Combinatorics, Probability and Computing (2019) 28, 130–158.
doi:10.1017/S0963548318000251

On the Number of Solutions in Random Graph k-Colouring

FELICIA RASSMANN†

Institute of Mathematics, Goethe University, 10 Robert-Mayer-Straße, Frankfurt 60325, Germany
(e-mail: rassmann@math.uni-frankfurt.de)

Received 20 September 2016; revised 19 March 2018; first published online 21 May 2018

Let $k \geqslant 3$ be a fixed integer. We exactly determine the asymptotic distribution of $\ln Z_k(G(n,m))$, where $Z_k(G(n,m))$ is the number of k-colourings of the random graph $G(n,m)$. A crucial observation to this end is that the fluctuations in the number of colourings can be attributed to the fluctuations in the number of small cycles in $G(n,m)$. Our result holds for a wide range of average degrees, and for k exceeding a certain constant k_0 it covers all average degrees up to the so-called *condensation phase transition*.

2010 *Mathematics subject classification*: Primary 05C80
Secondary 05C15

1. Introduction

1.1. Background and motivation

Going back to the ground-breaking paper of Erdős and Rényi [15] in 1960, the study of the random graph colouring problem has attained a lot of attention and innumerable articles have been published in this area of research over the years. In the most frequently studied model, a random graph $G(n,m)$ on the vertex set $[n] = \{1,\dots,n\}$ with precisely m edges is drawn uniformly at random from all such graphs.

A question that has turned out to be a very challenging one is how to choose n and m to obtain a random graph that is colourable w.h.p. Or, put differently, whether a random graph with given n and m can be coloured with a fixed number of colours, thus determining its chromatic number.

Beginning in the 1990s, considerable progress has been made in the case of *sparse* random graphs, where $m = O(n)$ as $n \to \infty$. Much effort has been devoted to studying the typical value of the chromatic number of $G(n,m)$ [3, 9, 21, 22] and its concentration

† The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement no. 278857–PTCC.

[4, 20, 29]. Several experiments and simulations led to the hypothesis, that, when changing the ratio of edges to variables, there is a transition from a regime where the random graph is colourable w.h.p. to the one where it is not w.h.p. Furthermore, the observation was that this transition does not happen smoothly, suggesting the existence of a sharp satisfiability threshold. Indeed, in 1999, Achlioptas and Friedgut [2] proved the existence of a *sharp threshold sequence* $d_{k,\mathrm{col}}(n)$ for any $k \geqslant 3$, meaning that for any fixed $\varepsilon > 0$ the random graph $G(n,m)$ is k-colourable w.h.p. if $2m/n < d_{k,\mathrm{col}}(n) - \varepsilon$, whereas $G(n,m)$ fails to be k-colourable w.h.p. if $2m/n > d_{k,\mathrm{col}}(n) + \varepsilon$. This threshold sequence is non-uniform, namely, it is a function of n, and although it is broadly believed to converge for n tending to infinity, this has not yet been established. Also, in spite of continued efforts, the exact value of this threshold remains unknown to date. The best current bounds [11, 13] on $d_{k,\mathrm{col}}(n)$ show that there is a sequence $(\gamma_k)_{k\geqslant 3}$, $\lim_{k\to\infty}\gamma_k = 0$, such that

$$(2k-1)\ln k - 2\ln 2 - \gamma_k \leqslant \liminf_{n\to\infty} d_{k,\mathrm{col}}(n) \leqslant \limsup_{n\to\infty} d_{k,\mathrm{col}}(n) \leqslant (2k-1)\ln k - 1 + \gamma_k.$$

Nonetheless, there exist predictions by statistical physicists regarding the precise location of this threshold. They developed a method called the *cavity method* that allowed them to gain insights into the combinatorial structure of the random graph colouring problem and to understand the significance of *typical* k-colourings, that is, k-colourings chosen uniformly at random from the set of all k-colourings, on both the combinatorial and algorithmic aspects of the problem [19]. What is more, this method has also been used to predict a further phase transition shortly before the colouring threshold. This transition $d_{k,\mathrm{cond}}$ has been named *condensation* and its existence and location have been rigorously determined in 2014 by Bapst, Coja-Oghlan, Hetterich, Rassmann and Vilenchik [7]. We will formally introduce $d_{k,\mathrm{cond}}$ in the next subsection; for the moment we content ourselves with stating that it plays a very important role for several reasons. It marks the point where the behaviour of the number of solutions changes significantly, as does the geometry of the solution space [1, 23]. The prediction states that while two k-colourings chosen uniformly at random tend to be uncorrelated before the condensation threshold, they typically exhibit long-range correlations afterwards [24]. Furthermore, in contrast to the colourability transition, the condensation transition persists for finite inverse temperatures as well [8]. In recent work, it has been proved that the condensation transition is also related to the information theoretic threshold in the stochastic block model [5], where it marks the point from which it is possible to decide whether or not a random graph has been drawn from a planted distribution.

By obtaining an exact expression for the asymptotic distribution of the logarithm of the number of solutions up to the condensation threshold $d_{k,\mathrm{cond}}$, in the present paper we give a definite and complete answer to the question about the relationship between the planted model and the Gibbs distribution. Furthermore, we show that the fluctuations in the number of solutions can be completely attributed to the presence of short cycles, thereby eliminating the possibility of other influencing factors.

For a graph G on n vertices, we let $Z_k(G)$ be the number of *k-colourings* (also called *solutions*) of G, which are maps $\sigma : [n] \to [k]$ such that $\sigma(i) \neq \sigma(j)$ for all edges $\{i,j\}$ of G. We always consider *sparse* random graphs $G(n,m)$ where $m = O(n)$. We are especially interested in the case where m is of the same order of magnitude as n, more precisely, where $d = 2m/n$ does not grow with n. As we are going to need a very neat computation of the first and second moment of the number of k-colourings of $G(n,m)$, in addition to d, which arises naturally in the computations of the first and second moment, we introduce d', which is an arbitrary but fixed value such that $m = \lceil d'n/2 \rceil$. We note that $d' \sim d$, although $d = d(n)$ might vary slightly with n, whereas d' is assumed to be fixed as $n \to \infty$.

1.2. Results

We show that under certain conditions the number $Z_k(G(n,m))$ of k-colourings of the random graph is tightly concentrated and determines the distribution of

$$\ln Z_k(G(n,m)) - \ln \mathbb{E}[Z_k(G(n,m))]$$

asymptotically in a broad density regime.

Before we state the result, we introduce the following notation. For $k \geqslant 3$, we define the condensation transition

$$d_{k,\text{cond}} = \sup\{d' > 0 : \liminf_{n\to\infty} \mathbb{E}[Z_k(G(n,m))^{1/n}] = k(1-1/k)^{d'/2}\}, \tag{1.1}$$

where $m = \lceil d'n/2 \rceil$ is a function of d'. Indeed, the supremum in the definition of $d_{k,\text{cond}}$ is well defined, as the set

$$\{d' > 0 : \liminf_{n\to\infty} \mathbb{E}[Z_k(G(n,m))^{1/n}] = k(1-1/k)^{d'/2}\}$$

is non-empty due to the fact that for $d' \in (0,1]$ the graph $G(n,m)$ decomposes into tree components w.h.p., and thus in this regime

$$\liminf_{n\to\infty} Z_k(G(n,m))^{1/n} = k(1-1/k)^{d'/2},$$

and it is also bounded, as for densities d' above some $d_{k,\text{col}}$ the graph will have no more colourings. The definition of $d_{k,\text{cond}}$ is motivated by the well-known fact that

$$\mathbb{E}[Z_k(G(n,m))] = \Theta(k^n(1-1/k)^m).$$

This statement will reappear in Proposition 2.1 and will also be proved in this context. Jensen's inequality shows that

$$\limsup_{n\to\infty} \mathbb{E}[Z_k(G(n,m))^{1/n}] \leqslant k(1-1/k)^{d'/2} \quad \text{for all } d',$$

and this upper bound is tight up to the density $d_{k,\text{cond}}$.

Under the assumption that $k \geqslant k_0$ for a certain constant k_0, it is possible to calculate the number $d_{k,\text{cond}}$ precisely [7], and an asymptotic expansion in k yields

$$d_{k,\text{cond}} = (2k-1)\ln k - 2\ln 2 + \gamma_k, \quad \text{where } \lim_{k\to\infty} \gamma_k = 0.$$

To state the main theorem of the paper, we let

$$\lambda_l = \frac{d^l}{2l} \quad \text{and} \quad \delta_l = \frac{(-1)^l}{(k-1)^{l-1}}$$

for $k, l \in \mathbb{N}$. With these definitions, we have the following result.

Theorem 1.1. *There is a constant $k_0 > 3$ such that the following is true. Assume either that $k \geqslant 3$ and $d' \leqslant 2(k-1)\ln(k-1)$ or that $k \geqslant k_0$ and $d' < d_{k,\text{cond}}$.*

For $l \geqslant 2$, let $(X_l)_l$ be a family of independent Poisson variables with $\mathbb{E}[X_l] = \lambda_l$, all defined on the same probability space. Then

$$\ln Z_k(G(n,m)) - \ln \mathbb{E}[Z_k(G(n,m))] \xrightarrow{D} W,$$

where the random variable W is given by

$$W = \sum_{l \geqslant 3} X_l \ln(1+\delta_l) - \lambda_l \delta_l$$

and satisfies $\mathbb{E}|W| < \infty$. □

Remark. By definition, W has an infinitely divisible distribution. It was shown in [17] that the random variable $W' = \exp[W]$ converges almost surely and in L^2 with $\mathbb{E}[W'] = 1$ and $\mathbb{E}[W'^2] = \exp\left[\sum_l \lambda_l \delta_l^2\right]$. Thus, by Jensen's inequality it follows that $\mathbb{E}[W] \leqslant 0$ and $\mathbb{E}[W^2] \leqslant \sum_l \lambda_l \delta_l^2$. □

1.3. Discussion and further related work.

The crucial observation that the proof of Theorem 1.1 builds upon is that the fluctuations of $\ln Z_k(G(n,m))$ can be attributed to variations in the number of cycles of bounded length in the random graph and that this is their only significant influencing factor. As a consequence, conditioning on the cycle count for cycles up to some preselected length reduces the variance of $\ln Z_k(G(n,m))$ enormously.

This was first observed in [6], and it has been used to determine the order of magnitude of the fluctuations of $\ln Z_k(G(n,m))$ in the random graph colouring problem. Following this result, the asymptotic distribution of the logarithm of the number of solutions has been established for random regular k-SAT [14] and random hypergraph 2-colouring [26]. Our result Theorem 1.1 refines the analysis from [14]. We are the first to determine the asymptotic distribution of the logarithm of the number of solutions in random graph k-colouring in a broad density regime up to the condensation transition $d_{k,\text{cond}}$ (for large values of k).

The proof combines the second moment arguments from Achlioptas and Naor [3] and its enhancements from [7, 13] with the 'small subgraph conditioning'. This method was originally developed in [27, 28] and extended by Janson [17] to obtain limiting distributions. It has frequently been used in random regular graph problems (see [30] for an enlightening survey), for example in [18] and [12] to upper-bound the chromatic number of the random d-regular graph, as the sharp threshold result [2] does not apply for this problem. More recently, it has also been used to obtain results in the stochastic block model [5] and to determine the satisfiability threshold for positive 1-in-k-SAT [25].

Unfortunately, Janson's result does not apply directly in our case and instead we have to perform a variance analysis along the lines of [28], analogous to [14, 26]. The reason for this is that in contrast to [6], where only *bounds* on the fluctuation of $\ln Z_k$ were proved, we aim at a statement about its asymptotic *distribution* and thus we need an asymptotically tight expression for the second moment. Thus, in the present paper it does not suffice to consider colourings with *balanced* colour classes (with a deviation of $o(n^{-1/2})$ from their typical value), but we have to get a handle on *all* colourings providing a positive contribution. To this end, we collect together colourings exhibiting similar colour class sizes. This results in the need to consider not only one random variable but a growing number of random variables, and to develop methods to deal with them simultaneously.

We expect that it is possible to apply a combination of the second moment method and small subgraph conditioning to a variety of further random constraint problems, for example random k-NAESAT, random k-XORSAT or random hypergraph k-colourability. However, for asymmetric problems such as the well-known benchmark problem random k-SAT, we expect the logarithm of the number of satisfying assignments to exhibit stronger fluctuations, and we doubt that a result similar to ours can be established.

1.4. Preliminaries and notation

We always assume that $n \geqslant n_0$ is sufficiently large for our various estimates to hold and let $[n]$ denote the set $\{1, \ldots, n\}$.

We use the standard O-notation when referring to the limit $n \to \infty$. Thus, $f(n) = O(g(n))$ means that there exist $C > 0$, $n_0 > 0$ such that for all $n > n_0$ we have $|f(n)| \leqslant C \cdot |g(n)|$. In addition, we use the standard symbols $o(\cdot), \Omega(\cdot), \Theta(\cdot)$. In particular, $o(1)$ stands for a term that tends to 0 as $n \to \infty$. Furthermore, the notation $f(n) \sim g(n)$ means that $f(n) = g(n)(1 + o(1))$ or equivalently $\lim_{n\to\infty} f(n)/g(n) = 1$. Besides taking the limit $n \to \infty$, at some point we need to consider the limit $v \to \infty$ for some number $v \in \mathbb{N}$. Thus, we introduce $f(n, v) \sim_v g(n, v)$, meaning that $\lim_{v\to\infty} \lim_{n\to\infty} f(n, v)/g(n, v) = 1$.

If $p = (p_1, \ldots, p_l)$ is a vector with entries $p_i \geqslant 0$, then we let

$$\mathcal{H}(p) = -\sum_{i=1}^{l} p_i \ln p_i.$$

Here and throughout, we use the convention that $0 \ln 0 = 0$. Hence, if $\sum_{i=1}^{l} p_i = 1$, then $\mathcal{H}(p)$ is the entropy of the probability distribution p. Further, for a number x and an integer $h > 0$ we let $(x)_h = x(x-1)\cdots(x-h+1)$ denote the hth falling factorial of x.

For the sake of simplicity, we choose to prove Theorem 1.1 using the random graph model $\mathfrak{G}(n, m)$. This is a random (multi-)graph on the vertex set $[n]$ obtained by choosing m edges $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m$ of the complete graph on n vertices uniformly and independently at random (*i.e.* with replacement). In this model we may choose the same edge more than once. However, the following statement (proved in [16, Chapter 1.3], for example) shows that for sparse random graphs the probability of this event is bounded away from 1.

Fact 1.2. *Assume that $m = m(n)$ is a sequence such that $m = O(n)$ and let $\mathcal{A}_n$ be the event that $\mathfrak{G}(n, m)$ has no multiple edges. Then there is a constant $c > 0$ such that $\lim_{n\to\infty} \mathbb{P}[\mathcal{A}_n] > c$.*

2. Outline of the proof

To determine bounds on $Z_k(\mathfrak{G}(n, m))$, it will be necessary to control the size of the colour classes. To formalize this, we introduce the following notation. For a map $\sigma : [n] \to [k]$, we define

$$\rho(\sigma) = (\rho_1(\sigma), \dots, \rho_k(\sigma)), \quad \text{where } \rho_i(\sigma) = |\sigma^{-1}(i)|/n \quad \text{for } i = 1, \dots, k.$$

Thus, $\rho(\sigma)$ is a probability distribution on $[k]$, which we refer to as the *colour density* of σ.

Let $\mathcal{A}_k(n)$ signify the set of all possible colour densities $\rho(\sigma)$ for $\sigma : [n] \to [k]$. Further, let $\mathcal{A}_k$ be the set of all probability distributions $\rho = (\rho_1, \dots, \rho_k)$ on $[k]$, and let $\rho^\star = (1/k, \dots, 1/k)$ signify the barycentre of $\mathcal{A}_k$.

In order to simplify the notation, for the rest of the paper we assume that ω, ν are odd natural numbers; formally we define $N = \{2i - 1 : i \in \mathbb{N}\}$ and let $\omega, \nu \in N$. We say that $\rho = (\rho_1, \dots, \rho_k) \in \mathcal{A}_k(n)$ is (ω, n)-*balanced* if

$$\rho_i \in \left[\frac{1}{k} - \frac{\omega}{\sqrt{n}}, \frac{1}{k} + \frac{\omega}{\sqrt{n}}\right) \quad \text{for all } i \in [k]$$

and let $\mathcal{A}_{k,\omega}(n)$ denote the set of all (ω, n)-balanced $\rho \in \mathcal{A}_k(n)$. As we will see, in order to prove statements about the number Z_k of all solutions, it suffices to consider solutions σ with $\rho(\sigma) \in \mathcal{A}_{k,\omega}(n)$. We let $Z_{k,\omega}(G)$ signify the number of (ω, n)-*balanced* k-*colourings* of a graph G on $[n]$, that is, k-colourings σ such that $\rho(\sigma) \in \mathcal{A}_{k,\omega}(n)$.

Since verifying the required properties to apply small subgraph conditioning directly for the random variable $Z_{k,\omega}$ is very intricate, we break $Z_{k,\omega}$ down into smaller contributions, for which we determine the first and second moment in the following sections.

To this end, we decompose the set $\mathcal{A}_{k,\omega}(n)$ into smaller sets. We define

$$S_{k,\omega,\nu} = \left\{ s \in \mathbb{Z}^k : \|s\|_1 = 2i, i \in \mathbb{N}, i \leqslant \frac{\omega\nu - 1}{2} \right\}. \tag{2.1}$$

$S_{k,\omega,\nu}$ contains vectors that we use as centres of disjoint 'balls' to partition the set $\mathcal{A}_{k,\omega}(n)$. For $s = (s_1, \dots, s_k) \in S_{k,\omega,\nu}$, we let $\rho^{k,\omega,\nu,s} \in \mathbb{R}^k$ be the vector with components

$$\rho_i^{k,\omega,\nu,s} = \frac{1}{k} + \frac{s_i}{\nu\sqrt{n}}. \tag{2.2}$$

Further, we let $\mathcal{A}^s_{k,\omega,\nu}(n)$ be the set of all colour densities $\rho \in \mathcal{A}_{k,\omega}(n)$ such that

$$\rho_i \in \left[\rho_i^{k,\omega,\nu,s} - \frac{1}{\nu\sqrt{n}}, \rho_i^{k,\omega,\nu,s} + \frac{1}{\nu\sqrt{n}}\right).$$

For a graph G, we let $Z^s_{k,\omega,\nu}(G)$ denote the number of k-colourings σ such that $\rho(\sigma) \in \mathcal{A}^s_{k,\omega,\nu}(n)$. For each fixed ν, we have $Z_{k,\omega} = \sum_{s \in S_{k,\omega,\nu}} Z^s_{k,\omega,\nu}$ and our strategy is to apply small subgraph conditioning to the random variables $Z^s_{k,\omega,\nu}$ rather than directly to Z_k. But first, we will calculate the first moments of Z_k and $Z_{k,\omega}$ in Section 3 to obtain the following.

Proposition 2.1. *Fix an integer $k \geqslant 3$ and a number $d' \in (0, \infty)$ as defined at the end of Section 1.1. Let $\omega > 0$. Then*

$$\mathbb{E}[Z_k(\mathfrak{G}(n,m))] = \Theta(k^n(1 - 1/k)^m) \quad \text{and} \quad \lim_{\omega \to \infty} \liminf_{n \to \infty} \frac{\mathbb{E}[Z_{k,\omega}(\mathfrak{G}(n,m))]}{\mathbb{E}[Z_k(\mathfrak{G}(n,m))]} = 1.$$

□

https://doi.org/10.1017/S0963548318000251 Published online by Cambridge University Press

As discussed in Section 1.3, the key observation the proof is based on is that the fluctuations of $Z_k(\mathfrak{G}(n,m))$ can be attributed to fluctuations in the number of cycles of a bounded length. Hence, for an integer $l \geqslant 2$ we let $C_{l,n}$ denote the number of cycles of length exactly l in $\mathfrak{G}(n,m)$. Let

$$\lambda_l = \frac{d^l}{2l} \quad \text{and} \quad \delta_l = \frac{(-1)^l}{(k-1)^{l-1}}. \tag{2.3}$$

We will see that λ_l asymptotically denotes the expected number of cycles of length l in a random graph, whereas δ_l is a correction factor taking into account that we do not allow for edges connecting two vertices of the same colour. It is easy to verify that for $k \geqslant 3$ we have

$$\sum_{l=2}^{\infty} \lambda_l \delta_l^2 < \infty.$$

The following fact shows that $C_{2,n}, \dots$ are asymptotically independent Poisson variables (this is a standard application of [10, Theorem 1.23] together with a monotonicity argument).

Fact 2.2. *If $c_2, \dots, c_L$ are non-negative integers, then*

$$\lim_{n\to\infty} \mathbb{P}[\forall 2 \leqslant l \leqslant L : C_{l,n} = c_l] = \prod_{l=2}^{L} \mathbb{P}[\mathrm{Po}(\lambda_l) = c_l].$$

In [6] the impact of the cycle counts $C_{l,n}$ on the first moment of $Z^s_{k,\omega,\nu}(\mathfrak{G}(n,m))$ was investigated. The result is the following.

Proposition 2.3. *Assume that $k \geqslant 3$ and $d' \in (0,\infty)$. Moreover, let $\omega, \nu \in N$ and $c_2, \dots, c_L$ be non-negative integers. Then*

$$\frac{\mathbb{E}[Z^s_{k,\omega,\nu}(\mathfrak{G}(n,m)) | \forall 2 \leqslant l \leqslant L : C_{l,n} = c_l]}{\mathbb{E}[Z^s_{k,\omega,\nu}(\mathfrak{G}(n,m))]} \sim \prod_{l=2}^{L} [1+\delta_l]^{c_l} \exp[-\delta_l \lambda_l]. \tag{2.4}$$

□

Thus, this proposition quantifies the influence of the number of short cycles on the expectation of $Z^s_{k,\omega,\nu}(\mathfrak{G}(n,m))$. In addition, to apply small subgraph conditioning, we have to determine the second moment of $Z^s_{k,\omega,\nu}(\mathfrak{G}(n,m))$ very precisely. This step constitutes the main technical work of this paper. We consider two regimes of d' and k separately. In the simpler case, based on the second moment argument from [3], we obtain the following result.

Proposition 2.4. *Assume that $k \geqslant 3$ and $d' < 2(k-1)\ln(k-1)$. Then*

$$\frac{\mathbb{E}[Z^s_{k,\omega,\nu}(\mathfrak{G}(n,m))^2]}{\mathbb{E}[Z^s_{k,\omega,\nu}(\mathfrak{G}(n,m))]^2} \sim \exp\left[\sum_{l\geqslant 2} \lambda_l \delta_l^2\right] = \left(1 - \frac{d}{(k-1)^2}\right)^{-(k-1)^2/2} \exp\left[-\frac{d}{2}\right]. \qquad \square$$

The second regime of d' and k is that $k \geqslant k_0$ for a certain constant $k_0 \geqslant 3$ and $d' < d_{k,\mathrm{cond}}$ (with $d_{k,\mathrm{cond}}$ the number defined in (1.1)). In this case, we replace $Z^s_{k,\omega,v}$ with the slightly tweaked random variable $\widetilde{Z}^s_{k,\omega,v}$ used in the second moment arguments from [7, 13].

Proposition 2.5. *There is a constant $k_0 \geqslant 3$ such that the following is true. Assume that $k \geqslant k_0$ and $2(k-1)\ln(k-1) \leqslant d' < d_{k,\mathrm{cond}}$. Then for each $\omega, v \in N$ and $s \in S_{k,\omega,v}$ there exists an integer-valued random variable $\widetilde{Z}^s_{k,\omega,v}$ which is such that $0 \leqslant \widetilde{Z}^s_{k,\omega,v} \leqslant Z^s_{k,\omega,v}$ and for which it holds that*

$$\mathbb{E}[\widetilde{Z}^s_{k,\omega,v}(\mathfrak{G}(n,m))] \sim \mathbb{E}[Z^s_{k,\omega,v}(\mathfrak{G}(n,m))] \quad \text{and} \tag{2.5}$$

$$\frac{\mathbb{E}[\widetilde{Z}^s_{k,\omega,v}(\mathfrak{G}(n,m))^2]}{\mathbb{E}[\widetilde{Z}^s_{k,\omega,v}(\mathfrak{G}(n,m))]^2} \leqslant (1+o(1))\ \exp\left[\sum_{l\geqslant 2}\lambda_l\delta_l^2\right]. \qquad \square$$

The proofs of Propositions 2.4 and 2.5 appear at the end of Section 4. In order to apply small subgraph conditioning to the random variable $\widetilde{Z}^s_{k,\omega,v}$, we need to investigate the impact of $C_{l,n}$ on the first moment of $\widetilde{Z}^s_{k,\omega,v}$. Thus, we need a similar result to Proposition 2.3 for $\widetilde{Z}^s_{k,\omega,v}$. Fortunately, instead of having to reiterate the proof of Proposition 2.3, we obtain the following by combining Proposition 2.3 with (2.5).

Corollary 2.6. *Let $c_2, \ldots, c_L$ be non-negative integers. With the assumptions and notation of Proposition 2.5 we have*

$$\frac{\mathbb{E}[\widetilde{Z}^s_{k,\omega,v}(\mathfrak{G}(n,m)) | \forall 2 \leqslant l \leqslant L : C_{l,n} = c_l]}{\mathbb{E}[\widetilde{Z}^s_{k,\omega,v}(\mathfrak{G}(n,m))]} \sim \prod_{l=2}^{L}[1+\delta_l]^{c_l}\exp[-\delta_l\lambda_l]. \qquad \square$$

The proof of this statement is nearly identical to the one in [6].

The aim is now to derive Theorem 1.1 from Propositions 2.1–2.4. The key observation is that the variance of the random variables $Z^s_{k,\omega,v}$ is affected by the presence of cycles of bounded length and that this is the only significant influence. As a consequence, conditioning on the small cycle counts up to some preselected length reduces the variance of $Z^s_{k,\omega,v}$. What is maybe surprising is that conditioning on the number of sufficiently small cycles reduces the variance to any desired fraction of $\mathbb{E}[Z^s_{k,\omega,v}]^2$.

As in [14, 26], the arguments we use are similar to the small subgraph conditioning from [17, 28]. But we do not refer to any technical statements from [17, 28] directly, because instead of working only with the random variable Z_k we need to control all $Z^s_{k,\omega,v}$ for fixed $\omega, v \in N$ simultaneously. In fact, ultimately we have to take $v \to \infty$ and $\omega \to \infty$ as well. Our line of argument follows the path beaten in [14, 26], and the following three lemmas are nearly identical to the ones derived there.

For $L > 2$, let $\mathcal{F}_L = \mathcal{F}_{L,n}(d,k)$ be the σ-algebra generated by the random variables $C_{l,n}$ with $2 \leqslant l \leqslant L$. The set of all graphs can be divided into groups according to the small cycle counts. For each $L \geqslant 2$, the decomposition of the variance of $Z^s_{k,\omega,v}$ yields

$$\mathrm{Var}[Z^s_{k,\omega,v}(\mathfrak{G}(n,m))] = \mathrm{Var}[\mathbb{E}[Z^s_{k,\omega,v}(\mathfrak{G}(n,m))|\mathcal{F}_L]] + \mathbb{E}[\mathrm{Var}[Z^s_{k,\omega,v}(\mathfrak{G}(n,m))|\mathcal{F}_L]],$$

https://doi.org/10.1017/S0963548318000251 Published online by Cambridge University Press

meaning that the variance can be written as the variance of the group mean plus the expected value of the variance within a group. The term $\mathrm{Var}[\mathbb{E}[Z^s_{k,\omega,v}(\mathfrak{G}(n,m))|\mathcal{F}_L]]$ accounts for the amount of variance induced by the fluctuations of the number of cycles of length at most L. The strategy when using small subgraph conditioning is to bound the second summand, which is the expected conditional variance

$$\mathbb{E}[\mathrm{Var}[Z^s_{k,\omega,v}(\mathfrak{G}(n,m))|\mathcal{F}_L]] = \mathbb{E}[\mathbb{E}[Z^s_{k,\omega,v}(\mathfrak{G}(n,m))^2|\mathcal{F}_L] - \mathbb{E}[Z^s_{k,\omega,v}(\mathfrak{G}(n,m))|\mathcal{F}_L]^2].$$

In the following lemma we show that in fact in the limit of large L and n this quantity is negligible. This implies that conditioned on the number of short cycles the variance vanishes, and thus the limiting distribution of $\ln Z^s_{k,\omega,v}$ is just the limit of $\ln \mathbb{E}[Z^s_{k,\omega,v}|\mathcal{F}_L]$ as $n, L \to \infty$. This limit is determined by the joint distribution of the number of short cycles.

Lemma 2.7. *Let $k \geqslant 3$ and $d' < 2(k-1)\ln(k-1)$. For any $\omega, v \in N$ and $s \in S_{k,\omega,v}$, we have*

$$\limsup_{L\to\infty} \limsup_{n\to\infty} \frac{\mathbb{E}[\mathbb{E}[Z^s_{k,\omega,v}(\mathfrak{G}(n,m))^2|\mathcal{F}_L] - \mathbb{E}[Z^s_{k,\omega,v}(\mathfrak{G}(n,m))|\mathcal{F}_L]^2]}{\mathbb{E}[Z^s_{k,\omega,v}(\mathfrak{G}(n,m))]^2} = 0.$$

Proof. Fix $\omega, v \in N$ and set $Z_s = Z^s_{k,\omega,v}(\mathfrak{G}(n,m))$. Using Fact 2.2 and Proposition 2.3, we can choose for any $\varepsilon > 0$ a constant $B = B(\varepsilon)$ and $L \geqslant L_0(\varepsilon)$ sufficiently large that, for each sufficiently large $n \geqslant n_0(\varepsilon, B, L)$, we have for any $s \in S_{k,\omega,v}$

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[Z_s|\mathcal{F}_L]^2] &\geqslant \sum_{c_1,\dots,c_L \leqslant B} \mathbb{E}[Z_s|\forall 2 \leqslant l \leqslant L : C_{l,n} = c_l]^2\, \mathbb{P}[\forall 2 \leqslant l \leqslant L : C_{l,n} = c_l] \\
&\geqslant \exp[-\varepsilon]\mathbb{E}[Z_s]^2 \sum_{c_1,\dots,c_L \leqslant B} \prod_{l=2}^{L} [(1+\delta_l)^{c_l} \exp[-\lambda_l\delta_l]]^2\, \mathbb{P}[\mathrm{Po}(\lambda_l) = c_l] \\
&= \exp[-\varepsilon]\mathbb{E}[Z_s]^2 \sum_{c_1,\dots,c_L \leqslant B} \prod_{l=2}^{L} \frac{[(1+\delta_l)^2\lambda_l]^{c_l}}{c_l!\exp[2\lambda_l\delta_l + \lambda_l]} \\
&= \exp[-\varepsilon]\mathbb{E}[Z_s]^2 \prod_{l=2}^{L} \exp[-2\lambda_l\delta_l - \lambda_l] \sum_{c_1,\dots,c_L \leqslant B} \frac{[(1+\delta_l)^2\lambda_l]^{c_l}}{c_l!} \\
&\geqslant \mathbb{E}[Z_s]^2 \exp\left[-2\varepsilon + \sum_{l=2}^{L} \delta_l^2\lambda_l\right],
\end{aligned} \tag{2.6}$$

where we used that $\exp(j) = \sum_{n=0}^{\infty} j^n/n!$ in the last step. The tower property for conditional expectations and the standard formula for the decomposition of the variance yields

$$\mathbb{E}[Z_s^2] = \mathbb{E}[\mathbb{E}[Z_s^2|\mathcal{F}_L]] = \mathbb{E}[\mathbb{E}[Z_s^2|\mathcal{F}_L] - \mathbb{E}[Z_s|\mathcal{F}_L]^2] + \mathbb{E}[\mathbb{E}[Z_s|\mathcal{F}_L]^2],$$

and thus using (2.6) we have

$$\frac{\mathbb{E}[\mathbb{E}[Z_s^2|\mathcal{F}_L] - \mathbb{E}[Z_s|\mathcal{F}_L]^2]}{\mathbb{E}[Z_s]^2} \leqslant \frac{\mathbb{E}[Z_s^2]}{\mathbb{E}[Z_s]^2} - \exp\left[-2\varepsilon + \sum_{l=2}^{L} \delta_l^2\lambda_l\right]. \tag{2.7}$$

Finally, the estimate $\exp[-x] \geqslant 1 - x$ combined with (2.7) and Proposition 2.4 implies that for sufficiently large v, n, L and each $s \in S_{k,\omega,v}$ we have

$$\frac{\mathbb{E}[\mathbb{E}[Z_s^2|\mathcal{F}_L] - \mathbb{E}[Z_s|\mathcal{F}_L]^2]}{\mathbb{E}[Z_s]^2} \leqslant 2\varepsilon \exp\left[\sum_{l=2}^{\infty} \delta_l^2 \lambda_l\right].$$

As this holds for any $\varepsilon > 0$ and by equation (2.3) the expression $\exp\left[\sum_{l=2}^{\infty} \delta_l^2 \lambda_l\right]$ is finite, the proof of the lemma is completed by first taking $n \to \infty$ and then $L \to \infty$. □

Lemma 2.8. *For any $\alpha > 0$, we have*

$$\limsup_{L\to\infty} \limsup_{n\to\infty} \mathbb{P}[|Z_k(\mathfrak{G}(n,m)) - \mathbb{E}[Z_k(\mathfrak{G}(n,m))|\mathcal{F}_L]| > \alpha \mathbb{E}[Z_k(\mathfrak{G}(n,m))]] = 0.$$

Proof. To unclutter the notation, we set $Z_k = Z_k(\mathfrak{G}(n,m))$ and $Z_{k,\omega} = Z_{k,\omega}(\mathfrak{G}(n,m))$. First we observe that Proposition 2.1 implies that for any $\alpha > 0$ we can choose $\omega \in N$ sufficiently large that

$$\liminf_{n\to\infty} \frac{\mathbb{E}[Z_{k,\omega}]}{\mathbb{E}[Z_k]} > (1 - \alpha^2). \tag{2.8}$$

We let $v \in N$. To prove the statement, we need to get a handle on the cases where the variables $Z_{k,\omega,v}^s(\mathfrak{G}(n,m))$ deviate strongly from their conditional expectation $\mathbb{E}[Z_{k,\omega,v}^s(\mathfrak{G}(n,m))|\mathcal{F}_L]$. We let $Z_s = Z_{k,\omega,v}^s(\mathfrak{G}(n,m))$ and define

$$X_s = |Z_s - \mathbb{E}[Z_s|\mathcal{F}_L]| \cdot \mathbf{1}_{\{|Z_s - \mathbb{E}[Z_s|\mathcal{F}_L]| > \alpha\mathbb{E}[Z_s]\}}$$

and $X = \sum_{s \in S_{k,\omega,v}} X_s$. Then these definitions directly yield

$$\mathbb{P}[X < \alpha \mathbb{E}[Z_{k,\omega}]] \leqslant \mathbb{P}[|Z_{k,\omega} - \mathbb{E}[Z_{k,\omega}|\mathcal{F}_L]| < 2\alpha \mathbb{E}[Z_{k,\omega}]]. \tag{2.9}$$

By the definition of the X_s and Chebyshev's inequality it is true for every s that

$$\mathbb{E}[X_s|\mathcal{F}_L] \leqslant \sum_{j\geqslant 0} 2^{j+1}\alpha\mathbb{E}[Z_s]\,\mathbb{P}[|Z_s - \mathbb{E}[Z_s|\mathcal{F}_L]| > 2^j\alpha\mathbb{E}[Z_s]] \leqslant \frac{4\mathrm{Var}[Z_s|\mathcal{F}_L]}{\alpha\mathbb{E}[Z_s]}.$$

Hence, using that with Proposition 2.1 there is a number $\beta = \beta(\alpha, \omega)$ such that $\mathbb{E}[Z_s]/\mathbb{E}[Z_k] \leqslant \beta/(|S_{k,\omega,v}|)$ for all $s \in S_{k,\omega,v}$ and n sufficiently large, we have

$$\mathbb{E}[X|\mathcal{F}_L] \leqslant \sum_{s\in S_{k,\omega,v}} \frac{4\mathrm{Var}[Z_s|\mathcal{F}_L]}{\alpha\mathbb{E}[Z_s]} \leqslant \frac{4\beta\mathbb{E}[Z_k]}{\alpha|S_{k,\omega,v}|} \sum_{s\in S_{k,\omega,v}} \frac{\mathrm{Var}[Z_s|\mathcal{F}_L]}{\mathbb{E}[Z_s]^2}.$$

Taking expectations, choosing $\varepsilon = \varepsilon(\alpha, \beta, \omega)$ sufficiently small and applying Lemma 2.7, we obtain

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_L]] \leqslant \frac{4\beta\mathbb{E}[Z_k]}{\alpha|S_{k,\omega,v}|} \sum_{s\in S_{k,\omega,v}} \frac{\mathbb{E}[\mathrm{Var}[Z_s|\mathcal{F}_L]]}{\mathbb{E}[Z_s]^2} \leqslant \frac{4\beta\varepsilon\mathbb{E}[Z_k]}{\alpha} \leqslant \alpha^2\mathbb{E}[Z_k]. \tag{2.10}$$

Using (2.9), Markov's inequality, (2.10) and (2.8), it follows that

$$\mathbb{P}[|Z_{k,\omega} - \mathbb{E}[Z_{k,\omega}|\mathcal{F}_L]| < 2\alpha\mathbb{E}[Z_{k,\omega}]] \geqslant 1 - 2\alpha. \tag{2.11}$$

https://doi.org/10.1017/S0963548318000251 Published online by Cambridge University Press

Finally, the triangle inequality combined with Markov's inequality and equations (2.8) and (2.11) yields

$$\begin{aligned}&\mathbb{P}[|Z_k - \mathbb{E}[Z_k|\mathcal{F}_L]| > \alpha\mathbb{E}[Z_k]] \\ &\quad \leqslant \mathbb{P}[|Z_k - Z_{k,\omega}| + |Z_{k,\omega} - \mathbb{E}[Z_{k,\omega}|\mathcal{F}_L]| + |\mathbb{E}[Z_{k,\omega}|\mathcal{F}_L] - \mathbb{E}[Z_k|\mathcal{F}_L]| > \alpha\mathbb{E}[Z_k]] \\ &\quad \leqslant 3\alpha + \alpha/3 + 3\alpha < 7\alpha,\end{aligned}$$

which proves the statement. □

Lemma 2.9. *Let*

$$U_L = \sum_{l=2}^{L} C_{l,n} \ln(1+\delta_l) - \lambda_l\delta_l. \tag{2.12}$$

Then $\limsup_{L\to\infty} \limsup_{n\to\infty} \mathbb{E}[|U_L|] < \infty$*, and further, for any* $\varepsilon > 0$ *we have*

$$\limsup_{L\to\infty} \limsup_{n\to\infty} \mathbb{P}[|\ln \mathbb{E}[Z_k(\mathfrak{G}(n,m))|\mathcal{F}_L] - \ln \mathbb{E}[Z_k(\mathfrak{G}(n,m))] - U_L| > \varepsilon] = 0 \tag{2.13}$$

Proof. In a first step we show that $\mathbb{E}[|U_L|]$ is uniformly bounded. As $x - x^2 \leqslant \ln(1+x) \leqslant x$ for $|x| \leqslant 1/8$, we have for every $l \leqslant L$

$$\mathbb{E}[|C_{l,n} \ln(1+\delta_l) - \lambda_l\delta_l|] \leqslant \delta_l\mathbb{E}[|C_{l,n} - \lambda_l|] + \delta_l^2\mathbb{E}[C_{l,n}].$$

Therefore, Fact 2.2 implies that

$$\mathbb{E}[|U_L|] \leqslant \sum_{l=2}^{L} \delta_l\sqrt{\lambda_l} + \delta_l^2\lambda_l. \tag{2.14}$$

Proposition 2.3 ensures that $\sum_l \delta_l^2\lambda_l < \infty$. Furthermore, as $d' \leqslant (2k-1)\ln k$, we have

$$\sum_l \delta_l\sqrt{\lambda_l} \leqslant \sum_l k^l 2^{-(k-1)l/2} < \infty$$

and thus (2.14) shows that $\mathbb{E}[|U_L|]$ is uniformly bounded.

To prove (2.13), for given n and a constant $B > 0$ we let $\mathcal{C}_B$ be the event that $C_{l,n} < B$ for all $l \leqslant L$. Referring to Fact 2.2, we can find for each $L, \varepsilon > 0$ a $B > 0$ such that

$$\mathbb{P}[\mathcal{C}_B] > 1 - \varepsilon. \tag{2.15}$$

To simplify the notation we set $Z_k = Z_k(\mathfrak{G}(n,m))$ and $Z_{k,\omega} = Z_{k,\omega}(\mathfrak{G}(n,m))$. By Proposition 2.1 we can choose for any $\alpha > 0$ a $\omega > 0$ sufficiently large that $\mathbb{E}[Z_{k,\omega}] > (1-\alpha)\mathbb{E}[Z_k]$ for sufficiently large n. Then Propositions 2.1 and 2.3 combined with Fact 2.2 imply that, for any $c_1, \ldots, c_L \leqslant B$ and sufficiently small $\alpha = \alpha(\varepsilon, L, B)$, we have for sufficiently large n

$$\begin{aligned}\mathbb{E}[Z_k|\forall 2 \leqslant l \leqslant L : C_{l,n} = c_l] &\geqslant \mathbb{E}[Z_{k,\omega}|\forall 2 \leqslant l \leqslant L : C_{l,n} = c_l] \\ &\geqslant \exp[-\varepsilon]\mathbb{E}[Z_k] \prod_{l=2}^{L} (1+\delta_l)^{c_l} \exp[-\delta_l\lambda_l].\end{aligned} \tag{2.16}$$

On the other hand, for sufficiently small α and sufficiently large n we have

$$\begin{aligned}
&\mathbb{E}[Z_k|\forall 2 \leqslant l \leqslant L : C_{l,n} = c_l] \\
&\quad = \mathbb{E}[Z_k - Z_{k,\omega}|\forall 2 \leqslant l \leqslant L : C_{l,n} = c_l] + \mathbb{E}[Z_{k,\omega}|\forall 2 \leqslant l \leqslant L : C_{l,n} = c_l] \\
&\quad \leqslant \frac{2\alpha \mathbb{E}[Z_k]}{\prod_{l=2}^{L} \mathbb{P}[\mathrm{Po}(\lambda_l) = c_l]} + \mathbb{E}[Z_{k,\omega}|\forall 2 \leqslant l \leqslant L : C_{l,n} = c_l] \\
&\quad \leqslant \exp[\varepsilon]\mathbb{E}[Z_k] \prod_{l=2}^{L} (1+\delta_l)^{c_l} \exp[-\delta_l \lambda_l]
\end{aligned} \tag{2.17}$$

Thus, the proof of (2.13) is completed by combining (2.15), (2.16) and (2.17) and taking logarithms. □

Proof of Theorem 1.1. From Lemmas 2.9 and 2.8 it follows that for any $\varepsilon > 0$ we have

$$\limsup_{L\to\infty} \limsup_{n\to\infty} \mathbb{P}[|\ln Z_k(\mathfrak{G}(n,m)) - \ln \mathbb{E}[Z_k(\mathfrak{G}(n,m))] - U_L| > \varepsilon] = 0, \tag{2.18}$$

with U_L defined as in (2.12). We now let

$$U'_L = \sum_{l=3}^{L} C_{l,n} \ln(1+\delta_l) - \lambda_l \delta_l,$$

and let S denote the event that $\mathfrak{G}(n,m)$ consists of m distinct edges, or, equivalently, that no cycles of length 2 exist in $\mathfrak{G}(n,m)$. Given that S occurs, $\mathfrak{G}(n,m)$ is identical to $G(n,m)$ and U'_L is identical to U_L for any $L \geqslant 3$. Furthermore, Fact 1.2 implies that $\mathbb{P}[S] = \Omega(1)$. Consequently, (2.18) yields

$$\begin{aligned}
0 &= \limsup_{L\to\infty} \lim_{n\to\infty} \mathbb{P}[|\ln Z_k(\mathfrak{G}(n,m)) - \ln \mathbb{E}[Z_k(\mathfrak{G}(n,m))] - U_L| > \varepsilon|S] \\
&= \limsup_{L\to\infty} \lim_{n\to\infty} \mathbb{P}[|\ln Z_k(G(n,m)) - \ln \mathbb{E}[Z_k(\mathfrak{G}(n,m))] - U'_L| > \varepsilon].
\end{aligned} \tag{2.19}$$

In a next step, we define

$$W_L = \sum_{l=3}^{L} X_l \ln(1+\delta_l) - \lambda_l \delta_l$$

and remember that

$$W = \sum_{l\geqslant 3} X_l \ln(1+\delta_l) - \lambda_l \delta_l.$$

As $(X_l)_l$ are independent Poisson random variables by definition, Fact 2.2 implies that for each L the random variables U'_L converge in distribution to W_L as $n \to \infty$. Furthermore, because

$$\mathbb{E}[W_L^+] \leqslant \mathbb{E}[|W_L|] \leqslant \sum_{l=3}^{L} \delta_l \sqrt{\lambda_l} + \delta_l^2 \lambda_l < \infty$$

by a reasoning analogous to (2.14) and the explanation thereafter, the martingale convergence theorem implies that W is well-defined and that the W_L converge to W

https://doi.org/10.1017/S0963548318000251 Published online by Cambridge University Press

almost surely as $L \to \infty$. As Lemma 3.1 implies that

$$\mathbb{E}[Z_k(\mathfrak{G}(n,m))], \mathbb{E}[Z_k(G(n,m)] = \Theta(k^n(1-1/k)^m),$$

we have

$$\mathbb{E}[Z_k(\mathfrak{G}(n,m))] = \Theta(\mathbb{E}[Z_k(G(n,m)]),$$

and with (2.19) Theorem 1.1 follows. □

3. The first moment

The aim in this section is to prove Proposition 2.1. The calculations that have to be done follow the path beaten in [3, 13, 18, 26] and are in fact very similar to [6]. Furthermore, at the end of the section we state a result that we need for the proof of Proposition 2.4.

Let $Z_{k,\rho}(G)$ be the number of k-colourings of the graph G with colour density ρ. We define

$$f_1 : \rho \in \mathcal{A}_k \mapsto \mathcal{H}(\rho) + \frac{d}{2} \ln\left(1 - \sum_{i=1}^{k} \rho_i^2\right),$$

where $\mathcal{H}(\rho)$ denotes the entropy function introduced in Section 1.4. In order to determine the expectation of $Z_{k,\rho}$, we have to analyse the function $f_1(\rho)$. Let $\rho^\star$ be a k-dimensional vector with all entries set to $1/k$. The following lemma was obtained in [6].

Lemma 3.1. *Let $k \geqslant 3$ and $d' \in (0,\infty)$. Then there exist numbers*

$$C_1 = C_1(k,d'), C_2 = C_2(k,d') > 0$$

such that for any $\rho \in \mathcal{A}_k(n)$ we have

$$C_1 n^{(1-k)/2} \exp[nf_1(\rho)] \leqslant \mathbb{E}[Z_{k,\rho}(\mathfrak{G}(n,m))] \leqslant C_2 \exp[nf_1(\rho)]. \tag{3.1}$$

Moreover, if $\|\rho - \rho^\star\|_2 = o(1)$ and $d = 2m/n$, then

$$\mathbb{E}[Z_{k,\rho}(\mathfrak{G}(n,m))] \sim (2\pi n)^{(1-k)/2}\, k^{k/2} \exp[d/2 + nf_1(\rho)]. \tag{3.2}$$

□

We can now state the expectation of Z_k.

Corollary 3.2. *For any $k \geqslant 3, d' \in (0,\infty)$ and $d = 2m/n$, we have*

$$\mathbb{E}[Z_k(\mathfrak{G}(n,m))] \sim \exp[d/2 + nf_1(\rho^\star)]\left(1 + \frac{d}{k-1}\right)^{-(k-1)/2}. \tag{3.3}$$

Furthermore, for $\omega > 0$ we have

$$\lim_{\omega\to\infty} \liminf_{n\to\infty} \frac{\mathbb{E}[Z_{k,\omega}(\mathfrak{G}(n,m))]}{\mathbb{E}[Z_k(\mathfrak{G}(n,m))]} = 1. \tag{3.4}$$

https://doi.org/10.1017/S0963548318000251 Published online by Cambridge University Press

Proof of Proposition 2.1. The proposition is immediate from Corollary 3.2, as evaluating $nf_1(\rho^\star)$ yields

$$nf_1(\rho^\star) = \Theta(k^n(1-1/k)^m)$$

and d is chosen to be constant. □

Finally, as our approach requires the analysis of the random variables $Z^s_{k,\omega,v}(\mathfrak{G}(n,m))$, we derive an expression for $\mathbb{E}[Z^s_{k,\omega,v}(\mathfrak{G}(n,m))]$ that we will need to prove Proposition 2.4.

Lemma 3.3. *Let $k \geqslant 3, \omega, v \in N, d' \in (0,\infty)$ and $d = 2m/n$. For $s \in S_{k,\omega,v}$ and $\rho^{k,\omega,v,s}$ as defined in (2.2), we have*

$$\mathbb{E}[Z^s_{k,\omega,v}(\mathfrak{G}(n,m))] \sim_v |\mathcal{A}^s_{k,\omega,v}(n)|(2\pi n)^{(1-k)/2}\, k^{k/2} \exp[d/2 + nf_1(\rho^{k,\omega,v,s})].$$

Proof. Using a Taylor expansion of $f_1(\rho)$ around $\rho = \rho^{k,\omega,v,s}$, we get

$$f_1(\rho) = f_1(\rho^{k,\omega,v,s}) + \Theta\left(\frac{\omega}{\sqrt{n}}\right)\|\rho - \rho^{k,\omega,v,s}\|_1 + \Theta(\|\rho - \rho^{k,\omega,v,s}\|_2^2). \tag{3.5}$$

Since

$$\|\rho - \rho^{k,\omega,v,s}\|_1 = O\left(\frac{1}{v\sqrt{n}}\right) \quad \text{for } \rho \in \mathcal{A}^s_{k,\omega,v}(n)$$

and

$$\|\rho - \rho^{k,\omega,v,s}\|_2^2 = O\left(\frac{1}{v^2 n}\right),$$

we conclude that

$$f_1(\rho) = f_1(\rho^{k,\omega,v,s}) + O\left(\frac{\omega}{vn}\right),$$

and as this is independent of ρ, the assertion follows by inserting (3.5) in (3.2) and multiplying by $|\mathcal{A}^s_{k,\omega,v}(n)|$. □

4. The second moment

The aim of this section is to prove Propositions 2.4 and 2.5. These two propositions yield the important result that the second moment of the adequately chosen number of solutions can be bounded by a constant times the square of its first moment. The proof of Proposition 2.4 constitutes the main technical contribution of this work and is implemented in the following way. In Section 4.1 we introduce the overlap and some further important notation and crucial estimates on the second moment of two colourings with fixed overlap. In Section 4.2 we divide the set of overlaps into smaller subsets more convenient to work with and prove that it suffices to consider overlaps near some canonical overlap corresponding to the barycentre of the set of all overlaps. Subsequently, in Section 4.3 we show that we can further reduce the number of overlaps contributing to the second moment and finally in Section 4.4 we calculate the relevant constants to complete the proof on Proposition 2.4. After that, the proof of Proposition 2.5 is carried out in the last subsection and is based on and an enhancement of results derived in [3].

4.1. Classifying the overlap

To standardize the notation, we define the *overlap matrix* $\rho(\sigma,\tau) = (\rho_{ij}(\sigma,\tau))_{i,j\in[k]}$ for two colour assignments $\sigma,\tau : [n] \to [k]$ as the doubly stochastic $k \times k$-matrix with entries

$$\rho_{ij}(\sigma,\tau) = \frac{1}{n} \cdot |\sigma^{-1}(i) \cap \tau^{-1}(j)|,$$

that is, the fraction of vertices receiving colour i in colouring σ and colour j in colouring τ. Analogously to our notation in Section 2, we let $\mathcal{B}_k(n)$ denote the set of all overlap matrices and let $\mathcal{B}_k$ denote the set of all probability measures $\rho = (\rho_{ij})_{i,j\in[k]}$ on $[k] \times [k]$. To unclutter the notation and shorten the proofs, for a $k \times k$-matrix $\rho = (\rho_{ij})$, we introduce the shorthand

$$\rho_{i\star} = \sum_{j=1}^{k} \rho_{ij}, \quad \rho_{\cdot\star} = (\rho_{i\star})_{i\in[k]}, \quad \rho_{\star j} = \sum_{i=1}^{k} \rho_{ij}, \quad \rho_{\star\cdot} = (\rho_{\star i})_{i\in[k]}.$$

With the notation from Section 2, we observe that for any $\sigma,\tau : [n] \to [k]$ we have $\rho_{\cdot\star}, \rho_{\star\cdot} \in \mathcal{A}_k(n)$. For a given graph G on $[n]$, we let $Z^{(2)}_{k,\rho}(G)$ be the number of pairs (σ,τ) of k-colourings of G whose overlap is ρ and let further $\bar\rho$ signify the $k \times k$-matrix with all entries equal to k^{-2}, the barycentre of $\mathcal{B}_k$. We will need the following elementary estimates.

Fact 4.1. *For any $k \geqslant 3$, $d' \in (0,\infty)$ and $d = 2m/n$, the following estimates are true.*

(1) *Let $\rho \in \mathcal{B}_k(n)$. Then*

$$\mathbb{E}[Z^{(2)}_{k,\rho}(\mathcal{G}(n,m))] \sim \frac{\sqrt{2\pi}n^{(1-k^2)/2}}{\prod_{i,j=1}^{k}\sqrt{2\pi\rho_{ij}}} \exp[d/2 + n\mathcal{H}(\rho) + m\ln(1 - \|\rho_{\cdot\star}\|_2^2 - \|\rho_{\star\cdot}\|_2^2 + \|\rho\|_2^2)]. \tag{4.1}$$

(2) *For any $\rho \in \mathcal{B}_k(n)$ with $\|\rho - \bar\rho\|_2^2 = o(1)$, we have*

$$\mathbb{E}[Z^{(2)}_{k,\rho}(\mathcal{G}(n,m))] \sim k^{k^2}(2\pi n)^{(1-k^2)/2} \exp[d/2 + n\mathcal{H}(\rho) + m\ln(1 - \|\rho_{\cdot\star}\|_2^2 - \|\rho_{\star\cdot}\|_2^2 + \|\rho\|_2^2)]. \tag{4.2}$$

To avoid writing down the expressions repeatedly, we introduce the function $f_2 : \mathcal{B}_k \to \mathbb{R}$ defined as

$$f_2(\rho) = \mathcal{H}(\rho) + \frac{d}{2}\ln(1 - \|\rho_{\cdot\star}\|_2^2 - \|\rho_{\star\cdot}\|_2^2 + \|\rho\|_2^2). \tag{4.3}$$

A direct consequence of Fact 4.1 is that for every $\rho \in \mathcal{B}_k(n)$ we have

$$\mathbb{E}[Z^{(2)}_{k,\rho}(\mathfrak{G}(n,m))] = \exp[nf_2(\rho) + O(\ln n)]. \tag{4.4}$$

4.2. Dividing up the hypercube

We now split up the set of overlap matrices $\mathcal{B}_k(n)$ in the following way. We introduce the set

$$\mathcal{B}_{k,\omega}(n) = \left\{\rho \in \mathcal{B}_k(n) : \rho_{i\star}, \rho_{\star i} \in \left[\frac{1}{k} - \frac{w}{\sqrt{n}}, \frac{1}{k} + \frac{w}{\sqrt{n}}\right) \text{ for all } i \in [k]\right\},$$

which corresponds to $\mathcal{A}_{k,\omega}(n)$ insofar as for $\rho \in \mathcal{B}_{k,\omega}(n)$ we have $\rho_{i\star}, \rho_{\star i} \in \mathcal{A}_{k,\omega}(n)$ for all $i \in [k]$. In a next step, we refine these sets into even smaller subsets. We recall $S_{k,\omega,v}$ from (2.1) and $\rho^{k,\omega,v,s}$ from (2.2). For $s \in S_{k,\omega,v}$ we set

$$\mathcal{B}^s_{k,\omega,v}(n) = \left\{\rho \in \mathcal{B}_{k,\omega}(n) : \rho_{i\star}, \rho_{\star i} \in \left[\rho_i^{k,\omega,v,s} - \frac{1}{v\sqrt{n}}, \rho_i^{k,\omega,v,s} + \frac{1}{v\sqrt{n}}\right) \text{ for all } i \in [k]\right\},$$

an analogue of the quantity $\mathcal{A}^s_{k,\omega,v}(n)$. Thus, for any fixed v, $\mathcal{B}_{k,\omega}(n)$ is a disjoint union of all $\mathcal{B}^s_{k,\omega,v}(n)$ for $s \in S_{k,\omega,v}$. By linearity of expectation,

$$\mathbb{E}[Z^s_{k,\omega,v}(\mathfrak{G}(n,m))^2] = \sum_{\rho \in \mathcal{B}^s_{k,\omega,v}(n)} \mathbb{E}[Z^{(2)}_{k,\rho}(\mathfrak{G}(n,m))]. \tag{4.5}$$

To prove Proposition 2.4, this is the quantity we need to determine. To proceed with the calculations, we introduce, for each $\omega, v \in N, s \in S_{k,\omega,v}$ and $\eta > 0$,

$$\mathcal{B}^s_{k,\omega,v,\eta}(n) = \{\rho \in \mathcal{B}^s_{k,\omega,v}(n) : \|\rho - \bar{\rho}\|_2 \leqslant \eta\}.$$

We are going to show that the right-hand side of (4.5) is dominated by the contributions with ρ 'close to' $\bar{\rho}$ in terms of the Euclidean norm. More precisely, for a graph G let

$$Z^{s\,(2)}_{k,\omega,v,\eta}(G) = \sum_{\rho \in \mathcal{B}^s_{k,\omega,v,\eta}(n)} Z^{(2)}_{k,\rho}(G) \quad \text{for any } \eta > 0.$$

Then the second moment argument performed in [3] fairly directly yields the following statement showing that overlap matrices that are far apart from $\bar{\rho}$ asymptotically do not contribute to the second moment.

Proposition 4.2. *Assume that $k \geqslant 3$ and $d' < 2(k-1)\ln(k-1)$. Further, let $\omega, v \in N$. Then for any fixed $\eta > 0$ and any $s \in S_{k,\omega,v}$, it holds that*

$$\mathbb{E}[Z^s_{k,\omega,v}(\mathfrak{G}(n,m))^2] \sim \mathbb{E}[Z^{s\,(2)}_{k,\omega,v,\eta}(\mathfrak{G}(n,m))].$$

□

The rest of this subsection will be dedicated to proving this proposition. We first define a function

$$\bar{f}_2 : \mathcal{B}_{k,\omega}(n) \to \mathbb{R}, \quad \rho \mapsto \mathcal{H}(\rho) + \frac{d}{2}\ln\left(1 - \frac{2}{k} + \|\rho\|_2^2\right).$$

The following lemma shows how f_2 defined in (4.3) relates to $\bar{f}_2$.

Lemma 4.3. *For $\rho = (\rho_{ij}) \in \mathcal{B}_{k,\omega}(n)$, we have*

$$\exp[nf_2(\rho)] = (1 + o(1))\ \exp[n\bar{f}_2(\rho) + O(\omega^2)].$$

https://doi.org/10.1017/S0963548318000251 Published online by Cambridge University Press

Proof. We define the function

$$\zeta(\rho) = f_2(\rho) - \bar{f}_2(\rho)$$

and derive an upper bound on $\zeta(\rho)$. By definition, for each $\rho \in \mathcal{B}_{k,\omega}(n)$ there exist $\alpha = (\alpha_i)_{i\in[k]}$ and $\beta = (\beta_j)_{j\in[k]}$ such that $\rho_{i\star} = 1/k + \alpha_i$ and $\rho_{\star j} = 1/k + \beta_j$ for all $i, j \in [k]$ with $|\alpha_i|, |\beta_j| \leqslant \omega/\sqrt{n}$. Thus,

$$f_2(\rho) = \mathcal{H}(\rho) + \frac{d}{2}\ln(1 - \|\bar{\rho}_{\cdot\star} + \alpha\|_2^2 - \|\bar{\rho}_{\star\cdot} + \beta\|_2^2 + \|\rho\|_2^2).$$

As we are only interested in the difference between f_2 and $\bar{f}_2$, we can reparametrize ζ as

$$\zeta(\alpha, \beta) = \frac{d}{2}\ln\left(\frac{1 - \|\bar{\rho}_{\cdot\star} + \alpha\|_2^2 - \|\bar{\rho}_{\star\cdot} + \beta\|_2^2 + \|\rho\|_2^2}{1 - 2/k + \|\rho\|_2^2}\right).$$

Differentiating and simplifying the expression yields

$$\frac{\partial\zeta}{\partial\alpha_i}(\alpha, \beta), \quad \frac{\partial\zeta}{\partial\beta_j}(\alpha, \beta) = O\left(\frac{\omega}{\sqrt{n}}\right) \quad \text{for all } i, j \in [k].$$

According to the fundamental theorem of calculus, it follows that

$$\max_{\rho\in\mathcal{B}_{k,\omega}(n)} |\zeta(\rho)| = \int_{-\omega/\sqrt{n}}^{\omega/\sqrt{n}} O\left(\frac{\omega}{\sqrt{n}}\right) d\alpha_1 = O\left(\frac{\omega^2}{n}\right),$$

completing the proof. □

Equation (4.4) combined with Lemma 4.3 reduces our task to studying the function $\bar{f}_2(\rho)$. For the range of d covered by Proposition 4.2, this analysis is the main technical achievement of [3], where the following statement is proved.

Lemma 4.4. *Assume that $k \geqslant 3, \omega \in N$ as well as $d' \leqslant 2(k-1)\ln(k-1)$ and $d = 2m/n$. For any $n > 0$ and any overlap matrix $\rho \in \mathcal{B}_{k,\omega}(n)$, we have*

$$\bar{f}_2(\rho) \leqslant \bar{f}_2(\bar{\rho}) - \frac{2(k-1)\ln(k-1) - d}{4(k-1)^2}(k^2\|\rho\|_2^2 - 1) + o(1). \tag{4.6}$$

Proof. For ρ such that $\sum_{i=1}^k \rho_{ij} = \sum_{i=1}^k \rho_{ji} = 1/k$, the bound (4.6) is proved in [3, Section 3]. This implies that (4.6) also holds for $\rho \in \mathcal{B}_{k,\omega}(n)$, because $\bar{f}_2$ is uniformly continuous on the compact set $\mathcal{B}_{k,\omega}(n)$. □

Proof of Proposition 4.2. Assume that k and d satisfy the assumptions of Proposition 4.2 and let $\nu \in N$ and $\eta > 0$ be any fixed number. Then, for any $\hat{\rho} \in \mathcal{B}^s_{k,\omega,\nu}(n)$, we have

$$\|\hat{\rho} - \bar{\rho}\|_2 = O\left(\frac{\omega}{\sqrt{n}}\right).$$

Consequently, we obtain with (4.4) that

$$\sum_{\substack{\rho\in\mathcal{B}^s_{k,\omega,\nu}(n)\\ \|\rho-\bar{\rho}\|_2\leqslant\eta}} \mathbb{E}[Z^{(2)}_{k,\rho}(\mathfrak{G}(n,m))] \geqslant \mathbb{E}[Z^{(2)}_{k,\hat{\rho}}(\mathfrak{G}(n,m))] \geqslant \exp[nf_2(\bar{\rho}) + O(\ln n)]. \tag{4.7}$$

https://doi.org/10.1017/S0963548318000251 Published online by Cambridge University Press

On the other hand, the function $\mathcal{B} \to \mathbb{R}$, $\rho \to k^2\|\rho\|_2$ is smooth, strictly convex and attains its global minimum of 1 at $\rho = \bar{\rho}$. Consequently, there exist $(c_k)_k > 0$ such that if $\|\rho - \bar{\rho}\|_2 > \eta$, then $(k^2\|\rho\|_2 - 1) \geqslant c_k$. Hence, Fact 4.1, Lemma 4.3 and Lemma 4.4 yield

$$\sum_{\substack{\rho \in \mathcal{B}^s_{k,\omega,v}(n) \\ \|\rho-\bar{\rho}\|_2 > \eta}} \mathbb{E}[Z^{(2)}_{k,\rho}(\mathfrak{G}(n,m))] \leqslant \exp[nf_2(\bar{\rho}) - nc_k d_k + o(n)], \tag{4.8}$$

$$\text{where } d_k = \frac{2(k-1)\ln(k-1) - d}{4(k-1)^2} > 0.$$

Combining (4.8) and (4.7), we conclude that

$$\mathbb{E}[Z^s_{k,\omega,v}(\mathfrak{G}(n,m))^2] \sim \mathbb{E}[Z^{s\,(2)}_{k,\omega,v,\eta}(\mathfrak{G}(n,m))],$$

thereby completing the proof of Proposition 4.2. □

4.3. Reducing the number of overlaps

Having reduced our task to studying overlaps ρ such that $\|\rho - \bar{\rho}\|_2 \leqslant \eta$ for a small but fixed $\eta > 0$, in this section we are going to argue that, in fact, it suffices to consider ρ such that $\|\rho - \bar{\rho}\|_2 \leqslant n^{-3/8}$ (where the constant 3/8 is somewhat arbitrary; any number smaller than 1/2 would do). More precisely, we have the following result.

Proposition 4.5. *Assume that $k \geqslant 3$ and that $d' < d_{k,\mathrm{cond}}$. Let $v, \omega \in N$ and $s \in S_{k,\omega,v}$. There exists a number $\eta_0 = \eta_0(d', k)$ such that for any $0 < \eta < \eta_0$ we have*

$$\mathbb{E}[Z^{s\,(2)}_{k,\omega,v,\eta}(\mathfrak{G}(n,m))] \sim \mathbb{E}[Z^{s\,(2)}_{k,\omega,v,n^{-3/8}}(\mathfrak{G}(n,m))]. \qquad \square$$

The key to proving this proposition is the following lemma. It specifies the expected number of pairs of solutions in the cases where the overlap matrices $\rho \in \mathcal{B}^s_{k,\omega,v}(n)$ satisfy $\|\rho - \bar{\rho}\|_2 \leqslant n^{-3/8}$ or $\|\rho - \bar{\rho}\|_2 \in (n^{-3/8}, \eta)$.

Lemma 4.6. *Let $k \geqslant 3, d' < (k-1)^2$ and $d = 2m/n$. Set*

$$C_n(d,k) = \exp[d/2] k^{k^2} (2\pi n)^{(1-k^2)/2} \quad \text{and} \quad D(d,k) = k^2\left(1 - \frac{d}{(k-1)^2}\right). \tag{4.9}$$

- *If $\rho \in \mathcal{B}^s_{k,\omega,v,\eta}(n)$ satisfies $\|\rho - \bar{\rho}\|_2 \leqslant n^{-3/8}$, then*

$$\mathbb{E}[Z^{(2)}_{k,\rho}(\mathfrak{G}(n,m))] \sim C_n(d,k) \exp\left[2nf_1(\rho^\star) - n\frac{D(d,k)}{2}\|\rho - \bar{\rho}\|_2^2\right]. \tag{4.10}$$

- *There exist numbers $\eta = \eta(d,k) > 0$ and $A = A(d,k) > 0$ such that if $\rho \in \mathcal{B}^s_{k,\omega,v,\eta}(n)$ satisfies $\|\rho - \bar{\rho}\|_2 \in (n^{-3/8}, \eta)$, then*

$$\mathbb{E}[Z^{(2)}_{k,\rho}(\mathfrak{G}(n,m))] \leqslant \exp[2nf_1(\rho^\star) - An^{1/4}]. \tag{4.11}$$

Proof. As Fact 4.1 yields

$$\mathbb{E}[Z^{(2)}_{k,\rho}(\mathcal{G}(n,m))] \sim C_n(d,k)\exp[nf_2(\rho)],$$

https://doi.org/10.1017/S0963548318000251 Published online by Cambridge University Press

we have to analyse f_2. Expanding this function around $\bar{\rho}$ yields

$$f_2(\rho) = f_2(\bar{\rho}) - \frac{D(d,k)}{2}\|\rho - \bar{\rho}\|_2^2 + O(\|\rho - \bar{\rho}\|_2^3). \tag{4.12}$$

Consequently, for $\|\rho - \bar{\rho}\|_2 \leqslant n^{-3/8}$,

$$\exp[nf_2(\rho)] = \exp\left[nf_2(\bar{\rho}) - n\frac{D(d,k)}{2}\|\rho - \bar{\rho}\|_2^2 + O(n^{-1/8})\right].$$

As f_2 satisfies $f_2(\bar{\rho}) = 2f_1(\rho^\star)$, the statement in (4.10) follows.

To prove (4.11), we observe that similarly to (4.12) and because f_2 is smooth in a neighbourhood of $\bar{\rho}$, there exist $\eta > 0$ and $A > 0$ such that for $\|\rho - \bar{\rho}\|_2 \leqslant \eta$,

$$f_2(\rho) \leqslant f_2(\bar{\rho}) - A\|\rho - \bar{\rho}\|_2^2.$$

Hence, if $\|\rho - \bar{\rho}\|_2 \in (n^{-3/8}, \eta)$, then

$$\mathbb{E}[Z^{(2)}_{k,\rho}(\mathfrak{G}(n,m))] = O(n^{(1-k^2)/2})\exp[nf_2(\rho)] \leqslant \exp[2nf_1(\rho^\star) - An^{1/4}],$$

as claimed. □

Proof of Proposition 4.5. We fix $s \in S_{k,\omega,v}$. Further, we fix $\eta > 0$ and $A > 0$ as given by Lemma 4.6. For each $\hat{\rho} \in \mathcal{B}^s_{k,\omega,v,\eta}(n)$, we have

$$\|\hat{\rho} - \bar{\rho}\|_2 = O\left(\frac{\omega}{\sqrt{n}}\right),$$

and obtain from the first part of Lemma 4.6 that

$$\mathbb{E}[Z^{s\,(2)}_{k,\omega,v,n^{-3/8}}(\mathfrak{G}(n,m))] \geqslant \mathbb{E}[Z^{(2)}_{k,\rho_0}(\mathfrak{G}(n,m))] = (1 + o(1))\; C_n(d,k)\exp[2nf_1(\rho^\star) + O(\omega^2)]. \tag{4.13}$$

On the other hand, because $|\mathcal{B}^s_{k,\omega,v,\eta}(n)|$ is bounded by a polynomial in n, the second part of Lemma 4.6 yields

$$\sum_{\substack{\rho\in\mathcal{B}^s_{k,\omega,v,\eta}(n)\\ \|\rho-\bar{\rho}\|_2>n^{-3/8}}} \mathbb{E}[Z^{(2)}_{k,\rho}(\mathfrak{G}(n,m))] \leqslant \exp[2nf_1(\rho^\star) - An^{1/6} + O(\ln n)]. \tag{4.14}$$

Combining (4.13) and (4.14), we obtain

$$\mathbb{E}[Z^{s\,(2)}_{k,\omega,v,\eta}(\mathfrak{G}(n,m))] \sim \sum_{\rho\in\mathcal{B}^s_{k,\omega,v,n^{-3/8}}(n)} \mathbb{E}[Z^{(2)}_{k,\rho}(\mathfrak{G}(n,m))] \sim \mathbb{E}[Z^{s\,(2)}_{k,\omega,v,n^{-3/8}}(\mathfrak{G}(n,m))],$$

as claimed. □

4.4. Calculating the constant

In this subsection we compute the contribution of the overlap matrices $\rho \in \mathcal{B}^s_{k,\omega,v,n^{-3/8}}(n)$, in order to give a very precise expression for the second moment. At the end of this subsection, the proof of Proposition 2.4 is finally completed.

https://doi.org/10.1017/S0963548318000251 Published online by Cambridge University Press

Proposition 4.7. *Assume that $k \geqslant 3, \omega, v \in N, d' < (k-1)^2$ and $d = 2m/n$. Let $s \in S_{k,\omega,v}$. Then*

$$\mathbb{E}[Z^{s\,(2)}_{k,\omega,v,n^{-3/8}}(\mathfrak{G}(n,m))]$$
$$\sim_v (|\mathcal{A}_{k,\omega}(n)|(2\pi n)^{(1-k)/2}\, k^{k/2} \exp[nf_1(\rho^{k,\omega,v,s})])^2 \exp[d/2]\left(1 - \frac{d}{(k-1)^2}\right)^{-(k-1)^2/2}.$$

□

The rest of this subsection will be dedicated to proving this proposition. We first show that in each region of the hypercube we can approximate f_2 by a function where the marginals are set to those of the centre of this region as defined in (2.2). More formally, let $f_2^s : \mathcal{B}_k \to \mathbb{R}$ be defined as

$$f_2^s : \rho \mapsto \mathcal{H}(\rho) + \frac{d}{2}\ln(1 - 2\|\rho^{k,\omega,v,s}\|_2^2 + \|\rho\|_2^2).$$

Then the following is true.

Lemma 4.8. *Let $k \geqslant 3, \omega, v \in N$ and $C_n(d,k)$ as in (4.9). Then for $\rho \in \mathcal{B}^s_{k,\omega,v,n^{-3/8}}(n)$ it holds that*

$$\mathbb{E}[Z^{(2)}_{k,\rho}(\mathfrak{G}(n,m))] = (1+o(1))\; C_n(d,k)\exp\left[nf_2^s(\rho) + O\left(\frac{\omega}{v}\right)\right].$$

Proof. Equation (4.2) of Fact 4.1 yields that

$$\mathbb{E}[Z^{(2)}_{k,\rho}(\mathfrak{G}(n,m))] \sim C_n(d,k)\exp[nf_2(\rho)]. \tag{4.15}$$

For $s \in S_{k,\omega,v}$, we define the function

$$\zeta^s(\rho) = f_2(\rho) - f_2^s(\rho).$$

To derive an upper bound on $\zeta^s(\rho)$ for all values $\rho \in \mathcal{B}^s_{k,\omega,v,n^{-3/8}}(n)$, we first observe that there exist $\alpha = (\alpha_i)_{i\in[k]}$ and $\beta = (\beta_j)_{j\in[k]}$ such that the function f_2 can be expressed by setting $\rho_{i\star} = \rho_i^{k,\omega,v,s} + \alpha_i$ and $\rho_{\star j} = \rho_j^{k,\omega,v,s} + \beta_j$ for all $i,j \in [k]$ with $|\alpha_i|, |\beta_j| \leqslant 1/(v\sqrt{n})$. Thus,

$$f_2 : \rho \mapsto \mathcal{H}(\rho) + \frac{d}{2}\ln(1 - \|\rho^{k,\omega,v,s} + \alpha\|_2^2 - \|\rho^{k,\omega,v,s} + \beta\|_2^2 + \|\rho\|_2^2).$$

As we are only interested in the difference between f_2 and f_2^s, we can reparametrize ζ^s as

$$\zeta^s(\alpha,\beta) = \frac{d}{2}\ln\left(\frac{1 - \|\rho^{k,\omega,v,s} + \alpha\|_2^2 - \|\rho^{k,\omega,v,s} + \beta\|_2^2 + \|\rho\|_2^2}{1 - 2\|\rho^{k,\omega,v,s}\|_2^2 + \|\rho\|_2^2}\right).$$

Differentiating and simplifying the expression yields

$$\frac{\partial\zeta^s}{\partial\alpha_i}(\alpha,\beta),\quad \frac{\partial\zeta^s}{\partial\beta_j}(\alpha,\beta) = O\left(\frac{\omega}{\sqrt{n}}\right) \quad \text{for all } i,j \in [k].$$

https://doi.org/10.1017/S0963548318000251 Published online by Cambridge University Press

According to the fundamental theorem of calculus, it follows for every $s \in S_{k,\omega,v}$ that

$$\max_{\rho \in \mathcal{B}^s_{k,\omega,v,n^{-3/8}}(n)} |\zeta^s(\rho)| = \int_{-(v\sqrt{n})^{-1}}^{(v\sqrt{n})^{-1}} O\left(\frac{\omega}{\sqrt{n}}\right) d\alpha_1 = O\left(\frac{\omega}{nv}\right).$$

Combining this with (4.15) yields the assertion. □

In due course we are going to need the set of matrices with coefficients in $n^{-1}\mathbb{Z}$ whose lines and columns sum to zero:

$$\mathcal{E}_n = \left\{ (\epsilon_{i,j})_{\substack{1 \leqslant i \leqslant k,\\ 1 \leqslant j \leqslant k}} \; \forall i,j \in [k],\ \epsilon_{i,j} \in \frac{1}{n}\mathbb{Z},\ \ \forall j \in [k],\ \sum_{i=1}^{k} \epsilon_{ij} = \sum_{i=1}^{k} \epsilon_{ji} = 0 \right\}. \tag{4.16}$$

The following result regards Gaussian summations over matrices in $\mathcal{E}_n$.

Lemma 4.9. *Let $k \geqslant 2$, $d' < (k-1)^2$ and $D > 0$ be fixed. Then*

$$\sum_{\epsilon \in \mathcal{E}_n} \exp\left[-n\frac{D}{2} \|\epsilon\|_2^2 + o(n^{1/2}) \|\epsilon\|_2 \right] \sim (\sqrt{2\pi n})^{(k-1)^2} D^{-(k-1)^2/2} k^{-(k-1)}.$$

□

Lemma 4.9 and its proof are very similar to an argument used in [18, Section 3]. In fact, Lemma 4.9 follows from the next result.

Lemma 4.10 ([18, Lemma 6 (b) and 7 (c)]). *There is a $(k-1)^2 \times (k-1)^2$-matrix*

$$\mathcal{H} = (\mathcal{H}_{(i,j),(k,l)})_{i,j,k,l \in [k-1]}$$

such that for any $\varepsilon = (\varepsilon_{ij})_{i,j \in [k]} \in \mathcal{E}_n$ we have

$$\sum_{i,j,i',j' \in [k-1]} \mathcal{H}_{(i,j),(i',j')} \varepsilon_{ij} \varepsilon_{i'j'} = \|\varepsilon\|_2^2.$$

This matrix $\mathcal{H}$ is positive definite and $\det \mathcal{H} = k^{2(k-1)}$.

Now we are ready to prove Proposition 4.7.

Proof of Proposition 4.7. Lemma 4.8 states that for every $\rho \in \mathcal{B}^s_{k,\omega,v,n^{-3/8}}(n)$ we have

$$\mathbb{E}[Z^{(2)}_{k,\rho}(\mathfrak{G}(n,m))] = (1+o(1))\ C_n(d,k) \exp\left[n f_2^s(\rho) + O\left(\frac{\omega}{v}\right)\right]. \tag{4.17}$$

Thus, all we have to do is analyse the function f_2^s for $s \in S_{k,\omega,v}$. To this end, we expand $f_2^s(\rho)$ around $\rho = \rho^s$ where $\rho^s = (\rho^s_{ij})_{i,j}$ with $\rho_{ij} = \rho_i^{k,\omega,v,s} \cdot \rho_j^{k,\omega,v,s}$. Then with $D(d,k)$ as defined in (4.9) we have

$$f_2^s(\rho) = f_2^s(\rho^s) + \Theta\left(\frac{\omega}{n}\right) \|\rho - \rho^s\|_2 - \frac{D(d,k)}{2} \|\rho - \rho^s\|_2^2 + o(n^{-1}). \tag{4.18}$$

https://doi.org/10.1017/S0963548318000251 Published online by Cambridge University Press

Combining (4.18) with (4.17), we find that

$$\mathbb{E}[Z_{k,\rho}^{(2)}(\mathfrak{G}(n,m))] \tag{4.19}$$
$$= (1+o(1))\; C_n(d,k)\exp\left[nf_2^s(\rho^s) + \Theta(\omega)\|\rho-\rho^s\|_2 - n\frac{D(d,k)}{2}\|\rho-\rho^s\|_2^2 + O\left(\frac{\omega}{v}\right)\right].$$

For two vectors of 'marginals' $\rho^0,\rho^1 \in \mathcal{B}^s_{k,\omega,v}(n)$, we introduce the set of overlap matrices

$$\mathcal{B}^s_{k,\omega,v,n^{-3/8}}(n,\rho^0,\rho^1) = \{\rho \in \mathcal{B}^s_{k,\omega,v,n^{-3/8}}(n) : \rho_{\cdot\star} = \rho^0, \rho_{\star\cdot} = \rho^1\}.$$

and observe that with this definition we have

$$\mathbb{E}[Z^{s\,(2)}_{k,\omega,v,n^{-3/8}}(\mathcal{G}(n,m))] = \sum_{\rho^0,\rho^1\in\mathcal{B}^s_{k,\omega,v}(n)}\; \sum_{\rho\in\mathcal{B}^s_{k,\omega,v,n^{-3/8}}(n,\rho^0,\rho^1)} \mathbb{E}[Z_{k,\rho}^{(2)}(\mathfrak{G}(n,m))]. \tag{4.20}$$

In particular, the set $\mathcal{B}^s_{k,\omega,v,n^{-3/8}}(n,\rho^0,\rho^1)$ contains the 'product' overlap $\rho^0\otimes\rho^1$ defined by $(\rho^0\otimes\rho^1)_{ij} = \rho^0_i\rho^1_j$ for $i,j\in[k]$. To proceed, we fix two colour densities $\rho^0,\rho^1 \in \mathcal{B}^s_{k,\omega,v}(n)$ and simplify the notation by writing

$$\widehat{\mathcal{B}} = \mathcal{B}^s_{k,\omega,v,n^{-3/8}}(n,\rho^0,\rho^1), \quad \widehat{\rho} = \rho^0\otimes\rho^1.$$

Thus, the inner sum from (4.20) simplifies to

$$\mathcal{S}_1 = \sum_{\rho\in\widehat{\mathcal{B}}} \mathbb{E}[Z_{k,\rho}^{(2)}(\mathfrak{G}(n,m))].$$

and we are going to evaluate this quantity. We observe that with $\mathcal{E}_n$ as defined in (4.16), for each $\rho\in\widehat{\mathcal{B}}$ we can find $\varepsilon\in\mathcal{E}_n$ such that

$$\rho = \widehat{\rho} + \varepsilon.$$

Hence, this gives $\|\rho-\rho^s\|_2 = \|\widehat{\rho}+\varepsilon-\rho^s\|_2$, and the triangle inequality yields

$$\|\varepsilon\|_2 - \|\widehat{\rho}-\rho^s\|_2 \leqslant \|\widehat{\rho}+\varepsilon-\rho^s\|_2 \leqslant \|\varepsilon\|_2 + \|\widehat{\rho}-\rho^s\|_2.$$

By definition of $\widehat{\rho}$ and ρ^s, we have $\|\widehat{\rho}-\rho^s\|_2 \leqslant 1/(v\sqrt{n})$ and consequently

$$\|\rho-\rho^s\|_2 = \|\boldsymbol{\varepsilon}\|_2 + O\left(\frac{1}{v\sqrt{n}}\right). \tag{4.21}$$

Observing that $f_2^s(\rho^s) = (f_1(\rho^{k,\omega,v,s}))^2$ and inserting (4.21) into (4.19) while taking first $n\to\infty$ and afterwards $v\to\infty$, we obtain

$$\mathcal{S}_1 \sim_v C_n(d,k)\exp[2nf_1^s(\rho^{k,\omega,v,s})]\sum_{\rho\in\widehat{\mathcal{B}}}\exp\left[-n\frac{D(d,k)}{2}\|\varepsilon\|_2^2 + o(n^{1/2})\|\varepsilon\|_2\right]. \tag{4.22}$$

To apply Lemma 4.9, we have to relate $\rho\in\widehat{\mathcal{B}}$ to $\varepsilon\in\mathcal{E}_n$. From the definitions we obtain

$$\{\widehat{\rho}+\varepsilon : \varepsilon\in\mathcal{E}_n, \|\varepsilon\|_2 \leqslant n^{-3/8}/2\} \subset \{\rho\in\widehat{\mathcal{B}}\} \subset \{\widehat{\rho}+\varepsilon : \varepsilon\in\mathcal{E}_n\}.$$

https://doi.org/10.1017/S0963548318000251 Published online by Cambridge University Press

We show that the contribution of $\varepsilon \in \mathcal{E}_n$ with $\|\varepsilon\|_2 > n^{-3/8}/2$ is negligible:

$$\begin{aligned}\mathcal{S}_2 &= C_n(d,k)\exp[2nf_1^s(\rho^{k,\omega,v,s})] \sum_{\substack{\epsilon\in\mathcal{S}_n\\ \|\epsilon\|_2>n^{-3/8}/2}} \exp\left[-n\frac{D(d,k)}{2}\|\epsilon\|_2^2(1+o(1))\right]\\ &= C_n(d,k)\exp[2nf_1^s(\rho^{k,\omega,v,s})] \sum_{\substack{l\in\mathbb{Z}/n\\ l>n^{-3/8}/2}} \sum_{\substack{\epsilon\in S_n\\ \|\epsilon\|_2=l}} \exp\left[-nl^2\frac{D(d,k)}{2}(1+o(1))\right]\\ &= C_n(d,k)\exp[2nf_1^s(\rho^{k,\omega,v,s})]O(n^{k^2})\exp\left[-\frac{D(d,k)}{2}n^{1/4}\right].\end{aligned}$$

Consequently, (4.22) yields $\Sigma_2 = o(\Sigma_1)$. Thus, we obtain from Lemma 4.9 that

$$\begin{aligned}\mathcal{S}_1 &\sim_v C_n(d,k)\exp[2nf_1^s(\rho^{k,\omega,v,s})]\sum_{\rho\in\widehat{\mathcal{B}}}\exp\left[-n\frac{D(d,k)}{2}\|\epsilon\|_2^2 + o(n^{1/2})\|\epsilon\|_2\right].\\ &\sim_v C_n(d,k)\exp[2nf_1^s(\rho^{k,\omega,v,s})](\sqrt{2\pi n})^{(k-1)^2}k^{-k(k-1)}\left(1-\frac{d}{(k-1)^2}\right)^{-(k-1)^2/2}. \qquad (4.23)\end{aligned}$$

In particular, the last expression is independent of the choice of the vectors ρ^0, ρ^1 that defined $\widehat{\mathcal{B}}$. Therefore, substituting (4.23) in the decomposition (4.20) completes the proof of Proposition 4.7. □

Proof of Proposition 2.4. The first part of the proposition is immediately obtained by combining Lemma 3.3 with Propositions 4.2, 4.5 and 4.7. The equation follows by applying the definitions ol λ_l and δ_l. □

4.5. Up to the condensation threshold

In this last subsection we prove Proposition 2.5. In the regime $2(k-1)\ln(k-1) \leqslant d' < d_{k,\mathrm{cond}}$ for $k \geqslant k_0$ for some big constant k_0, we consider random variables $\widetilde{Z}^s_{k,\omega,v}$ instead of $Z^s_{k,\omega,v}$. To prove the proposition we show the following result by adapting our setting in a way that we can apply the second moments argument from [13] and [7].

Proposition 4.11. *Let $\omega, v \in N$. There is a constant $k_0 > 3$ such that for $k \geqslant k_0$ and $2(k-1)\ln(k-1) \leqslant d' < d_{k,\mathrm{cond}}$, the following is true. For each $s \in S_{k,\omega,v}$, there exists an integer-valued random variable $0 \leqslant \widetilde{Z}^s_{k,\omega,v} \leqslant Z^s_{k,\omega,v}$ that satisfies*

$$\mathbb{E}[\widetilde{Z}^s_{k,\omega,v}(\mathfrak{G}(n,m))] \sim \mathbb{E}[Z^s_{k,\omega,v}(\mathfrak{G}(n,m))],$$

and such that for any fixed $\eta > 0$ we have

$$\mathbb{E}[\widetilde{Z}^s_{k,\omega,v}(\mathfrak{G}(n,m))^2] \leqslant (1+o(1))\ \mathbb{E}[Z^{s\,(2)}_{k,\omega,v,\eta}(\mathfrak{G}(n,m))].$$

□

In this section we work with the Erdős–Rényi random graph model $G(n,p)$, which is a random graph on $[n]$ vertices where every possible edge is present with probability $p = d/n$ independently. We further assume from now on that k divides n.

The use of results from [7, 13] is complicated by the fact that we are dealing with (ω, n)-balanced k-colourings that allow a larger discrepancy between the colour classes than [7, 13], where balanced colourings are defined such that in each colour class only a deviation of at most $\sqrt{n}$ from the typical value n/k is allowed. To circumvent this problem, we introduce the following.

Choose a map $\sigma : [n] \to [k]$ uniformly at random and generate a graph $G(n, p', \sigma)$ on $[n]$ by connecting any two vertices $v, w \in [n]$ such that $\sigma(v) \neq \sigma(w)$ with probability $p' = dk/(n(k-1))$ independently.

Given σ and $G(n, p', \sigma)$, we define

$$\alpha_i = |\sigma^{-1}(i) - n/k| \quad \text{for } i \in [k]$$

and let $\alpha = \max_{i \in [k]} \alpha_i$. Thus, by definition $\alpha \leqslant \omega\sqrt{n}$. We set $n' = n + k\lceil\alpha\rceil$. Further, we let

$$\beta_i = |\sigma^{-1}(i) - (n + k\lceil\alpha\rceil)/k| \quad \text{for } i \in [k].$$

We then construct a coloured graph $G'_{n',p',\sigma'}$ from $G(n, p', \sigma)$ in the following way.

- Add $k\lceil\alpha\rceil$ vertices to $G(n, p)$ and denote them by $n + 1, n + 2, \ldots, n + k\lceil\alpha\rceil$.
- Define a colouring $\sigma' : [n'] \to [k]$ by $\sigma'(i) = \sigma(i)$ for $i \in [n]$, $\sigma(i) = 1$ for $i \in n + 1, \ldots, n + \beta_1$ and $\sigma(i) = j$ for $j \in \{2, \ldots, k\}$ and $i \in n + \beta_{j-1} + 1, \ldots, n + \beta_j$.
- Add each possible edge (i, j) with $\sigma'(i) \neq \sigma'(j)$ involving a vertex $i \in \{n + 1, \ldots, n + k\lceil\alpha\rceil\}$ with probability $p' = dk/(n(k-1))$.

We call a colouring $\tau : [n] \to [k]$ of a graph G on $[n]$ *perfectly balanced* if $|\tau^{-1}(i)| = |\tau^{-1}(j)|$ for all $i, j \in [k]$ and we denote the set of all such perfectly balanced colourings by $\widetilde{\mathcal{B}}_k(n)$. Then the following holds by construction.

Fact 4.12. *$G'_{n',p',\sigma'}$ has the same distribution as $G(n', p', \tau)$ conditioned on the event that $\tau : [n'] \to k$ is perfectly balanced.*

Let $G''_{n,p',\sigma'|[n]}$ denote the graph obtained from $G'_{n',p',\sigma'}$ by deleting the vertices $n + 1, \ldots, n + k\lceil\alpha\rceil$ and the incident edges.

Fact 4.13. *$G''_{n,p',\sigma'|n}$ has the same distribution as $G(n, p', \tau)$ conditioned on the event that τ is (ω, n)-balanced.*

To proceed, we adopt the following notation from [13]. Let $\rho \in \mathcal{B}_k$ be called *s-stable* if it has precisely s entries bigger than $0.51/k$. Further, let $\bar{\mathcal{B}}_k$ be the set of all $\rho \in \mathcal{B}_k$ such that

$$\sum_{j=1}^{k} \rho_{ij} = \sum_{j=1}^{k} \rho_{ji} = 1/k \quad \text{for all } i \in [k].$$

Then any $\rho \in \bar{\mathcal{B}}_k$ is s-stable for some $s \in \{0, 1, \ldots, k\}$. In addition, let $\kappa = \ln^{20} k/k$ and let us call $\rho \in \mathcal{B}_k$ *separable* if $k\rho_{ij} \notin (0.51, 1 - \kappa)$ for all $i, j \in [k]$. A k-colouring σ of a graph G on $[n]$ is called *separable* if, for any other k-colouring τ of G, the overlap matrix $\rho(\sigma, \tau)$ is separable. We have the following result.

https://doi.org/10.1017/S0963548318000251 Published online by Cambridge University Press

Lemma 4.14. *Let $s \in S_{k,\omega,v}$. There exists $k_0 > 0$ such that for all $k > k_0$ and all d' such that $2(k-1)\ln(k-1) \leqslant d' \leqslant (2k-1)\ln k$, the following is true. Let $\widetilde{Z}^s_{k,\omega,v}(\mathfrak{G}(n,m))$ denote the number of (ω,n)-balanced k-colourings of $\mathfrak{G}(n,m)$ that fail to be separable. Then*

$$\mathbb{E}[\widetilde{Z}^s_{k,\omega,v}(\mathfrak{G}(n,m))] = o(\mathbb{E}[Z^s_{k,\omega,v}(\mathfrak{G}(n,m))]).$$

To prove this lemma, we combine Fact 4.12 with [13, Lemma 3.3]. This yields the following.†

Lemma 4.15 ([13]). *There is $k_0 > 0$ such that for all $k \geqslant k_0$ and all d' with $2(k-1)\ln(k-1) \leqslant d' \leqslant (2k-1)\ln k$, each $\tau \in \widetilde{\mathcal{B}}_k(n')$ is separable in $G'_{n',p',\tau}$ w.h.p.*

Proof of Lemma 4.14. Choose a map $\sigma : [n] \to [k]$ uniformly at random and generate a graph $G(n,p',\sigma)$ on $[n]$ by connecting any two vertices $v,w \in [n]$ such that $\sigma(v) \neq \sigma(w)$ with probability p' independently. Construct $G'_{n',p',\sigma'}$ from $G(n,p',\sigma)$ in the way defined above. Then $\sigma' \in \widetilde{\mathcal{B}}_k(n)$. By Lemma 4.15, σ' is separable in $G'_{n',p',\sigma'}$ w.h.p. Thus, σ is separable in $G''_{n,p',\sigma'|n}$ if we define separability using $\kappa' = (\ln^{21} k)/k$. By choosing k_0 sufficiently large and applying Fact 4.13, the assertion follows. □

For the next ingredient of the proof of Proposition 4.11, we need the following definition. For a graph G on $[n]$ and a k-colouring σ of G, we let $\mathcal{C}(G,\sigma)$ be the set of all $\tau \in \mathcal{B}_k$ that are k-colourings of G such that $\rho(\sigma,\tau)$ is k-stable.

Lemma 4.16. *Let $s \in S_{k,\omega,v}$. There is $k_0 > 0$ such that for all $k > k_0$ and all d' such that $(2k-1)\ln k - 2 \leqslant d' \leqslant d_{k,\mathrm{cond}}$, the following is true. There exists an $\varepsilon > 0$ such that if $\widetilde{Z}^s_{k,\omega,v}(\mathfrak{G}(n,m))$ denotes the number of (ω,n)-balanced k-colourings σ of $\mathfrak{G}(n,m)$ satisfying*

$$|\mathcal{C}(\mathfrak{G}(n,m),\sigma)| > \mathbb{E}[Z^s_{k,\omega,v}(\mathfrak{G}(n,m))]/\exp[\varepsilon n],$$

then

$$\mathbb{E}[\widetilde{Z}^s_{k,\omega,v}(\mathfrak{G}(n,m))] = o(\mathbb{E}[Z^s_{k,\omega,v}(\mathfrak{G}(n,m))]). \qquad \square$$

To prove this lemma, we combine 4.12 with [7, Corollary 1.1] and obtain the following.

Lemma 4.17 ([7]). *Let $s \in S_{k,\omega,v}$. There exists $k_0 > 0$ such that for all $k > k_0$ and all d' such that $(2k-1)\ln k - 2 \leqslant d' \leqslant d_{k,\mathrm{cond}}$, the following is true. Let $\tau \in \widetilde{\mathcal{B}}_k(n')$ be a perfectly balanced colour assignment. Then there exists $\varepsilon > 0$ such that if $\widetilde{Z}^s_{k,\omega,v}(G'_{n',p',\tau})$ denotes the number of (ω,n)-balanced k-colourings τ of $G'_{n',p',\tau}$ satisfying*

$$|\mathcal{C}(G'_{n',p',\tau},\tau)| > \mathbb{E}[Z^s_{k,\omega,v}(G'_{n',p',\tau})]/\exp[\varepsilon n],$$

† As a matter of fact, Lemma 3.2 in [13] also holds for densities $2(k-1)\ln(k-1) \leqslant d' \leqslant 2(k-1)\ln k - 2$, as all steps in the proof are also valid in this regime.

https://doi.org/10.1017/S0963548318000251 Published online by Cambridge University Press

then

$$\mathbb{E}[\widetilde{Z}^{s}_{k,\omega,v}(G'_{n',p',\tau})] = o(\mathbb{E}[Z^{s}_{k,\omega,v}(G'_{n',p',\tau})]).$$

Proof of Lemma 4.16. Choose a map $\sigma : [n] \to [k]$ uniformly at random and generate a graph $G(n,p',\sigma)$ on $[n]$ by connecting any two vertices $v,w \in [n]$ such that $\sigma(v) \neq \sigma(w)$ with probability p' independently. Construct $G'_{n',p',\sigma'}$ from $G(n,p',\sigma)$ in the way defined above. To construct $G''_{n,p',\sigma'_{|[n]}}$ from $G'_{n',p',\sigma'}$, we have to delete $O(\sqrt{n})$ many vertices. By [7, Section 6], for each of these vertices v we can bound the logarithm of the number of colourings that emerge when deleting v by $O(\ln n)$. Thus,

$$\ln|\mathcal{C}(G''_{n,p',\sigma'_{|[n]}},\sigma'_{|[n]})| = \ln|\mathcal{C}(G'_{n',p',\sigma'},\sigma')| + O(\sqrt{n}\ln n) = \ln|\mathcal{C}(G'_{n',p',\sigma'},\sigma')| + o(n). \quad (4.24)$$

Then Lemma 4.16 follows by combining Lemma 4.17 with (4.24) and Fact 4.13. □

To complete the proof, we have to analyse the function f_2 defined in (4.3), as we know from (4.4) that

$$\mathbb{E}[Z^{(2)}_{k,\rho}(\mathfrak{G}(n,m))] = \exp[nf_2(\rho) + O(\ln n)].$$

The following lemma shows that we can confine ourselves to the investigation of the function $\bar{f}_2$ defined in (4.2).

Lemma 4.18. *Let* $\lim_{n\to\infty}(\rho_n)_n = \rho_0$. *Then* $\lim_{n\to\infty} \ln \mathbb{E}[Z^{(2)}_{k,\rho_n}(\mathfrak{G}(n,m))] \leqslant \bar{f}_2(\rho_0)$.

Proof. Lemma 4.3 yields that

$$\exp[nf_2(\rho)] = (1+o(1))\ \exp[n\bar{f}_2(\rho) + O(\omega^2)].$$

Together with the uniform continuity of $\bar{f}_2$ this proves the assertion. □

We use results from [13] where an analysis of $\bar{f}_2$ was performed. The following lemma summarizes this analysis from [13, Section 4]. The same result was used in [6].

Lemma 4.19. *For any* $c > 0$, *there exists* $k_0 > 0$ *such that for all* $k > k_0$ *and all* d *such that* $(2k-1)\ln k - c \leqslant d' \leqslant (2k-1)\ln k$, *the following statements are true.*

(1) *If* $1 \leqslant s < k$, *then for all separable* s*-stable* $\rho \in \bar{\mathcal{B}}_k$ *we have* $\bar{f}_2(\rho) < \bar{f}_2(\bar{\rho})$.
(2) *If* $\rho \in \bar{\mathcal{B}}_k$ *is* 0*-stable and* $\rho \neq \bar{\rho}$, *then* $\bar{f}_2(\rho) < \bar{f}_2(\bar{\rho})$.
(3) *If* $d' = (2k-1)\ln k - 2$, *then for all separable,* k*-stable* $\rho \in \bar{\mathcal{B}}_k$ *we have* $\bar{f}_2(\rho) < \bar{f}_2(\bar{\rho})$.

Proof of Proposition 4.11. Assume that $k \geqslant k_0$ for a sufficiently large number k_0 and that $d' \geqslant 2(k-1)\ln(k-1)$. We consider two different cases.

Case 1. $d' \leqslant (2k-1)\ln k - 2$. Let $\widetilde{Z}^{s}_{k,\omega,v}$ be the number of (ω,n)-balanced separable k-colourings of $\mathfrak{G}(n,m)$. Then Lemma 4.15 implies that

$$\mathbb{E}[\widetilde{Z}^{s}_{k,\omega,v}(\mathfrak{G}(n,m))] \sim \mathbb{E}[Z^{s}_{k,\omega,v}(\mathfrak{G}(n,m))].$$

https://doi.org/10.1017/S0963548318000251 Published online by Cambridge University Press

Further, in the case that $d' = (2k-1)\ln k - 2$, the combination of the statements of Lemma 4.19 implies that $\bar{f}_2(\rho) < \bar{f}_2(\bar{\rho})$ for any separable $\rho \in \bar{\mathcal{B}}_k \setminus \{\bar{\rho}\}$. As $\bar{f}_2(\rho)$ is the sum of the concave function $\rho \mapsto \mathcal{H}(\rho)$ and the convex function $\rho \mapsto (d/2)\ln(1 - 2/k\|\rho\|_2^2)$, this implies that, in fact, for any $d' \leqslant (2k-1)\ln k - 2$ we have $\bar{f}_2(\rho) < \bar{f}_2(\bar{\rho})$ for any separable $\rho \in \bar{\mathcal{B}}_k \setminus \{\bar{\rho}\}$. Hence, the uniform continuity of $\bar{f}_2$ on $\mathcal{B}_k$ and (4.4) yield

$$\mathbb{E}[Z^s_{k,\omega,\nu}(\mathfrak{G}(n,m))^2] \leqslant (1+o(1)) \sum_{\substack{\rho \in \mathcal{B}^s_{k,\omega,\nu}(n) \\ \rho \text{ is 0-stable}}} \mathbb{E}[Z^{(2)}_{k,\rho}(\mathfrak{G}(n,m))]. \tag{4.25}$$

Additionally, as $\bar{\mathcal{B}}_k$ is a compact set, with the second statement of Lemma 4.19 it follows that for any $\eta > 0$ there exists $\varepsilon > 0$ such that

$$\max_{\substack{\rho \in \mathcal{B}^s_{k,\omega,\nu}(n) \\ \rho \text{ is 0-stable} \\ \|\rho - \bar{\rho}\|_2 > \eta}} \exp[n\bar{f}_2(\rho)] \leqslant \exp[n(\bar{f}_2(\bar{\rho}) - \varepsilon)]. \tag{4.26}$$

As on the other hand it holds that

$$\mathbb{E}[Z^{s\,(2)}_{k,\omega,\nu,\eta}(\mathfrak{G}(n,m))] \geqslant \exp[n\bar{f}_2(\bar{\rho})]/\mathrm{poly}(n), \tag{4.27}$$

combining (4.26) and (4.27) with (4.4) and the observation that $|\mathcal{B}^s_{k,\omega,\nu}(n)| \leqslant n^{k^2}$, we see that for any $\eta > 0$,

$$\sum_{\substack{\rho \in \mathcal{B}^s_{k,\omega,\nu}(n) \\ \rho \text{ is 0-stable} \\ \|\rho - \bar{\rho}\|_2 > \eta}} \mathbb{E}[Z^{(2)}_{k,\rho}(\mathfrak{G}(n,m))] \leqslant \sum_{\substack{\rho \in \mathcal{B}^s_{k,\omega,\nu}(n) \\ \rho \text{ is 0-stable} \\ \|\rho - \bar{\rho}\|_2 > \eta}} \exp[n\bar{f}_2(\rho) + O(\ln n)] = o(\mathbb{E}[Z^{s\,(2)}_{k,\omega,\nu,\eta}(\mathfrak{G}(n,m))]). \tag{4.28}$$

Case 2. $(2k-1)\ln k - 2 < d' < d_{k,\mathrm{cond}}$. For an appropriate $\varepsilon > 0$ let $\widetilde{Z}^s_{k,\omega,\nu}$ be the number of (ω, n)-balanced separable k-colourings σ of $\mathfrak{G}(n,m)$ such that

$$|\mathcal{C}(\mathfrak{G}(n,m),\sigma)| \leqslant \mathbb{E}[Z^s_{k,\omega,\nu}(\mathfrak{G}(n,m))]/\exp[\varepsilon n].$$

Then Lemmas 4.15 and 4.16 imply that

$$\mathbb{E}[\widetilde{Z}^s_{k,\omega,\nu}(\mathfrak{G}(n,m))] \sim \mathbb{E}[Z^s_{k,\omega,\nu}(\mathfrak{G}(n,m))].$$

Furthermore, the first part of Lemma 4.19 and equation (4.4) entail that (4.25) holds for this random variable $\widetilde{Z}^s_{k,\omega,\nu}$. Moreover, as in the previous case (4.26), (4.27), (4.4) and the second part of Lemma 4.19 show that (4.28) holds true for any fixed $\eta > 0$.

In either case the assertion follows by combining (4.25) and (4.28). □

Proof of Proposition 2.5. The assertion is obtained by combining Proposition 2.1 with Propositions 4.11, 4.5 and 4.7. □

Acknowledgements

I thank my PhD adviser Amin Coja-Oghlan for constant support and valuable suggestions and Samuel Hetterich for helpful discussions.

https://doi.org/10.1017/S0963548318000251 Published online by Cambridge University Press

References

[1] Achlioptas, D. and Coja-Oghlan, A. (2008) Algorithmic barriers from phase transitions. In *FOCS '08: IEEE 49th Annual IEEE Symposium on Foundations of Computer Science*, IEEE, pp. 793–802.

[2] Achlioptas, D. and Friedgut, E. (1999) A sharp threshold for k-colorability. *Random Struct. Alg.* **14** 63–70.

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

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

[5] Banks, J., Moore, C., Neeman, J. and Netrapalli, P. (2016) Information-theoretic thresholds for community detection in sparse networks. In *COLT: 29th Conference on Learning Theory*, MLR Press, pp. 383–416.

[6] Bapst, V., Coja-Oghlan, A. and Efthymiou, C. (2017) Planting colourings silently. *Combin. Probab. Comput.* **26** 338–366.

[7] Bapst, V., Coja-Oghlan, A., Hetterich, S., Raßmann, F. and Vilenchik, D. (2016) The condensation phase transition in random graph coloring. *Commun. Math. Phys.* **341** 543–606.

[8] Bapst, V., Coja-Oghlan, A. and Rassmann, F. (2016) A positive temperature phase transition in random hypergraph 2-colouring. *Ann. Appl. Probab.* **26** 1362–1406.

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

[10] Bollobás, B. (2001) *Random Graphs*, second edition, Cambridge University Press.

[11] Coja-Oghlan, A. (2013) Upper-bounding the k-colorability threshold by counting covers. *Electron. J. Combin.* **20** P32.

[12] Coja-Oghlan, A., Efthymiou, C. and Hetterich, S. (2016) On the chromatic number of random regular graphs. *J. Combin. Theory Ser. B* **116** 367–439.

[13] Coja-Oghlan, A. and Vilenchik, D. (2013) Chasing the k-colorability threshold. In *FOCS: IEEE 54th Annual Symposium on Foundations of Computer Science*, IEEE, pp. 380–389. A full version is available as arXiv:1304.1063.

[14] Coja-Oghlan, A. and Wormald, N. The number of satisfying assignments of random regular k-SAT formulas. arXiv:1611.03236

[15] Erdős, P. and Rényi, A. (1960) On the evolution of random graphs. *Magayar Tud. Akad. Mat. Kutato Int. Kozl.* **5** 17–61.

[16] Frieze, A. and Karónski, M. (2015) *Introduction to Random Graphs*, Cambridge University Press.

[17] Janson, S. (1995) Random regular graphs: Asymptotic distributions and contiguity. *Combin. Probab. Comput.* **4** 369–405.

[18] Kemkes, G., Perez-Gimenez, X. and Wormald, N. (2010) On the chromatic number of random d-regular graphs. *Adv. Math.* **223** 300–328.

[19] Krzakala, F., Montanari, A., Ricci-Tersenghi, F., Semerjian, G. and Zdeborova, L. (2007) Gibbs states and the set of solutions of random constraint satisfaction problems. *Proc. Natl Acad. Sci.* **104** 10318–10323.

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

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

[22] Matula, D. (1987) Expose-and-merge exploration and the chromatic number of a random graph. *Combinatorica* **7** 275–284.

[23] Molloy, M. (2012) The freezing threshold for k-colourings of a random graph. In *STOC: 44th Symposium on Theory of Computing*, ACM, pp. 921–930.

[24] Montanari, A., Restrepo, R. and Tetali, P. (2011) Reconstruction and clustering in random constraint satisfaction problems. *SIAM J. Discrete Math.* **25** 771–808.

[25] Moore, C. (2016) The phase transition in random regular exact cover. *Ann. Inst. Henri Poincaré* **3** 349–362.
[26] Rassmann, F. (2017) The Electronic Journal of Combinatorics **24**(3) #P3.11.
[27] Robinson, R. and Wormald, N. (1992) Almost all cubic graphs are Hamiltonian. *Random Struct. Alg.* **3** 117–125.
[28] Robinson, R. and Wormald, N. (1994) Almost all regular graphs are Hamiltonian. *Random Struct. Alg.* **5** 363–374.
[29] Shamir, E. and Spencer, J. (1987) Sharp concentration of the chromatic number of random graphs $G(n, p)$. *Combinatorica* **7** 121–129.
[30] Wormald, N. (1999) Models of random regular graphs. In *Surveys in Combinatorics*, Vol. 267 of London Mathematical Society Lecture Note Series, Cambridge University Press, pp. 239–298.