column so that every row sum and column sum is 1. Recall the definition of $\mathsf{B}_p(\delta)$ and define $\overline{\mathsf{B}_p(\delta)} = \{\overline{M} \mid M \in \mathsf{B}_p(\delta)\}$. Now we set

$$\delta = \frac{\epsilon \log n}{(k-1)^2 n^{1/2}},$$

and consider $M = \overline{M'}$ for $M' \in \mathsf{B}_{k-1}(\delta) \cap \frac{k}{n}\mathbb{Z}^{(k-1)^2}$. A straightforward application of the triangle inequality shows that $\overline{\mathsf{B}_{k-1}(\delta)} \subseteq \mathsf{B}_k((k-1)^2\delta)$, and therefore $M \in \mathsf{B}_k(\epsilon \log n/n^{1/2})$. So M is nonnegative (for large enough n) and hence $M \in \mathcal{M}$. Furthermore, the entries of M are in $\frac{k}{n}\mathbb{Z}$ because they are integer linear combinations of k/n and $1 = \frac{k}{n} \times \frac{n}{k}$, using the fact that k divides n. This shows that

$$\frac{|T(M)|}{|\mathcal{P}_{n,d}|}$$

is a term in the sum (3.1), suggesting that we express (3.1) as $\mathbf{E}(Y^2) = S_1 + S_2$ where

$$S_1 = \sum_{M' \in \mathsf{B}_{k-1}(\delta) \cap \frac{k}{n}\mathbb{Z}^{(k-1)^2}} \frac{|T(\overline{M'})|}{|\mathcal{P}_{n,d}|}$$

and S_2 is the sum of the remaining terms. Notice moreover that if $M' \in \mathsf{B}_{k-1}(\delta) \cap \frac{k}{n}\mathbb{Z}^{(k-1)^2}$ and $M = \overline{M'}$, then the matrix $A = M - (1/k)J_k$ has all entries $a_{p,q} < \epsilon n^{-1/2} \log n$. Hence the expansion given in (3.11) is valid, and we can express

$$S_1 \sim \gamma(k) \frac{(k-1)^{dn}}{n^{k^2/2-1/2} k^{(d-2)n}} \sum_{M' \in \mathsf{B}_{k-1}(\delta) \cap \frac{k}{n}\mathbb{Z}^{(k-1)^2}} e^{nf(M')}, \tag{3.12}$$

where

$$f(M') = -\frac{k^2-2k-d+2}{2(k^2-2k+2)} (\mathbf{vec}\tilde{A})^\top (J_{k-1} + I_{k-1})^{\otimes 2} \mathbf{vec}\tilde{A},$$

and $\tilde{A} = M' - (1/k)J_{k-1}$. By iterating the Euler–Maclaurin summation formula (see [1], p. 806), we have

$$\sum_{M' \in \mathsf{B}_{k-1}(\delta) \cap \frac{k}{n}\mathbb{Z}^{(k-1)^2}} e^{nf(M')} \sim \left(\frac{n}{k}\right)^{(k-1)^2} \int_{M' \in \mathsf{B}_{k-1}(\delta)} e^{nf(M')} dM'.$$

Setting $H = \frac{k^2-2k-d+2}{2(k^2-2k+2)}(J_{k-1} + I_{k-1})^{\otimes 2}$ and using Lemma 6 (b) for the eigenvalues of $(J_{k-1} + I_{k-1})^{\otimes 2}$, we deduce that H is positive definite and has determinant

$$k^{2k-2} \left(\frac{k^2-2k-d+2}{k^2-2k+2}\right)^{(k-1)^2}. \tag{3.13}$$

Here we used the fact that $k^2-2k-d+2>0$, which is guaranteed from our assumptions on d and k. Now we may apply Lemma 4 to conclude

$$\begin{aligned} S_1 &\sim \gamma(k)\frac{(k-1)^{dn}}{n^{k^2/2-1/2}k^{(d-2)n}}\left(\frac{n}{k}\right)^{(k-1)^2}\left(\int_{[-\infty,\infty]^{(k-1)^2}} e^{nf(M')}dM' + O\big(e^{-c\log^2 n}\big)\right) \\ &\sim \gamma(k)\frac{(k-1)^{dn}}{n^{k^2/2-1/2}k^{(d-2)n}}\left(\frac{n}{k}\right)^{(k-1)^2}\frac{(2\pi/n)^{(k-1)^2/2}}{|\det H|^{1/2}} \\ &= \frac{k^k(k-1)^{k(k-1)}}{(k^2-2k-d+2)^{(k-1)^2/2}(2\pi(k-2))^{k-1}}n^{-(k-1)}k^{2n}\left(1-\frac{1}{k}\right)^{dn}. \end{aligned} \tag{3.14}$$

To prove the proposition it suffices to show $S_2=o(S_1)$. Let M be an index of any term of S_2. This implies $M\notin\overline{\mathsf{B}_{k-1}(\delta)}$, so we must have $M\in\mathcal{M}\setminus\mathsf{B}_k(\delta)$ since $\mathsf{B}_k(\delta)\cap\mathcal{M}\subseteq\overline{\mathsf{B}_{k-1}(\delta)}$. Recall now the definition of φ in (3.8), and obtain by direct substitution

$$\varphi\left(\frac{1}{k}J_k\right)=\log k+d\log\left(1-\frac{1}{k}\right).$$

From Theorem 7 in [3] (see also (5) in the same paper) we have that if $d<d_{k-1}:=2(k-1)\log(k-1)$ for each $M\in\mathcal{M}$,

$$\begin{aligned} \varphi(M) &\le \varphi\left(\frac{1}{k}J_k\right)-\frac{d_{k-1}-d}{4(k-1)^2}\left(\sum_{p,q}{m_{p,q}}^2-1\right) \\ &= \log k+d\log\left(1-\frac{1}{k}\right)-\frac{d_{k-1}-d}{4(k-1)^2}\sum_{p,q}\left(m_{p,q}-\frac{1}{k}\right)^2. \end{aligned}$$

In particular, for each $M\in\mathcal{M}\setminus\mathsf{B}_k(\delta)$,

$$\varphi(M)\le\log k+d\log\left(1-\frac{1}{k}\right)-\frac{d_{k-1}-d}{4(k-1)^2}\delta^2. \tag{3.15}$$

By combining (3.15) with the bound on the general term obtained in (3.9) and also taking into account that the number of terms in S_2 is at most $O(n^{k^2})$, we conclude

$$\begin{aligned} S_2 &= O(n^{k^2})\,\mathrm{poly}(n)k^{2n}\left(1-\frac{1}{k}\right)^{dn}\exp\left(-\frac{d_{k-1}-d}{4(k-1)^2}\epsilon\log^2 n\right) \\ &\le \mathrm{poly}'(n)k^{2n}\left(1-\frac{1}{k}\right)^{dn}n^{-\Theta(\log n)}, \end{aligned}$$

for some polynomial $\mathrm{poly}'(n)$. Thus $S_2=o(S_1)$, and this completes the proof. ∎

It only remains to prove Lemmas 5, 7 and 8.

Proof of Lemma 5. Let $\mathcal{L}=\mathcal{L}(M)$ be the polytope consisting of all non-negative tuples $L=(\ell_{pqrs})_{\substack{1\le p,q,r,s\le k\\ p<r,q\neq s}}$ in $\mathbb{R}^{k^2(k-1)^2/2}$ such that

$$\sum_{\substack{1\le r,s\le k\\ r\neq p,s\neq q}}\ell_{pqrs}=dm_{p,q}\qquad\forall p,q\in\{1,\dots,k\}, \tag{3.16}$$

where for $p > r$ and $q \neq s$ we used the duplicate notation $\ell_{pqrs} = \ell_{rspq}$ to denote the coordinates. For each $L \in \mathcal{L}$, define

$$\psi(L) = \prod_{1 \leq p,q \leq k} \frac{{m_{p,q}}^{dm_{p,q}}}{\prod_{\substack{1 \leq r,s \leq k \\ p<r, q \neq s}} {\ell_{pqrs}}^{\ell_{pqrs}}}.$$

Our aim is to show that for all $L \in \mathcal{L}$

$$\psi(L) \leq \left(\frac{\sum_{\substack{1 \leq p,q,r,s \leq k \\ p \neq r, q \neq s}} m_{p,q} m_{r,s}}{dk} \right)^{dk/2}. \tag{3.17}$$

This is indeed equivalent to the statement of the lemma, after setting $b_{pqrs} = \ell_{pqrs} n/k$ and performing straightforward manipulations.

As a first case, assume that $m_{pq} > 0$ for all p, q in $\{1, \ldots, k\}$. We define $\widehat{\mathcal{L}}$ to be the polytope of all non-negative tuples $L = (\ell_{pqrs})_{\substack{1 \leq p,q,r,s \leq k \\ p<r, q \neq s}}$ in $\mathbb{R}^{k^2(k-1)^2/2}$ such that

$$\sum_{\substack{1 \leq p,q,r,s \leq k \\ p<r, q \neq s}} \ell_{pqrs} = dk/2, \tag{3.18}$$

and for each $L \in \widehat{\mathcal{L}}$, let

$$\widehat{\psi}(L) = \prod_{\substack{1 \leq p,q,r,s \leq k \\ p<r, q \neq s}} \left(\frac{m_{p,q} m_{r,s}}{\ell_{pqrs}} \right)^{\ell_{pqrs}}.$$

Observe that $\mathcal{L} \subset \widehat{\mathcal{L}}$, and moreover the restriction of $\widehat{\psi}$ to $\mathcal{L}$ is equal to ψ. Our goal is to maximise $\log \widehat{\psi}$ over $\widehat{\mathcal{L}}$, and thus provide a bound on ψ over $\mathcal{L}$. We first show that $\log \widehat{\psi}$ does not maximise on the boundary. Note that the boundary of $\widehat{\mathcal{L}}$ consists of points having some 0 coordinate (but at least some coordinate must be strictly positive). Let us choose an arbitrary boundary point L_0, and assume without loss of generality that $\ell_{1122} = 0$ and $\ell_{1221} > 0$. Let L_ϵ be the point with the same coordinates as L_0 but replacing ℓ_{1122} by ϵ and ℓ_{1221} by $\ell_{1221} - \epsilon$. For small enough $\epsilon > 0$, L_ϵ lies in $\widehat{\mathcal{L}}$, and moreover

$$\lim_{\epsilon \to 0^+} \frac{d}{d\epsilon} \log \widehat{\psi}(L_\epsilon) = +\infty,$$

so the maximum cannot occur at L_0. Hence, $\log \widehat{\psi}$ is maximised in the interior of $\widehat{\mathcal{L}}$ and the maximiser(s) must satisfy the following Lagrange equations

$$\log(m_{p,q} m_{r,s}) - \log \ell_{pqrs} - 1 = \lambda \qquad \forall p, q, r, s \in \{1, \ldots k\}, p < r, q \neq s.$$

These are equivalent to

$$\ell_{pqrs} = m_{p,q} m_{r,s} e^{-\lambda - 1} \qquad \forall p, q, r, s \in \{1, \ldots k\}, p < r, q \neq s,$$

which combined with (3.18) yield

$$\ell_{pqrs} = dk \frac{m_{p,q} m_{r,s}}{\sum_{\substack{1 \leq p', q', r', s' \leq k \\ p' \neq r', q' \neq s'}} m_{p',q'} m_{r',s'}} \qquad \forall p, q, r, s \in \{1, \ldots k\}, p < r, q \neq s.$$

From the uniqueness of the solution, we deduce that it must be the maximiser of $\log \widehat{\psi}$ (and $\widehat{\psi}$). The value of $\widehat{\psi}$ at this point can be easily computed by substitution

$$\left(\frac{\sum_{\substack{1\le p,q,r,s\le k\\ p\neq r, q\neq s}} m_{p,q}m_{r,s}}{dk}\right)^{dk/2},$$

which proves the bound in (3.17) under the assumption that $m_{pq} > 0$ for all p and q. To extend the argument to the other cases, we first define

$$\mathcal{P} = \{(M, L) \mid M \in \mathcal{M},\, L \in \mathcal{L}(M)\},$$

and with a mild abuse of notation denote by

$$\psi(M, L) = \prod_{1\le p,q\le k} \frac{{m_{p,q}}^{dm_{p,q}}}{\prod_{\substack{1\le r,s\le k\\ p<r, q\neq s}} {\ell_{pqrs}}^{\ell_{pqrs}}},$$

the natural extension of ψ to $\mathcal{P}$. Notice that ψ is continuous in $\mathcal{P}$. We showed so far that, for any M in the interior of $\mathcal{M}$ (i.e. $m_{pq} > 0$) and any $L \in \mathcal{L}(M)$,

$$\psi(M, L) \le \left(\frac{\sum_{\substack{1\le p,q,r,s\le k\\ p\neq r, q\neq s}} m_{p,q}m_{r,s}}{dk}\right)^{dk/2}.$$

Hence, this inequality can be extended by continuity to any M on the boundary of $\mathcal{M}$, and thus to any $(M, L) \in \mathcal{P}$. This concludes the proof of (3.17). ∎

Proof of Lemma 7. We begin by proving (a). Let $j \in \{1, 2, \ldots, k\}$. Since $f^{(k,j)} = f^{(k)} \otimes f^{(j)}$ is a linear combination of terms of the form $\mathbf{1} \otimes \mathbf{1}^{(q)} = \sum_{p=1}^{k}(\mathbf{1}^{(p)} \otimes \mathbf{1}^{(q)}), (1 \le q \le k)$, we have that $(\mathbf{vec}A)^\top f^{(k,j)}$ is a linear combination of terms of the form $\sum_{p=1}^{k}(\mathbf{vec}A)^\top(\mathbf{1}^{(p)} \otimes \mathbf{1}^{(q)}) = \sum_{p=1}^{k} a_{q,p} = 0$ since the row sums of A are 0. A similar argument shows $(\mathbf{vec}A)^\top f^{(i,k)} = 0$ for $i \in \{1, 2, \ldots, k\}$ using the fact that the column sums of A equal 0.

To prove (b) we apply (a) to write

$$\sum_{i=1}^{k-1}\sum_{j=1}^{k-1}\left((\mathbf{vec}A)^\top f^{(i,j)}\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{k}\left((\mathbf{vec}A)^\top f^{(i,j)}\right)^2,$$

which is the sum of the squares of the coordinates of $\mathbf{vec}A$ in the basis given by $\{f^{(i,j)}\}_{i,j=1}^{k}$. Since the basis is orthonormal, this expression is simply the square of the norm of $\mathbf{vec}A$ with respect to the standard basis, $\sum_{i=1}^{k}\sum_{j=1}^{k} a_{i,j}^2$.

To prove part (c) we begin by writing

$$\begin{aligned}
&\sum_{i=1}^{k}\sum_{j=1}^{k} a_{i,j}^2\\
=\;& a_{k,k}^2 + \sum_{i=1}^{k-1} a_{i,k}^2 + \sum_{j=1}^{k-1} a_{k,j}^2 + \sum_{i=1}^{k-1}\sum_{j=1}^{k-1} a_{i,j}^2\\
=\;& \left(\sum_{i=1}^{k-1}\sum_{j=1}^{k-1} a_{i,j}\right)^2 + \sum_{i=1}^{k-1}\left(-\sum_{j=1}^{k-1} a_{i,j}\right)^2 + \sum_{j=1}^{k-1}\left(-\sum_{i=1}^{k-1} a_{i,j}\right)^2 + \sum_{i=1}^{k-1}\sum_{j=1}^{k-1} a_{i,j}^2.
\end{aligned}$$

Since

$$\left(\sum_{i=1}^{k-1}\sum_{j=1}^{k-1} a_{i,j}\right)^2 = (\mathbf{vec}\tilde{A})^\top J_{k-1}^{\otimes 2}\mathbf{vec}\tilde{A},$$

$$\sum_{i=1}^{k-1}\left(\sum_{j=1}^{k-1} a_{i,j}\right)^2 = (\mathbf{vec}\tilde{A})^\top (J_{k-1}\otimes I_{k-1})\mathbf{vec}\tilde{A},$$

$$\sum_{j=1}^{k-1}\left(\sum_{i=1}^{k-1} a_{i,j}\right)^2 = (\mathbf{vec}\tilde{A})^\top (I_{k-1}\otimes J_{k-1})\mathbf{vec}\tilde{A},$$

$$\sum_{i=1}^{k-1}\sum_{j=1}^{k-1} a_{i,j}^2 = (\mathbf{vec}\tilde{A})^\top I_{k-1}^{\otimes 2}\mathbf{vec}\tilde{A},$$

and $(J_{k-1}+I_{k-1})^{\otimes 2} = J_{k-1}^{\otimes 2} + (J_{k-1}\otimes I_{k-1}) + (I_{k-1}\otimes J_{k-1}) + I_{k-1}^{\otimes 2}$, part (c) is proved. ∎

Before proceeding with the proof of Lemma 8, we need the following technical result.

Lemma 9 *Let $\delta \in (0, 2\pi/5)$ and fix an integer $k \geq 3$. For each $1 \leq p, q \leq k$, let $-\pi \leq \theta_{p,q} \leq \pi$. Suppose $\max_{p,q}|\theta_{p,q}| > \delta$ and $\min_{p,q}|\theta_{p,q}| < \pi - \delta$. Then there exist p, q, r, and s with $p \neq r$ and $q \neq s$ such that*

$$\frac{\delta}{2} \leq |\theta_{p,q} + \theta_{r,s}| \leq 2\pi - \frac{\delta}{2}.$$

Proof. There are two cases. In the first case, suppose $\delta < |\theta_{p,q}| < \pi - \delta$ for some p and q. Let S be the set of pairs

$$S = \{(r,s) \mid 1 \leq r \leq k, r \neq p, 1 \leq s \leq k, s \neq q\}.$$

The set S is nonempty as $k \geq 2$. If there exists $(r,s) \in S$ with $|\theta_{p,q} + \theta_{r,s}| > \delta/2$ then

$$\begin{aligned} \frac{\delta}{2} < |\theta_{p,q} + \theta_{r,s}| &\leq |\theta_{p,q}| + |\theta_{r,s}| \\ &< \pi - \delta + \pi \\ &< 2\pi - \frac{\delta}{2} \end{aligned}$$

and we are finished. Otherwise, all $\theta_{r,s}$ with $(r,s) \in S$ satisfy $|\theta_{p,q} + \theta_{r,s}| \leq \delta/2$; i.e. they are all within $\delta/2$ units of $-\theta_{p,q}$, and so, because $\delta < |\theta_{p,q}| < \pi - \delta$, they all have the same sign and satisfy $\delta/2 \leq |\theta_{r,s}| \leq \pi - \delta/2$. Now let $(r,s) \in S$ and choose any (t,u) with $t \in \{1,2,\ldots,k\} \setminus \{p,r\}$ and $u \in \{1,2,\ldots,k\} \setminus \{q,s\}$. This is possible because $k \geq 3$. Since $(t,u) \in S$ we have, using the above observations, $\delta < |\theta_{r,s} + \theta_{t,u}| < 2\pi - \delta$, which implies the required result.

For the remaining case, we must have $|\theta_{p,q}| \in [0,\delta] \cup [\pi - \delta, \pi]$ for all $1 \leq p, q \leq k$. We claim there exist p, q, r, s with $p \neq r$, $q \neq s$, $|\theta_{p,q}| \in [0,\delta]$, and $|\theta_{r,s}| \in [\pi - \delta, \pi]$. If we prove the claim then we are finished because

$$\frac{\delta}{2} < \pi - 2\delta \leq ||\theta_{p,q}| - |\theta_{r,s}|| \leq |\theta_{p,q} + \theta_{r,s}| \leq |\theta_{p,q}| + |\theta_{r,s}| \leq \delta + \pi < 2\pi - \frac{\delta}{2}.$$

Assume for contradiction that the claim is false. By the hypothesis of the proposition there exist p and q with $|\theta_{p,q}| < \pi - \delta$. Since $|\theta_{p,q}| \in [0,\delta] \cup [\pi-\delta,\pi]$ for every $1 \le p,q \le k$ we must have $|\theta_{p,q}| \in [0,\delta]$. Since we are assuming that the claim is false, we must have $|\theta_{r,s}| \in [0,\delta]$ for the $(k-1)^2$ pairs (r,s) with $r \neq p$ and $s \neq q$. But the hypothesis of the proposition also gives us (t,u) with $|\theta_{t,u}| > \delta$, so an argument analogous to the previous one shows there must exist $(k-1)^2$ pairs (v,w) with $|\theta_{v,w}| \in [\pi-\delta,\pi]$. Since $(k-1)^2 + (k-1)^2$ exceeds k^2, the total number of ordered pairs, we have a contradiction, as required. ∎

Proof of Lemma 8. We use the saddlepoint method to estimate the coefficient C in the statement. Using Cauchy's integral formula, C can be written in terms of an integral around the product of circles $z_{p,q} = \rho_{p,q}\exp(i\theta_{p,q})$, $-\pi \le \theta_{p,q} \le \pi$, $(1 \le p,q \le k)$, as follows,

$$\begin{aligned} C &= \frac{1}{(2\pi i)^{k^2}} \int \frac{\exp\left(\frac{1}{2}\sum_{\substack{p\neq r\\ q\neq s}} z_{p,q}z_{r,s}\right)}{\prod_{p,q} z_{p,q}^{dm_{p,q}n/k+1}} \prod_{p,q} dz_{p,q} \\ &= \frac{1}{(2\pi)^{k^2}\prod_{p,q}\rho_{p,q}^{dm_{p,q}n/k}} \\ &\quad \times \int_{\theta\in[-\pi,\pi]^{k^2}} \frac{\exp\left(\frac{1}{2}\sum_{\substack{p\neq r\\ q\neq s}} \rho_{p,q}\rho_{r,s}e^{i(\theta_{p,q}+\theta_{r,s})}\right)}{\exp(i\sum_{p,q}\theta_{p,q}dm_{p,q}n/k)} \prod_{p,q} d\theta_{p,q}. \end{aligned} \tag{3.19}$$

Viewing $\theta = \mathbf{vec}([\theta_{p,q}])$ as a k^2-dimensional vector, let $g(\theta)$ denote the integrand in the above expression. Consider

$$\begin{aligned} g(\theta+\pi\mathbf{1}) &= \frac{\exp\left(\frac{1}{2}\sum_{\substack{p\neq r\\ q\neq s}} \rho_{p,q}\rho_{r,s}e^{i(\theta_{p,q}+\theta_{r,s}+2\pi)}\right)}{\exp(i\sum_{p,q}\theta_{p,q}dm_{p,q}n/k + i\pi\sum_{p,q}dm_{p,q}n/k)} \\ &= \frac{\exp\left(\frac{1}{2}\sum_{\substack{p\neq r\\ q\neq s}} \rho_{p,q}\rho_{r,s}e^{i(\theta_{p,q}+\theta_{r,s})}\right)}{\exp(i\sum_{p,q}\theta_{p,q}dm_{p,q}n/k + i\pi dn)} \\ &= g(\theta), \end{aligned}$$

which holds since $\sum_{p,q} m_{p,q} = k$ and dn is even. Setting $\delta = \log n/\sqrt{n}$, this tells us that the integral over the region $\{\theta \mid |\theta_{p,q}| \le \delta \text{ for } 1 \le p,q \le k\}$ equals the integral over the region $\{\theta \mid \pi-\delta \le |\theta_{p,q}| \le \pi \text{ for } 1 \le p,q \le k\}$. Set each $\rho_{p,q}$ to be the common value

$$\rho_{p,q} = \rho = \frac{\sqrt{dn}}{k(k-1)}.$$

We will see that the integral over each of these regions is asymptotic to

$$\begin{aligned} I &= \frac{e^{dn/2}}{\sqrt{2}}\left(\frac{2\pi}{dn}\right)^{k^2/2} \frac{k^{k^2}(k-1)^{k(k-1)}\exp\left(\frac{-nd(k-1)^2(\mathbf{vec}\tilde{A})^\top(J_{k-1}+I_{k-1})^{\otimes 2}\mathbf{vec}\tilde{A}}{2(k^2-2k+2)}\right)}{(k^2-2k+2)^{(k-1)^2/2}(k-2)^{k-1}} \\ &\ge K\exp\left(\frac{dn}{2} - \epsilon'\log^2 n\right) \end{aligned} \tag{3.20}$$

(using $a_{p,q} < \epsilon \log n/n^{1/2}$) where K and ϵ' are constants depending on ϵ. We will also show that the integral over the remaining region is $o(I)$. The lemma then follows by combining these two facts with (3.19).

To see that the integral over the remaining region is $o(I)$, let θ be any vector in this region. By the definition of this region we must have $\min_{p,q} |\theta_{p,q}| < \pi - \delta$ and $\max_{p,q} |\theta_{p,q}| > \delta$. By Lemma 9 there exist $p^*, q^*, r^*, s^* \in \{1, \ldots, k\}$ with $p^* \neq r^*$ and $q^* \neq s^*$ such that

$$\frac{\delta}{2} \leq |\theta_{p^*,q^*} + \theta_{r^*,s^*}| \leq 2\pi - \frac{\delta}{2}.$$

Now

$$\cos(\theta_{p^*,q^*} + \theta_{r^*,s^*}) < \cos\left(\frac{\delta}{2}\right) = 1 - \frac{\delta^2}{8} + O(\delta^3)$$

so the absolute value of the integrand is

$$\begin{aligned}
|g(\theta)| &= \frac{\left|\exp\left(\frac{1}{2}\sum_{\substack{p\neq r\\ q\neq s}} \rho^2 e^{i(\theta_{p,q}+\theta_{r,s})}\right)\right|}{\left|\exp(i\sum_{p,q}\theta_{p,q} d m_{p,q} n/k)\right|} \\
&= \exp\left(\frac{1}{2}\sum_{\substack{p\neq r\\ q\neq s}} \rho^2 \cos(\theta_{p,q}+\theta_{r,s})\right) \\
&\leq \exp\left(\frac{1}{2}\rho^2\left((k^2(k-1)^2-1)1 + \cos(\theta_{p^*,q^*}+\theta_{r^*,s^*})\right)\right) \\
&= \exp\left(\frac{1}{2}\rho^2\left((k^2(k-1)^2-1)1 + 1 - \frac{\delta^2}{8} + O(\delta^3)\right)\right) \\
&= \exp\left(\frac{1}{2}\rho^2 k^2(k-1)^2 - \rho^2\frac{(\log n)^2}{16n} + O(\rho^2 n^{-3/2}(\log n)^3)\right) \\
&= \exp\left(\frac{dn}{2} - \frac{d(\log n)^2}{16k^2(k-1)^2} + o(1)\right)
\end{aligned}$$

recalling that we chose $\rho = k^{-1}(k-1)^{-1}\sqrt{dn}$ and $\delta = \log n/\sqrt{n}$. Hence $|g(\theta)| = o(I)$ if we choose ϵ sufficiently small so that the constant ϵ' in (3.20) satisfies $\epsilon' < d/\left(16k^2(k-1)^2\right)$.

It remains to show that

$$\int_{\theta\in[-\delta,\delta]^{k^2}} g(\theta)d\theta \sim I.$$

For $\theta \in [-\delta, \delta]^{k^2}$ we have

$$\log g(\theta) = \rho^2\frac{1}{2}\sum_{\substack{p\neq r\\ q\neq s}}\left(1 + i(\theta_{p,q}+\theta_{r,s}) - \frac{(\theta_{p,q}+\theta_{r,s})^2}{2} + O(|\theta|^3)\right) - i\frac{dn}{k}\sum_{p,q}\theta_{p,q}m_{p,q}.$$

Regrouping the terms and substituting $m_{p,q} = k^{-1} + a_{p,q}$ this becomes

$$\begin{aligned}
\log g(\theta) = & \quad \frac{\rho^2}{2} k^2 (k-1)^2 \\
& + \sum_{p,q} \theta_{p,q} \left(2i(k-1)^2 \frac{\rho^2}{2} - i \frac{dn}{k} \left(\frac{1}{k} + a_{p,q} \right) \right) \\
& - \frac{\rho^2}{2} \left((k-1)^2 \sum_{p,q} \theta_{p,q}^2 + \sum_{\substack{p \neq r \\ q \neq s}} \theta_{p,q} \theta_{r,s} \right) \\
& + O(\rho^2 |\theta|^3)
\end{aligned}$$

Let c be the constant $c = d/(2k^2(k-1)^2)$. Recalling $\rho = k^{-1}(k-1)^{-1}\sqrt{dn}$ we find

$$\log g(\theta) = \frac{dn}{2} - i \frac{dn}{k} (\mathbf{vec} A)^\top \theta - cn\theta^\top B \theta + O(n^{-1/2} (\log n)^3) \tag{3.21}$$

where B is the matrix

$$B = (k-1)^2 I_{k^2} + (J_k - I_k)^{\otimes 2}.$$

Define $h(\theta) = -i(dn/k)(\mathbf{vec} A)^\top \theta - cn\theta^\top B\theta$. Lemma 6 gives us an orthonormal basis $\{f^{(p,q)}\}_{p,q=1}^k$ of eigenvectors for B and corresponding sequence of eigenvalues $(\lambda_{p,q})_{p,q=1}^k$. Introduce the new variables $(\tau_{p,q})_{p,q=1}^k$ to perform the change of basis $\theta = \sum_{p,q} f^{(p,q)} \tau_{p,q}$. This gives

$$\begin{aligned}
h(\theta) &= -i \frac{dn}{k} (\mathbf{vec} A)^\top \sum_{p,q} f^{(p,q)} \tau_{p,q} - cn \sum_{p,q} \lambda_{p,q} \tau_{p,q}^2 \\
&= \sum_{p,q} \left(-i(dn/k)(\mathbf{vec} A)^\top f^{(p,q)} \tau_{p,q} - cn\lambda_{p,q} \tau_{p,q}^2 \right).
\end{aligned}$$

Let $p, q \in \{1, \ldots, k\}$. Using the identity

$$\int_{-\infty}^{\infty} e^{ax - bx^2} dx = \sqrt{\frac{\pi}{b}} \exp\left(\frac{a^2}{4b} \right)$$

(for $b > 0$), we have that $\int_{[-\infty,\infty]^{k^2}} \exp(h(\theta)) d\theta$ is a product of terms of the form

$$\begin{aligned}
& \int_{-\infty}^{\infty} \exp\left(-i \frac{dn}{k} (\mathbf{vec} A)^\top f^{(p,q)} \tau_{p,q} - cn\lambda_{p,q} \tau_{p,q}^2 \right) d\tau_{p,q} \\
= & \sqrt{\frac{\pi}{cn\lambda_{p,q}}} \exp\left(\frac{-d^2 n ((\mathbf{vec} A)^\top f^{(p,q)})^2}{4ck^2 \lambda_{p,q}} \right).
\end{aligned}$$

So by Lemma 4, for some constant $c' > 0$ we have

$$\begin{aligned}
\int_{[-\delta,\delta]^{k^2}} \exp(h(\theta))d\theta &= \int_{[-\infty,\infty]^{k^2}} \exp(h(\theta))d\theta + O(e^{-c'(\log n)^2}) \\
&= \prod_{p,q} \sqrt{\frac{\pi}{cn\lambda_{p,q}}} \exp\left(\frac{-d^2 n((\mathbf{vec}A)^\top f^{(p,q)})^2}{4ck^2\lambda_{p,q}}\right) + O(e^{-c'(\log n)^2}) \\
&\sim \prod_{p,q} \sqrt{\frac{\pi}{cn\lambda_{p,q}}} \exp\left(\frac{-d^2 n((\mathbf{vec}A)^\top f^{(p,q)})^2}{4ck^2\lambda_{p,q}}\right)
\end{aligned}$$

since the entries of A are less than $\epsilon n^{-1/2}\log n$, c' does not depend on ϵ and we will choose ϵ to be sufficiently small. Recalling (3.21) we now have

$$\int_{[-\delta,\delta]^{k^2}} g(\theta)d\theta \sim e^{dn/2} \prod_{p,q} \sqrt{\frac{\pi}{cn\lambda_{p,q}}} \exp\left(\frac{-d^2 n((\mathbf{vec}A)^\top f^{(p,q)})^2}{4ck^2\lambda_{p,q}}\right) \tag{3.22}$$

We will simplify the above product using the values of $\lambda_{p,q}$ given in Lemma 6. First, the contribution to the product from $1 \le p, q \le k-1$ is

$$\begin{aligned}
&\prod_{p=1}^{k-1}\prod_{q=1}^{k-1} \sqrt{\frac{\pi}{cn\lambda_{p,q}}} \exp\left(\frac{-d^2 n((\mathbf{vec}A)^\top f^{(p,q)})^2}{4ck^2\lambda_{p,q}}\right) \\
\sim\ & \left(\sqrt{\frac{\pi}{cn(k^2-2k+2)}}\right)^{(k-1)^2} \prod_{p=1}^{k-1}\prod_{q=1}^{k-1} \exp\left(\frac{-d^2 n((\mathbf{vec}A)^\top f^{(p,q)})^2}{4ck^2(k^2-2k+2)}\right) \\
=\ & \left(\sqrt{\frac{\pi}{cn(k^2-2k+2)}}\right)^{(k-1)^2} \exp\left(\frac{-d^2 n(\sum_{p=1}^{k-1}\sum_{q=1}^{k-1}(\mathbf{vec}A)^\top f^{(p,q)})^2}{4ck^2(k^2-2k+2)}\right) \\
=\ & \left(\sqrt{\frac{\pi}{cn(k^2-2k+2)}}\right)^{(k-1)^2} \exp\left(\frac{-d^2 n(\mathbf{vec}\tilde{A})^\top (J_{k-1}+I_{k-1})^{\otimes 2}\mathbf{vec}\tilde{A}}{4ck^2(k^2-2k+2)}\right)
\end{aligned}$$

where the last step used Lemmata 7(b) and 7(c). The contribution to the product when exactly one of p or q equals k is

$$\left(\sqrt{\frac{\pi}{cn(k-1)(k-2)}}\right)^{2(k-1)}$$

since Lemma 7(a) tells us that $((\mathbf{vec}A)^\top f^{(p,q)})^2 = 0$ when $p = k$ or $q = k$. When $p = q = k$ the contribution to the product is

$$\sqrt{\frac{\pi}{2cn(k-1)^2}}$$

Substituting these contributions into (3.22) we get

$$\begin{aligned}
&\int_{[-\delta,\delta]^{k^2}} g(\theta)d\theta \\
\sim\ & e^{dn/2}\left(\frac{\pi}{cn}\right)^{k^2/2} \frac{\exp\left(\frac{-nd^2(\mathbf{vec}\tilde{A})^\top(J_{k-1}+I_{k-1})^{\otimes 2}\mathbf{vec}\tilde{A}}{4ck^2(k^2-2k+2)}\right)}{(k^2-2k+2)^{(k-1)^2/2}\left((k-1)(k-2)\right)^{k-1}\sqrt{2}(k-1)} \\
=\ & e^{dn/2}\left(\frac{2\pi k^2(k-1)^2}{dn}\right)^{k^2/2} \frac{\exp\left(\frac{-nd(k-1)^2(\mathbf{vec}\tilde{A})^\top(J_{k-1}+I_{k-1})^{\otimes 2}\mathbf{vec}\tilde{A}}{2(k^2-2k+2)}\right)}{(k^2-2k+2)^{(k-1)^2/2}\left((k-1)(k-2)\right)^{k-1}\sqrt{2}(k-1)} \\
=\ & \frac{e^{dn/2}}{\sqrt{2}}\left(\frac{2\pi}{dn}\right)^{k^2/2} \frac{k^{k^2}(k-1)^{k(k-1)}\exp\left(\frac{-nd(k-1)^2(\mathbf{vec}\tilde{A})^\top(J_{k-1}+I_{k-1})^{\otimes 2}\mathbf{vec}\tilde{A}}{2(k^2-2k+2)}\right)}{(k^2-2k+2)^{(k-1)^2/2}(k-2)^{k-1}} \\
=\ & I,
\end{aligned}$$

as required. ∎

4 ...and for n not divisible by k

Define $k' = 2k$ if dk is odd or $k' = k$ otherwise. Note that the conditions we assumed so far (i.e. n divisible by k and dn even) can be rewritten as simply $n \equiv 0 \pmod{k'}$. Therefore we only need to consider the case $n \equiv r \pmod{k'}$ for each integer r such that $0 < r < k'$ and dr is even. One possibility is to rework the whole argument of this paper but with colourings that are not exactly equitable. Instead, the asymmetry in the argument can be somewhat reduced by using an argument relating different models of random regular graphs. We first treat the case $n \equiv 0 \pmod{k'}$ in more depth, and prove the following.

Theorem 10 *Fix nonnegative integers $d \geq 3$, k and ℓ such that $d < 2(k-1)\log(k-1)$. Consider the d-regular graphs with n vertices (n divisible by k and dn even) and a distinguished ordered set of ℓ edges whose endpoints induce a perfect matching (i.e. no two edges in the distinguished set are adjacent to the same edge or incident with the same vertex). Let G be chosen uniformly at random from such structures. Then G a.a.s. has a k-colouring in which all ℓ distinguished edges have end vertices coloured 1 and 2.*

Proof. Consider the probability space $\Omega_{n,d,\ell}$ with uniform probability distribution, and whose underlying set consists of pairings in $\mathcal{P}_{n,d}$ with an ordered set L of ℓ distinguished pairs of points, such that no two pairs in L are incident with the same vertex. Let $\hat{Y}$ denote the number of equitable k-colourings of a pairing containing L, in which the distinguished pairs join vertices of colours 1 and 2. We will show that

$$\mathbf{E}\hat{Y} \sim \binom{k}{2}^{-\ell}\mathbf{E}Y, \tag{4.1}$$

that (1.1) holds with Y replaced by $\hat{Y}$ (and no other adjustment), and that

$$\mathbf{E}(\hat{Y}^2) \sim \binom{k}{2}^{-2\ell}\mathbf{E}(Y^2). \tag{4.2}$$

The theorem then follows immediately by applying small subgraph conditioning as in the proof of Theorem 1 for n divisible by k, which was given in Section 1.

To show (4.1) and the analogue of (1.1), we define for integers $r \geq 0$ and $p_1, \ldots, p_r \geq 0$ $N(p_1, \ldots, p_r)$ to be the set of triples (P, C, Γ) such that P is a pairing in $\mathcal{P}_{n,d}$, C is a equitable k-colouring of P and $\Gamma = (\Gamma_{i,j})_{\substack{1 \leq i \leq r \\ 1 \leq j \leq p_i}}$ is an ordered $(p_1 + \cdots + p_r)$-tuple of different cycles in P. In view of that, we can express

$$\mathbf{E}(Y[X_1]_{p_1} \cdots [X_r]_{p_r}) = \frac{|N(p_1, \ldots, p_r)|}{|\mathcal{P}_{n,d}|} \quad \text{and}$$

$$\mathbf{E}(\hat{Y}[X_1]_{p_1} \cdots [X_r]_{p_r}) = \frac{1}{|\Omega_{n,d,\ell}|} \sum_{(P,C,\Gamma) \in N(p_1,\ldots,p_r)} h(P, C), \tag{4.3}$$

where $h(P, C)$ gives the number of choices of the ordered set L of ℓ pairs in P that have the required colours at their ends. Almost all triples in $N(p_1, \ldots, p_r)$ correspond to pairings with $dn/(k(k-1)) + O(n^{1/2} \log n)$ edges between each two colour classes, while the remaining triples contribute with at most a $O(n^{-\Theta(\log n)})$ factor of the total. To verify this claim, observe that for each fixed C and Γ, the number of pairings $P \in \mathcal{P}_{n,d}$ compatible with C and Γ has a factor $F = \sum_{\{b_{i,j}\}} 1/(\prod_{1 \leq i < j \leq k} b_{i,j}!)$, where $b_{i,j}$ denotes the number of edges between colour classes i and j excluding the edges of the cycles in Γ (see the computation of $\mathbf{E}Y$ and $\mathbf{E}(YX_1)$ in Section 2). After using Stirling's formula to estimate the factorials in F, it is easy to check that the main contribution to F is due to terms with all $b_{i,j} = dn/(k(k-1)) + O(n^{1/2} \log n)$ and that the weight of the remaining terms is $O(F/n^{\Theta(\log n)})$ as required. Next, observe that $h(P, C) \sim (dn/(k(k-1)))^\ell$ if P has $dn/(k(k-1)) + O(n^{1/2} \log n)$ edges with endpoints coloured 1 and 2, and that $h(P, C)$ is always $O(n^\ell)$. Therefore, we can estimate the sum in the right side of (4.3) and combine it with $|\Omega_{n,d,\ell}| \sim |\mathcal{P}_{n,d}|(dn/2)^\ell$ to deduce

$$\mathbf{E}(\hat{Y}[X_1]_{p_1} \cdots [X_r]_{p_r}) \sim \binom{k}{2}^{-\ell} \mathbf{E}(Y[X_1]_{p_1} \cdots [X_r]_{p_r}),$$

as required for (4.1) and the analogue of (1.1).

Similarly, to estimate $\mathbf{E}(\hat{Y}^2)$ we define N' to be the set of triples (P, C_1, C_2) such that P is a pairing in $\mathcal{P}_{n,d}$ and C_1, C_2 are equitable k-colourings of P, and write

$$\mathbf{E}(Y^2) = \frac{|N'|}{|\mathcal{P}_{n,d}|} \qquad \text{and} \qquad \mathbf{E}(\hat{Y}^2) = \frac{1}{|\Omega_{n,d,\ell}|} \sum_{(P,C_1,C_2) \in N'} h'(P, C_1, C_2), \tag{4.4}$$

where $h'(P, C_1, C_2)$ gives the number of choices of the ordered set L of ℓ pairs in P that have the required colours at their ends in both colourings C_1 and C_2. Given $P \in \mathcal{P}_{n,d}$, let b_{pqrs} denote the number of edges between points coloured (p, q) and points coloured (r, s). We will show that almost all triples in N' correspond to pairings with $b_{pqrs} = dn/(k^2(k-1)^2) + O(n^{1/2} \log n)$, and that the remaining triples are at most a $O(n^{-\Theta(\log n)})$ fraction of the total. Then (4.2) follows immediately from (4.4), since $h'(P, C_1, C_2)$ is always $O(n^\ell)$ and also $h'(P, C_1, C_2) \sim (2dn/(k^2(k-1)^2))^\ell$ whenever $b_{1122} \sim b_{1221} \sim dn/(k^2(k-1)^2)$. To prove the remaining claim, we first recall from the last lines of the proof of Proposition 2(b) that we can restrict our attention

to triples (P, C_1, C_2) with colour count $M = [m_{p,q}]$ where $|m_{p,q} - 1/k| < \epsilon \log n/n^{1/2}$, since all other triples contribute $O(|N'|/n^{\Theta(\log n)})$ to $|N'|$. Observe that while counting $|N'|$ we encounter a factor $F' = \sum_{\mathcal{B}(M,n)} \prod_{\substack{1 \le p,q,r,s \le k \\ p<r, q \ne s}} 1/b_{pqrs}!$. We can easily bound the weight in F' due to terms in which some b_{pqrs} is not $dn/(k^2(k-1)^2) + O(n^{1/2} \log n)$, and find an extra factor of $O(n^{-\Theta(\log n)})$ compared to the estimation of F' in Lemma 8. ∎

Proof of Theorem 1 (for n not divisible by k). It only remains to show that if $d \ge 3$ and $d < 2(k-1)\log(k-1)$, then $\mathcal{G}_{n,d}$ is a.a.s. k-colourable for n not divisible by k. We use the type of argument employed at the end of Section 3 of [18].

Recall the definition of k' in the beginning of the section, and let r be any integer such that $0 < r < k'$ and dr is even. Suppose $n \equiv r \pmod{k'}$. Take a random d-regular graph G with n vertices, and assume that the first r vertices $v_1, v_2, \ldots, v_r$ are at distance at least 4. This happens a.a.s. and we simply discard G if this property fails to hold. Delete $v_1, v_2, \ldots, v_r$ and then choose a random matching of the dr former neighbours of these vertices, and add the matching to the edges of the graph G. Give the matching edges a random order. Observe that no two matching edges are adjacent to the same edge by construction. It is easy to show and well known that a given vertex of a random d-regular graph is a.a.s. not in a cycle of length less than 4 (or 100, for that matter). It follows that a.a.s. no multiple edges occur due to the new edges. Throw the graph away if this last property fails to hold. The result is a random d-regular graph on $n - r \equiv 0 \pmod{k'}$ vertices with an ordered set of $dr/2$ distinguished edges, no two adjacent to the same edge or incident with the same vertex. Let us call this G'.

The distribution of G' is uniform, since for each G' the number of ways of reinstating the edges to $v_1, v_2, \ldots, v_r$ is exactly $\binom{dr}{d,d,\ldots,d}$. Thus, by Theorem 10, G' a.a.s. has a k-colouring such that the $dr/2$ distinguished edges join vertices of colours 1 and 2. We can use exactly this colouring on $V(G) \setminus \{v_1, v_2, \ldots, v_r\}$, and colour $v_1, v_2, \ldots, v_r$ with colour 3 to obtain a k-colouring of G. (Note that $k \ge 3$ from our assumptions on d and k). ∎

References

[1] M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, volume 55 of *National Bureau of Standards Applied Mathematics Series*. U.S. Government Printing Office, 1972. 10th printing.

[2] D. Achlioptas and C. Moore. The chromatic number of random regular graphs. In *Proc. 7th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems and 8th International Workshop on Randomization and Computation*, volume 3122 of *Lect. Notes Comput. Sc.*, pages 219–228. Springer, 2004.

[3] D. Achlioptas and A. Naor. The two possible values of the chromatic number of a random graph. *Ann. Math.*, 162(3):1333–1349, 2005.

[4] N. Alon and M. Krivelevich. The concentration of the chromatic number of random graphs. *Combinatorica*, 17(3):303–313, 1997.

[5] S. Ben-Shimon and M. Krivelevich. Random regular graphs of non-constant degree: Concentration of the chromatic number. *Discrete Math.*, To appear.

[6] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *Eur. J. Combin.*, 1(4):311–316, 1980.

[7] B. Bollobás. The chromatic number of random graphs. *Combinatorica*, 8(1):49–55, 1988.

[8] A. Coja-Oghlan, K. Panagiotou, and A. Steger. On the chromatic number of random graphs. *J. Comb. Theory B*, 98(5):980–993, 2008.

[9] C. Cooper, A. M. Frieze, B. A. Reed, and O. Riordan. Random regular graphs of non-constant degree: Independence and chromatic number. *Comb. Probab. Comput.*, 11(4):323–341, 2002.

[10] N. de Bruijn. *Asymptotic methods in analysis.* North Holland Publishing Co., third edition, 1970.

[11] J. Díaz, A. C. Kaporis, G. D. Kemkes, L. M. Kirousis, X. Pérez, and N. C. Wormald. On the chromatic number of a random 5-regular graph. *J. Graph Theor.*, To appear.

[12] A. M. Frieze and T. Łuczak. On the independence and chromatic numbers of random regular graphs. *J. Comb. Theory B*, 54(1):123–132, 1992.

[13] S. Janson, T. Łuczak, and A. Ruciński. *Random Graphs.* John Wiley, 2000.

[14] M. Krivelevich, B. Sudakov, V. H. Vu, and N. C. Wormald. Random regular graphs of high degree. *Random Struct. Algor.*, 18(4):346–363, 2001.

[15] T. Łuczak. The chromatic number of random graphs. *Combinatorica*, 11(1):45–54, 1991.

[16] T. Łuczak. A note on the sharp concentration of the chromatic number of random graphs. *Combinatorica*, 11(3):295–297, 1991.

[17] M. Molloy and B. A. Reed. The chromatic number of sparse random graphs. Master's thesis, University of Waterloo, 1992.

[18] R. W. Robinson and N. C. Wormald. Almost all regular graphs are hamiltonian. *Random Struct. Algor.*, 5(2):363–374, 1994.

[19] E. Shamir and J. Spencer. Sharp concentration of the chromatic number on random graphs $G_{n,p}$. *Combinatorica*, 7(1):121–129, 1987.

[20] L. Shi and N. C. Wormald. Colouring random 4-regular graphs. *Comb. Probab. Comput.*, 16(2):309–344, 2007.

[21] L. Shi and N. C. Wormald. Colouring random regular graphs. *Comb. Probab. Comput.*, 16(3):459–494, 2007.

[22] N. C. Wormald. Models of random regular graphs. In J. D. Lamb and D. A. Preece, editors, *Surveys in Combinatorics 1999*, volume 267 of *LMS Lecture Note Series*, pages 239–298. Cambridge University Press, 1999.