

# On the number of solutions in random hypergraph 2-colouring

Felicia Rassmann\*

Goethe University, Institute of Mathematics  
10 Robert-Mayer-Str, Frankfurt 60325, Germany

rassmann@math.uni-frankfurt.de

Submitted: Mar 25, 2016; Accepted: Jul 13, 2017; Published: Jul 28, 2017

Mathematics Subject Classifications: 05C15, 05C65, 05C80, 60C05

## Abstract

We determine the limiting distribution of the logarithm of the number of satisfying assignments in the random  $k$ -uniform hypergraph 2-colouring problem in a certain density regime for all  $k \geq 3$ . As a direct consequence we obtain that in this regime the random colouring model is contiguous with respect to the planted model, a result that helps simplifying the transfer of statements between these two models.

## 1 Introduction

### 1.1 Background and motivation

A main focus when studying random constraint satisfaction problems is on determining the expected value of the number of solutions and understanding how this number evolves when the constraint density changes. However, up to now the distribution of the number of solutions remains elusive in any of the standard examples of random constraint satisfaction problems.

In this paper we consider *random  $k$ -uniform hypergraphs*  $H_k(n, m)$  on the vertex set  $[n] = \{1, \dots, n\}$  with exactly  $m$  hyperedges each comprising of  $k$  distinct vertices - where  $k$  is a fixed number independent of  $n$  - and chosen uniformly at random from all possible subsets of  $[n]$  of size  $k$ . The *random hypergraph 2-colouring problem* is a random constraint satisfaction problem where one is interested in the number  $Z(H_k(n, m))$  of *2-colourings* (also called *solutions*) of  $H_k(n, m)$ , which are maps  $\sigma : [n] \rightarrow \{0, 1\}$  that generate no *monochromatic* edges (i.e. edges  $e$  such that  $|\sigma(e)| = 1$ ).

---

\*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 n. 278857-PTCC

In the following we only consider *sparse* random hypergraphs, meaning that  $m = O(n)$  as  $n \rightarrow \infty$ . We call the parameter  $d = km/n$  the *hyperedge density*.

As for many other random constraint satisfaction problems, there is a conjecture as to a sharp threshold for the existence of solutions in terms of the hyperedge density. The best currently known bounds on this threshold  $d_{k,\text{col}}$  are from Achlioptas and Moore [3] and Coja-Oghlan and Zdeborová [10]. Furthermore, there was a prediction for this density by statistical physicists [11, 13] suggesting that

$$d_{k,\text{col}}/k = 2^{k-1} \ln 2 - \frac{\ln 2}{2} - \frac{1}{4} + \varepsilon_k \quad \text{with } \lim_{k \rightarrow \infty} \varepsilon_k \rightarrow 0.$$

This prediction was proved by Coja-Oghlan and Panagiotou [8] for the problem of NAE- $k$ -SAT which is almost equivalent to hypergraph 2-colouring and it should be possible to transfer the result without major difficulties.

For a long period of time, the best rigorous upper and lower bounds on  $d_{k,\text{col}}$  were based on the non-constructive *first* and *second moment method* applied to the random variable  $Z$ . Achlioptas and Moore [3] proved that there is a critical density  $d_{k,\text{sec}}$  such that  $\mathbb{E}[Z^2] \leq C \cdot \mathbb{E}[Z]^2$  for some constant  $C = C(d, k) > 0$  for all  $d < d_{k,\text{sec}}$  but is violated for  $d > d_{k,\text{sec}}$ . Via the Paley-Zygmund inequality it can be established that  $d_{k,\text{sec}}$  is a non-constructive lower bound on  $d_{k,\text{col}}$ . In [3] it was shown that

$$d_{k,\text{sec}}/k = 2^{k-1} \ln 2 - \frac{\ln 2}{2} - \frac{1}{2} + \varepsilon_k, \quad \text{with } \lim_{k \rightarrow \infty} \varepsilon_k = 0.$$

## 1.2 Results

The main result in the present paper is to show that under certain conditions the number  $Z(H_k(n, m))$  of 2-colourings of the random  $k$ -uniform hypergraph is concentrated remarkably tightly and to even obtain the distribution of  $\ln Z(H_k(n, m)) - \ln \mathbb{E}[Z(H_k(n, m))]$  asymptotically in a density regime essentially up to  $d_{k,\text{sec}}$ .

As we are going to need a very neat computation of the first and second moment of the number of 2-colourings of  $Z(H_k(n, m))$ , additionally 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  $m = \lceil d'n/k \rceil$ . We note that  $d' \sim d$ , although  $d = d(n)$  might slightly vary with  $n$ , whereas  $d'$  is assumed to be fixed as  $n \rightarrow \infty$ . We also always require that  $k \geq 3$ .

**Theorem 1.** *Let  $k \geq 3$  and  $d'/k \leq 2^{k-1} \ln 2 - 2$  as well as*

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

*for  $l \geq 2$ . Further 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 the random variable*

$$W = \sum_{l=2}^{\infty} X_l \ln(1 + \delta_l) - \lambda_l \delta_l$$

satisfies  $\mathbb{E}|W| < \infty$  and  $\ln Z(H_k(n, m)) - \ln \mathbb{E}[Z(H_k(n, m))]$  converges in distribution to  $W$ .

*Remark 2.* By definition,  $W$  has an infinitely divisible distribution. It was shown in [12] that the random variable  $W' = \exp[W]$  converges almost surely and in  $L^2$  with  $\mathbb{E}[W'] = 1$  and  $\mathbb{E}[W'^2] = \exp[\sum_l \lambda_l \delta_l^2]$ . Thus, by Jensen's inequality it follows that  $\mathbb{E}[W] \leq 0$  and  $\mathbb{E}[W^2] \leq \sum_l \lambda_l \delta_l^2$ .

As a direct consequence of Theorem 1 we obtain the following.

**Corollary 3.** *Assume that  $k \geq 3$  and  $d'/k \leq 2^{k-1} \ln 2 - 2$ . Then*

$$\lim_{\omega \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}[|\ln Z(H_k(n, m)) - \ln \mathbb{E}[Z(H_k(n, m))]| \leq \omega] = 1. \quad (1.1)$$

On the other hand, for any fixed number  $\omega > 0$  we have

$$\lim_{n \rightarrow \infty} \mathbb{P}[|\ln Z(H_k(n, m)) - \ln \mathbb{E}[Z(H_k(n, m))]| \leq \omega] < 1.$$

For  $d', k$  covered by Corollary 3 we have  $\ln Z_k(G(n, m)) = \Theta(n)$  w.h.p.. Thus, it would be reasonable to expect that  $\ln Z_k(G(n, m))$  has fluctuations of order e. g.  $\sqrt{n}$ . However, the first part of Corollary 3 shows that actually  $\ln Z_k(G(n, m))$  fluctuates w.h.p. by no more than  $\omega = \omega(n)$  for *any*  $\omega(n) \rightarrow \infty$ . Moreover, the second part shows that this is best possible.

### 1.2.1 Planted model and silent planting

When proving results about random hypergraphs and investigating their properties, it turns out very useful and often essential to have the notion of *typical* 2-colourings at hand. By a typical 2-colouring of a random hypergraph we mean a 2-colouring chosen uniformly at random from all its 2-colourings. To make this formal, let  $\Lambda_{k,n,m}$  be the set of all pairs  $(H_k, \sigma)$  with  $H_k = H_k(n, m)$  and  $\sigma$  a 2-colouring of  $H_k$ . Now we define a probability distribution  $\pi_{k,n,m}^{\text{rc}}[H_k, \sigma]$  on  $\Lambda_{k,n,m}$  by letting

$$\pi_{k,n,m}^{\text{rc}}[H_k, \sigma] = \left[ Z(H_k) \binom{n}{k} \binom{k}{m} \mathbb{P}[H_k \text{ is 2-colourable}] \right]^{-1}.$$

We call this distribution the *random colouring model* or *Gibbs distribution*. It can also be described as the distribution produced by the following experiment.

**RC1** Generate a random hypergraph  $H_k = H_k(n, m)$  provided that  $Z(H_k) > 0$ .

**RC2** Choose a 2-colouring  $\sigma$  of  $H_k$  uniformly at random. The result of the experiment is  $(H_k, \sigma)$ .

However, up to now there is no known method to implement this experiment efficiently for a wide range of hyperedge densities. In fact, the first step **RC1** is easy to process, because we are only interested in values of  $d$  where  $H_k(n, m)$  is 2-colourable

w.h.p. and thus the conditioning on  $Z(H_k) > 0$  does not cause problems (the probability  $\mathbb{P}[H_k(n, m)$  is 2-colourable] is close to 1). But what turns the direct study of the distribution  $\pi_{k,n,m}^{\text{rc}}$  into a challenge is step **RC2**, because in the interesting density regimes we cannot even find one 2-colouring algorithmically let alone sample one uniformly: The currently best-performing algorithms for sampling a 2-colouring of  $H_k(n, m)$  are known to succeed up to density  $d'/k = c2^{k-1}/k$  for some constant  $c > 0$  [2], which is about a factor of  $k$  below the colourability threshold.

To circumvent these difficulties, we consider an alternative probability distribution on  $\Lambda_{k,n,m}$  called the *planted model*, which is much easier to approach. To describe this experiment, for  $\sigma : [n] \rightarrow \{0, 1\}$  let

$$\mathcal{F}(\sigma) = \binom{|\sigma^{-1}(0)|}{k} + \binom{|\sigma^{-1}(1)|}{k}$$

be the number of hyperedges of the complete hypergraph that are monochromatic under  $\sigma$ . Then the planted distribution is induced by the following experiment:

**PL1** Choose a map  $\sigma : [n] \rightarrow \{0, 1\}$  uniformly at random provided that that  $\mathcal{F}(\sigma) \leq \binom{n}{k} - m$ .

**PL2** Generate a  $k$ -uniform hypergraph  $\mathbf{H}$  on  $[n]$  consisting of  $m$  hyperedges that are bichromatic under  $\sigma$  uniformly at random. The result of the experiment is  $(\mathbf{H}, \sigma)$ .

Thus, the probability that the planted model assigns to a pair  $(H_k, \sigma)$  is

$$\pi_{k,n,m}^{\text{pl}}[H_k, \sigma] \sim \left[ 2^n \binom{\binom{n}{k}}{m} \mathbb{P}[\sigma \text{ is a 2-colouring of } H_k] \right]^{-1}.$$

We observe that in contrast to the "difficult" step **RC2**, step **PL2** is much easier to implement.

Of course, the two probability distributions  $\pi_{k,n,m}^{\text{rc}}$  and  $\pi_{k,n,m}^{\text{pl}}$  differ. Under  $\pi_{k,n,m}^{\text{rc}}$ , the hypergraph is chosen uniformly at random, whereas under  $\pi_{k,n,m}^{\text{pl}}$  it comes up with a probability that is proportional to its number of solutions, meaning that hypergraphs exhibiting many 2-colourings are "favoured" by the planted model.

However, the two models are related if  $m = m(n)$  is such that

$$\ln Z(H_k(n, m)) = \ln \mathbb{E}[Z(H_k(n, m))] + o(n) \quad \text{w.h.p.} \quad (1.2)$$

Coja-Oghlan and Achlioptas showed in [1] (where it was actually formulated for the problem of graph  $k$ -colouring, but the authors asserted that it also holds for hypergraph 2-colouring), that if (1.2) is satisfied, then the following is true.

If  $(\mathcal{E}_n)$  is a sequence of events  $\mathcal{E}_n \subset \Lambda_{k,n,m}$  such that  $\pi_{k,n,m}^{\text{pl}}[\mathcal{E}_n] \leq \exp[-\Omega(n)]$ , then  $\pi_{k,n,m}^{\text{rc}}[\mathcal{E}_n] = o(1)$ .

$$(1.3)$$

The statement (1.3) was baptised “quiet planting” by Krzakala and Zdeborová [14] and has ever since been used to study the behaviour of the set of colourings and its geometrical structure in various random constraint satisfaction problems [1, 5, 15, 16, 17]. Yet a significant complication in the use of (1.3) is that  $\mathcal{E}_n$  must not only have a small probability but is required to be *exponentially* unlikely in the planted model. This has caused substantial difficulties in several applications (e.g., [6, 5, 15]).

Theorem 1 enables us to establish a very strong connection between the random colouring model and the planted model. To state this, we recall the following definition. Suppose that  $\boldsymbol{\mu} = (\mu_n)_{n \geq 1}$ ,  $\boldsymbol{\nu} = (\nu_n)_{n \geq 1}$  are two sequences of probability measures such that  $\mu_n, \nu_n$  are defined on the same probability space  $\Omega_n$  for every  $n$ . Then  $(\mu_n)_{n \geq 1}$  is *contiguous* with respect to  $(\nu_n)_{n \geq 1}$ , in symbols  $\boldsymbol{\mu} \triangleleft \boldsymbol{\nu}$ , if for any sequence  $(\mathcal{E}_n)_{n \geq 1}$  of events such that  $\lim_{n \rightarrow \infty} \nu_n(\mathcal{E}_n) = 0$  we have  $\lim_{n \rightarrow \infty} \mu_n(\mathcal{E}_n) = 0$ .

We show that as a consequence of (1) the statement (1.3) can be sharpened in the strongest possible sense. Roughly speaking, we are going to show that in a density regime nearly up to the second moment lower bound the random colouring model is contiguous with respect to the planted model, i.e., in (1.3) it suffices that  $\pi_{k,n,m}^{\text{pl}}[\mathcal{E}_n] = o(1)$ .

**Corollary 4.** *Assume that  $d/k \leq 2^{k-1} \ln 2 - 2$ . Then  $(\pi_{k,n,m}^{\text{rc}})_{n \geq 1} \triangleleft (\pi_{k,n,m}^{\text{pl}})_{n \geq 1}$ .*

As done in [4], we refer to this contiguity statement as *silent planting*.

### 1.3 Discussion and further related work.

The ideas for the proofs follow the way beaten in [4], where statements analogue to Corollary 3 and Corollary 4 are shown for the problem of  $k$ -colouring random graphs. However, Theorem 1 is stronger than the results obtained in [4] because we determine the exact distribution of  $\ln Z - \mathbb{E}[\ln Z]$  asymptotically. The main observation used in the proofs is that the variance in the logarithm of the number of 2-colourings can be attributed to the fluctuations of the number of cycles of bounded length. The same phenomenon was observed in [4] and also in [9], where a combination of the second moment method and small subgraph conditioning was applied to derive a result similar to ours for the problem of random regular  $k$ -SAT.

Small subgraph conditioning was originally developed by Robinson and Wormald in [20] to investigate the Hamiltonicity of random regular graphs of degree at least three. Janson showed in [12] that the method can be used to obtain limiting distributions. Neeman und Netrapalli [19] used the method to obtain a result on non-distinguishability of the Erdős-Rényi model and the stochastic block model. Moore [18] used the method to determine the satisfiability threshold for positive 1-in- $k$ -SAT, a Boolean satisfiability problem where each clause contains  $k$  variables and demands that exactly one of them is true.

Similar to [12], we aim at obtaining a limiting distribution. Unfortunately, Jansons result does not apply directly in our case for the following reason. Since verifying the required properties to apply small subgraph conditioning directly for the random variable  $Z$  is very intricate, we break  $Z$  down into  $Z = \sum_s Z^s$  for some number  $s > 0$  and determine

the first and second moment of these smaller variables  $Z^s$ . However, it is not evident how to apply Jansons result to that growing number of variables simultaneously. Instead, we choose to perform a variance analysis along the lines of [20]. The same approach was pursued in [9], and thus our proof technique is similar to theirs in flavour, but we have the advantage of only having to deal with a very moderately growing number  $s$  of variables, which simplifies matters slightly.

## 1.4 Preliminaries and notation

We always assume that  $n \geq n_0$  is large enough for our various estimates to hold and denote by  $[n]$  the set  $\{1, \dots, n\}$ .

We use the standard  $O$ -notation when referring to the limit  $n \rightarrow \infty$ . Thus,  $f(n) = O(g(n))$  means that there exist  $C > 0$ ,  $n_0 > 0$  such that for all  $n > n_0$  we have  $|f(n)| \leq C \cdot |g(n)|$ . 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  $n \rightarrow \infty$ . Furthermore, the notation  $f(n) \sim g(n)$  means that  $f(n) = g(n)(1 + o(1))$  or equivalently  $\lim_{n \rightarrow \infty} f(n)/g(n) = 1$ . Besides taking the limit  $n \rightarrow \infty$ , at some point we need to consider the limit  $\nu \rightarrow \infty$  for some number  $\nu \in \mathbb{N}$ . Thus, we introduce  $f(n, \nu) \sim_\nu g(n, \nu)$  meaning that  $\lim_{\nu \rightarrow \infty} \lim_{n \rightarrow \infty} f(n, \nu)/g(n, \nu) = 1$ .

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

$$\mathcal{H}(p) = - \sum_{i=1}^l p_i \ln p_i. \quad (1.4)$$

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  $h$ th falling factorial of  $x$ .

For the sake of simplicity we choose to prove Theorem 1 using the random hypergraph model  $\mathcal{H}(n, m)$ . This is a random  $k$ -uniform (multi-)hypergraph on the vertex set  $[n]$  obtained by choosing  $m$  hyperedges  $e_1, \dots, e_m$  of the complete hypergraph 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 shows that this is quite unlikely.

**Fact 5.** *Assume that  $m = m(n)$  is a sequence such that  $m = O(n)$  and let  $\mathcal{A}$  be the event that  $\mathcal{H}(n, m)$  has no multiple edges. Then  $\mathbb{P}[\neg \mathcal{A}] = O(n^{2-k})$ .*

All statements and all proofs are valid for any  $k \geq 3$ , where  $k$  is always assumed to be independent of  $n$ .

## 2 Outline of the proof

We classify the 2-colourings according to their proportion of assigned colours: For a map  $\sigma : [n] \rightarrow \{0, 1\}$  we define

$$\rho(\sigma) = |\sigma^{-1}(0)|/n \quad (2.1)$$

and call this value the *colour density* of  $\sigma$ . We let  $\mathcal{A}(n)$  signify the set of all possible colour densities  $\rho(\sigma)$  for  $\sigma : [n] \rightarrow \{0, 1\}$ . We will later show that when bounding the moments of  $Z(\mathcal{H}(n, m))$  we can confine ourselves to colourings such that the proportion of the two colours does not deviate too much from  $1/2$ . Formally, we say that  $\rho \in [0, 1]$  is  $(\omega, n)$ -balanced for  $\omega \in \mathbb{N}$  if

$$\rho \in \left[ \frac{1}{2} - \frac{\omega}{\sqrt{n}}, \frac{1}{2} + \frac{\omega}{\sqrt{n}} \right)$$

and we denote by  $\mathcal{A}_\omega(n)$  the set of all  $(\omega, n)$ -balanced colour densities  $\rho \in \mathcal{A}(n)$ . For a hypergraph  $H$  on  $[n]$  we let  $Z_\omega(H)$  signify the number of  $(\omega, n)$ -balanced colourings, which are 2-colourings  $\sigma$  such that  $\rho(\sigma) \in \mathcal{A}_\omega(n)$ . It will turn out useful to split up the set  $\mathcal{A}_\omega(n)$  into smaller sets in the following way. For  $\nu \in \mathbb{N}$  and  $s \in [\omega\nu]$  let

$$\rho_{\omega, \nu}^s = \frac{1}{2} - \frac{\omega}{\sqrt{n}} + \frac{2s - 1}{\nu\sqrt{n}}. \tag{2.2}$$

Let  $\mathcal{A}_{\omega, \nu}^s(n)$  be the set of all colour densities  $\rho \in \mathcal{A}(n)$  such that

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

For a hypergraph  $H$  let  $Z_{\omega, \nu}^s(H)$  denote the number of 2-colourings  $\sigma$  of  $H$  such that  $\rho(\sigma) \in \mathcal{A}_{\omega, \nu}^s(n)$ . We are going to apply small subgraph conditioning to  $Z_{\omega, \nu}^s$  rather than directly to  $Z$ . We observe that for each fixed  $\nu$  we have  $Z_\omega = \sum_{s=1}^{\omega\nu} Z_{\omega, \nu}^s$ . In Section 3 we will calculate the first moments of  $Z$  and  $Z_\omega$  to obtain the following.

**Proposition 6.** *Let  $k \geq 3, d' \in (0, \infty)$  and  $\omega > 0$ . Then*

$$\mathbb{E}[Z(\mathcal{H}(n, m))] = \Theta(2^n (1 - 2^{1-k})^m) \quad \text{and} \quad \lim_{\omega \rightarrow \infty} \liminf_{n \rightarrow \infty} \frac{\mathbb{E}[Z_\omega(\mathcal{H}(n, m))]}{\mathbb{E}[Z(\mathcal{H}(n, m))]} = 1.$$

As outlined in Section 1.3, our basic strategy is to show that the fluctuations of  $\ln Z$  can be attributed to fluctuations in the number of cycles of a bounded length. By a cycle of length  $l$  for an integer  $l \geq 2$  we mean a structure of  $l$  vertices and  $l$  distinct edges in the hypergraph such that two consecutive vertices in the cycle are connected by one of the edges. We let  $C_{l, n}$  denote the number of cycles of length exactly  $l$  in  $\mathcal{H}(n, m)$ . Let

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

We will see that  $\lambda_l$  denotes the expected number of cycles of length  $l$  in a random  $k$ -uniform hypergraph, whereas  $\delta_l$  is a correction factor taking into account that we only allow for bichromatic edges. It is well-known that  $C_{2, n}, \dots$  are asymptotically independent Poisson variables [7, Theorem 5.16]. More precisely, we have the following.

**Fact 7.** If  $c_2, \dots, c_L$  are non-negative integers, then

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

Next, we investigate the impact of the cycle counts  $C_{l,n}$  on the first moment of  $Z_{\omega,\nu}^s$ . In Section 4 we prove the following.

**Proposition 8.** Assume that  $k \geq 3$  and  $d'(k-1) < (2^{k-1} - 1)^2$ . Then

$$\sum_{l=2}^{\infty} \lambda_l \delta_l^2 < \infty. \tag{2.4}$$

Moreover, let  $L \geq 2$  be an integer and  $\omega, \nu \in \mathbb{N}$ . If  $c_2, \dots, c_L$  are non-negative integers, then for any  $s \in [\omega\nu]$ :

$$\frac{\mathbb{E}[Z_{\omega,\nu}^s(\mathcal{H}(n, m)) | \forall 2 \leq l \leq L : C_{l,n} = c_l]}{\mathbb{E}[Z_{\omega,\nu}^s(\mathcal{H}(n, m))]} \sim \prod_{l=2}^L [1 + \delta_l]^{c_l} \exp[-\delta_l \lambda_l]. \tag{2.5}$$

Additionally, we need to know the second moment of  $Z_{\omega,\nu}^s$  very precisely. The following proposition is the key result of our approach and the one that requires the most of the technical work. Its proof can be found at the end of Section 5.

**Proposition 9.** Assume that  $k \geq 3$  and  $d'/k < 2^{k-1} \ln 2 - 2$  and let  $\omega, \nu \in \mathbb{N}$ . Then for every  $s \in [\omega\nu]$  we have

$$\frac{\mathbb{E}[Z_{\omega,\nu}^s(\mathcal{H}(n, m))^2]}{\mathbb{E}[Z_{\omega,\nu}^s(\mathcal{H}(n, m))]^2} \sim_{\nu} \exp\left[\sum_{l \geq 2} \lambda_l \delta_l^2\right].$$

We now derive Theorem 1 from Propositions 6-9. The key observation we will need is that the variance of the random variables  $Z_{\omega,\nu}^s$  can almost entirely be attributed to the fluctuations of the number of short cycles. As done in [9], the arguments we use are similar to the small subgraph conditioning from [12, 20]. But we do not refer to any technical statements from [12, 20] directly because instead of working only with the random variable  $Z$  we need to control all  $Z_{\omega,\nu}^s$  for fixed  $\omega, \nu \in \mathbb{N}$  simultaneously. In fact, ultimately we have to take  $\nu \rightarrow \infty$  and  $\omega \rightarrow \infty$  as well. Our line of argument follows the path beaten in [9] and the following three lemmas are an adaption of the ones there.

For  $L > 2$  let  $\mathcal{F}_L = \mathcal{F}_{L,n}(d, k)$  be the  $\sigma$ -algebra generated by the random variables  $C_{l,n}$  with  $2 \leq l \leq L$ . For each  $L \geq 2$  the standard decomposition of the variance yields

$$\text{Var}[Z_{\omega,\nu}^s(\mathcal{H}(n, m))] = \text{Var}[\mathbb{E}[Z_{\omega,\nu}^s(\mathcal{H}(n, m)) | \mathcal{F}_L]] + \mathbb{E}[\text{Var}[Z_{\omega,\nu}^s(\mathcal{H}(n, m)) | \mathcal{F}_L]].$$

The term  $\text{Var}[\mathbb{E}[Z_{\omega,\nu}^s(\mathcal{H}(n, m)) | \mathcal{F}_L]]$  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

$$\mathbb{E} [\text{Var} [Z_{\omega,\nu}^s(\mathcal{H}(n, m)) | \mathcal{F}_L]] = \mathbb{E} \left[ \mathbb{E} [Z_{\omega,\nu}^s(\mathcal{H}(n, m))^2 | \mathcal{F}_L] - \mathbb{E} [Z_{\omega,\nu}^s(\mathcal{H}(n, m)) | \mathcal{F}_L]^2 \right].$$

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  $\ln Z_{\omega,\nu}^s$  is just the limit of  $\ln \mathbb{E} [Z_{\omega,\nu}^s | \mathcal{F}_L]$  as  $n, L \rightarrow \infty$ . This limit is determined by the joint distribution of the number of short cycles.

**Lemma 10.** *For  $d'/k < 2^{k-1} \ln 2 - 2$  and any  $\omega, \nu \in \mathbb{N}$  and  $s \in [\omega\nu]$  we have*

$$\limsup_{L \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{E} \left[ \frac{\mathbb{E} [Z_{\omega,\nu}^s(\mathcal{H}(n, m))^2 | \mathcal{F}_L] - \mathbb{E} [Z_{\omega,\nu}^s(\mathcal{H}(n, m)) | \mathcal{F}_L]^2}{\mathbb{E} [Z_{\omega,\nu}^s(\mathcal{H}(n, m))]^2} \right] = 0.$$

*Proof.* Fix  $\omega, \nu \in \mathbb{N}$  and set  $Z_s = Z_{\omega,\nu}^s(\mathcal{H}(n, m))$ . Using Fact 7 and equation (2.5) from Proposition 8 we can choose for any  $\varepsilon > 0$  a constant  $B = B(\varepsilon)$  and  $L \geq L_0(\varepsilon)$  large enough such that for each large enough  $n \geq n_0(\varepsilon, B, L)$  we have for any  $s \in [\omega\nu]$ :

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

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

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

Finally, the estimate  $\exp[-x] \geq 1 - x$  for  $|x| < 1/8$  combined with (2.7) and Proposition 9 implies that for large enough  $\nu, n, L$  and each  $s \in [\omega\nu]$  we have

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

As this holds for any  $\varepsilon > 0$  and as the expression  $\exp [\sum_{l=2}^{\infty} \delta_l^2 \lambda_l]$  is bounded by equation (2.4) from Proposition 8, the proof of the lemma is completed by first taking  $n \rightarrow \infty$  and then  $L \rightarrow \infty$ .  $\square$

**Lemma 11.** *For  $d'/k < 2^{k-1} \ln 2 - 2$  and any  $\alpha > 0$  we have*

$$\limsup_{L \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} [ |Z(\mathcal{H}(n, m)) - \mathbb{E} [Z(\mathcal{H}(n, m)) | \mathcal{F}_L]| > \alpha \mathbb{E} [Z(\mathcal{H}(n, m))] ] = 0.$$

*Proof.* To unclutter the notation, we set  $Z = Z(\mathcal{H}(n, m))$  and  $Z_\omega = Z_\omega(\mathcal{H}(n, m))$ . First we observe that Proposition 6 implies that for any  $\alpha > 0$  we can choose  $\omega \in \mathbb{N}$  large enough such that

$$\liminf_{n \rightarrow \infty} \frac{\mathbb{E} [Z_\omega]}{\mathbb{E} [Z]} > (1 - \alpha^2). \quad (2.8)$$

We let  $\nu \in \mathbb{N}$ . To prove the statement, we need to get a handle on the cases in which the random variables  $Z_{\omega, \nu}^s(\mathcal{H}(n, m))$  deviate significantly from their conditional expectation  $\mathbb{E} [Z_{\omega, \nu}^s(\mathcal{H}(n, m)) | \mathcal{F}_L]$ . We let  $Z_s = Z_{\omega, \nu}^s(\mathcal{H}(n, m))$  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  $X = \sum_{s=1}^{\omega\nu} X_s$ . Then these definitions directly yield

$$\mathbb{P} [X < \alpha \mathbb{E} [Z_\omega]] \leq \mathbb{P} [|Z_\omega - \mathbb{E} [Z_\omega | \mathcal{F}_L]| < 2\alpha \mathbb{E} [Z_\omega]]. \quad (2.9)$$

To get an upper bound on  $\mathbb{E} [X_s | \mathcal{F}_L]$ , we cleverly decompose the  $X_s$ 's and together with Chebyshev's inequality it is true for every  $s$  that

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

Hence, using that with Proposition 6 there is a number  $\beta = \beta(\alpha, \omega)$  such that  $\mathbb{E} [Z_s] / \mathbb{E} [Z] \leq \beta / (\omega\nu)$  for all  $s \in [\omega\nu]$  and  $n$  large enough, we have

$$\mathbb{E} [X | \mathcal{F}_L] \leq \sum_{s=1}^{\omega\nu} \frac{4 \text{Var} [Z_s | \mathcal{F}_L]}{\alpha \mathbb{E} [Z_s]} \leq \frac{2\beta \mathbb{E} [Z]}{\alpha \nu \omega} \sum_{s=1}^{\omega\nu} \frac{\text{Var} [Z_s | \mathcal{F}_L]}{\mathbb{E} [Z_s]^2}.$$

Taking expectations, choosing  $\varepsilon = \varepsilon(\alpha, \beta, \omega)$  small enough and applying Lemma 10, we obtain

$$\mathbb{E} [X] = \mathbb{E} [\mathbb{E} [X | \mathcal{F}_L]] \leq \frac{2\beta \mathbb{E} [Z]}{\alpha \nu \omega} \sum_{s=1}^{\omega\nu} \frac{\mathbb{E} [\text{Var} [Z_s | \mathcal{F}_L]]}{\mathbb{E} [Z_s]^2} \leq \frac{4\beta \varepsilon \mathbb{E} [Z]}{\alpha} \leq \alpha^2 \mathbb{E} [Z]. \quad (2.10)$$

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

$$\mathbb{P} [|Z_\omega - \mathbb{E} [Z_\omega | \mathcal{F}_L]| < 2\alpha \mathbb{E} [Z_\omega]] \geq 1 - 2\alpha. \quad (2.11)$$

As (2.8) is equivalent to

$$\liminf_{n \rightarrow \infty} \frac{\mathbb{E} [Z] - \mathbb{E} [Z_\omega]}{\mathbb{E} [Z]} \leq \alpha^2,$$

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

$$\begin{aligned} & \mathbb{P} [|Z - \mathbb{E} [Z | \mathcal{F}_L]| > \alpha \mathbb{E} [Z]] \\ & \leq \mathbb{P} [|Z - Z_\omega| + |Z_\omega - \mathbb{E} [Z_\omega | \mathcal{F}_L]| + |\mathbb{E} [Z_\omega | \mathcal{F}_L] - \mathbb{E} [Z | \mathcal{F}_L]| > \alpha \mathbb{E} [Z]] \\ & \leq \mathbb{P} \left[ |Z - Z_\omega| > \frac{\alpha}{3} \mathbb{E} [Z] \right] + \mathbb{P} \left[ |Z_\omega - \mathbb{E} [Z_\omega | \mathcal{F}_L]| > \frac{\alpha}{3} \mathbb{E} [Z] \right] \\ & \quad + \mathbb{P} \left[ |\mathbb{E} [Z_\omega | \mathcal{F}_L] - \mathbb{E} [Z | \mathcal{F}_L]| > \frac{\alpha}{3} \mathbb{E} [Z] \right] \\ & \leq 3\alpha + \alpha/3 + 3\alpha < 7\alpha, \end{aligned}$$

which proves the statement. □

**Lemma 12.** *Let*

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

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

$$\limsup_{L \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} [|\ln \mathbb{E} [Z(\mathcal{H}(n, m)) | \mathcal{F}_L] - \ln \mathbb{E} [Z(\mathcal{H}(n, m))] - U_L| > \varepsilon] = 0 \quad (2.13)$$

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

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

Therefore, Fact 7 implies that

$$\mathbb{E} [|U_L|] \leq \sum_{l=2}^L |\delta_l| \sqrt{\lambda_l} + \delta_l^2 \lambda_l. \quad (2.14)$$

Proposition 8 ensures that  $\sum_l \delta_l^2 \lambda_l < \infty$ . Furthermore, as we are in the regime  $d'/k \leq 2^{k-1} \ln 2$ , we have  $\sum_l |\delta_l| \sqrt{\lambda_l} \leq \sum_l k^l 2^{-(k-1)l/2} < \infty$  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  $l \leq L$ . Referring to Fact 7, we can find for each  $L, \varepsilon > 0$  a  $B > 0$  such that

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

To simplify the notation we set  $Z = Z(\mathcal{H}(n, m))$  and  $Z_\omega = Z_\omega(\mathcal{H}(n, m))$ . By Proposition 6 we can choose for any  $\alpha > 0$  a  $\omega > 0$  large enough such that  $\mathbb{E}[Z_\omega] > (1 - \alpha)\mathbb{E}[Z]$  for large enough  $n$ . Then Propositions 6 and 8 combined with Fact 7 imply that for any  $c_1, \dots, c_L \leq B$  and small enough  $\alpha = \alpha(\varepsilon, L, B)$  we have for  $n$  large enough:

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

On the other hand, for  $\alpha$  sufficiently small and large enough  $n$  we have

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

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

*Proof of Theorem 1.* For  $L \geq 2$  we define

$$W_L = \sum_{l=2}^L X_l \ln(1 + \delta_l) - \lambda_l \delta_l.$$

Then Fact 7 implies that for each  $L$  the random variable  $U_L$  defined in (2.12) converges in distribution to  $W_L$  as  $n \rightarrow \infty$ . Furthermore, because  $\sum_l \delta_l \sqrt{\lambda_l}, \sum_l \delta_l^2 \lambda_l < \infty$ , the martingale convergence theorem implies that  $W$  is well-defined and that the  $W_L$  converge to  $W$  almost surely as  $L \rightarrow \infty$ . Therefore, from Lemmas 12 and 11 it follows that  $\ln Z(\mathcal{H}(n, m)) - \ln \mathbb{E}[Z(\mathcal{H}(n, m))]$  converges to  $W$  in distribution, meaning that for any  $\varepsilon > 0$  we have

$$\lim_{n \rightarrow \infty} \mathbb{P}[|\ln Z(\mathcal{H}(n, m)) - \ln \mathbb{E}[Z(\mathcal{H}(n, m))]| - W| > \varepsilon] = 0. \quad (2.18)$$

To derive Theorem 1 from (2.18) let  $S$  be the event that  $\mathcal{H}(n, m)$  consists of  $m$  distinct edges. Given that  $S$  occurs,  $\mathcal{H}(n, m)$  is identical to  $H_k(n, m)$ . Furthermore, Fact 5 implies that  $\mathbb{P}[S] = \Omega(1)$ . Consequently, (2.18) yields

$$\begin{aligned} 0 &= \lim_{n \rightarrow \infty} \mathbb{P}[|\ln Z(\mathcal{H}(n, m)) - \ln \mathbb{E}[Z(\mathcal{H}(n, m))]| - W| > \varepsilon | S] \\ &= \lim_{n \rightarrow \infty} \mathbb{P}[|\ln Z(H_k(n, m)) - \ln \mathbb{E}[Z(\mathcal{H}(n, m))]| - W| > \varepsilon]. \end{aligned} \quad (2.19)$$

Furthermore, Lemma 13 implies that  $\mathbb{E}[Z(\mathcal{H}(n, m))], \mathbb{E}[Z(H_k(n, m))] = \Theta(2^n(1 - 2^{1-k})^m)$ . Thus, it holds that  $\mathbb{E}[Z(\mathcal{H}(n, m))] = \Theta(\mathbb{E}[Z(H_k(n, m))])$  and with (2.19) it follows that

$$\lim_{n \rightarrow \infty} \mathbb{P}[|\ln Z(H_k(n, m)) - \ln \mathbb{E}[Z(H_k(n, m))]| - W| > \varepsilon] = 0,$$

which proves Theorem 1. □

*Proof of Corollary 3.* The first part of the proof follows directly from Theorem 1 and the properties of  $W$ . By the definition of convergence in distribution and Markov's inequality we have

$$\lim_{n \rightarrow \infty} \mathbb{P}[|\ln Z(H_k(n, m)) - \ln \mathbb{E}[Z(H_k(n, m))]| \leq \omega] = \mathbb{P}[|W| \leq \omega] \geq 1 - \frac{\mathbb{E}|W|}{\omega}$$

and (1.1) follows.

To prove the second part, we construct an event whose probability is bounded away from 0 and that is such that conditioned on this event, the number of solutions of the random hypergraph  $H_k(n, m)$  is not concentrated very strongly.

We consider the event  $\mathcal{T}_t$  that the random hypergraph  $H_k(n, m)$  contains  $t$  isolated triangles, i. e.  $t$  connected components such that each component consists of  $3k - 3$  vertices and 3 edges and the intersection of each pair of edges contains exactly one vertex. It is well-known that for  $t \geq 0$  there exists  $\varepsilon = \varepsilon(d, t) > 0$  such that

$$\liminf_{n \rightarrow \infty} \mathbb{P}[\mathcal{T}_t] > \varepsilon. \tag{2.20}$$

Given  $\mathcal{T}_t$ , we let  $H_k^*(n, m)$  denote the random hypergraph obtained by choosing a set of  $t$  isolated triangles randomly and removing them. Then  $H_k^*(n, m)$  is identical to  $H_k(n - (3k - 3)t, m - 3t)$  and with Proposition 6 there exists a constant  $C = C(d, k)$  such that

$$\mathbb{E}[Z(H_k^*(n, m))] = \mathbb{E}[Z(H_k(n - (3k - 3)t, m - 3t))] \leq C \cdot 2^{n - (3k - 3)t} (1 - 2^{1-k})^{m - 3t}.$$

A very accurate calculation of the number of 2-colourings of a triangle in a hypergraph yields that this number is given by  $(2^{k-2} - 1)(2^{2k-1} - 2^k + 2)$ . Thus, we obtain

$$\begin{aligned} \mathbb{E}[Z(H_k(n, m)) | \mathcal{T}_t] &\leq \mathbb{E}[Z(H_k(n - (3k - 3)t, m - 3t))] ((2^{k-2} - 1)(2^{2k-1} - 2^k + 2))^t \\ &\leq C \cdot 2^n (1 - 2^{1-k})^{m - 3t} (1 - 2^{2-k})^t (1 - 2^{1-k} + 2^{2-2k})^t \\ &\leq C \cdot 2^n (1 - 2^{1-k})^m (1 - 8(2^k - 2)^{-3}) \\ &\leq O(\mathbb{E}[Z(H_k(n, m))]) (1 - 8(2^k - 2)^{-3}), \end{aligned}$$

implying that for any  $\omega > 0$  we can choose  $t$  large enough so that  $\mathbb{E}[Z(H_k(n, m)) | \mathcal{T}_t] \leq \mathbb{E}[Z(H_k(n, m))] / (2 \exp[\omega])$ . Using Markov's inequality, we obtain

$$\begin{aligned} \mathbb{P}[\ln Z(H_k(n, m)) \geq \ln \mathbb{E}[Z(H_k(n, m))] - \omega | \mathcal{T}_t] \\ = \mathbb{P}[Z(H_k(n, m)) / \mathbb{E}[Z(H_k(n, m))] \geq \exp[-\omega] | \mathcal{T}_t] \leq 1/2. \end{aligned} \tag{2.21}$$

Thus, combining (2.20) and (2.21) yields that for any finite  $\omega > 0$  there is  $\varepsilon > 0$  such that for large enough  $n$  we have

$$\begin{aligned} & \mathbb{P} [ | \ln Z(H_k(n, m)) - \mathbb{E} [\ln Z(H_k(n, m))] | > \omega ] \\ & \geq \mathbb{P} [ \ln Z(H_k(n, m)) < \mathbb{E} [\ln Z(H_k(n, m))] - \omega ] \\ & \geq \mathbb{P} [ \ln Z(H_k(n, m)) \geq \ln \mathbb{E} [Z(H_k(n, m))] - \omega | \mathcal{T}_t ] \mathbb{P} [\mathcal{T}_t] \\ & > \varepsilon/2, \end{aligned}$$

thereby completing the proof of the second claim.  $\square$

*Proof of Corollary 4.* This proof is nearly identical to the one in [4]. Assume for contradiction that  $(\mathcal{A}_n)_{n \geq 1}$  is a sequence of events such that for some fixed number  $0 < \varepsilon < 1/2$  we have

$$\lim_{n \rightarrow \infty} \pi_{k,n,m}^{\text{pl}} [\mathcal{A}_n] = 0 \quad \text{while} \quad \limsup_{n \rightarrow \infty} \pi_{k,n,m}^{\text{rc}} [\mathcal{A}_n] > \varepsilon. \quad (2.22)$$

Let  $\sigma : [n] \rightarrow \{0, 1\}$  be a map and  $H_k(n, m, \sigma)$  denote a  $k$ -uniform hypergraph on  $[n]$  with precisely  $m$  edges chosen uniformly at random from all edges that are bichromatic under  $\sigma$ . Let  $\mathcal{V}(\sigma)$  be the event that  $\sigma$  is a 2-colouring of  $H_k(n, m)$  and  $Z(H_k(n, m))\mathbf{1}_{\mathcal{A}_n}$  be the random variable counting the number of colourings  $\sigma$  of  $H_k(n, m)$  such that  $(H_k(n, m), \sigma) \in \mathcal{A}_n$ . Then

$$\begin{aligned} \mathbb{E} [Z(H_k(n, m))\mathbf{1}_{\mathcal{A}_n}] &= \sum_{\sigma: [n] \rightarrow \{0,1\}} \mathbb{P} [\mathcal{V}(\sigma) \text{ and } (H_k(n, m), \sigma) \in \mathcal{A}_n] \\ &= \sum_{\sigma: [n] \rightarrow \{0,1\}} \mathbb{P} [(H_k(n, m), \sigma) \in \mathcal{A}_n | \mathcal{V}(\sigma)] \mathbb{P} [\mathcal{V}(\sigma)] \\ &= \sum_{\sigma: [n] \rightarrow \{0,1\}} \mathbb{P} [H_k(n, m, \sigma) \in \mathcal{A}_n] \mathbb{P} [\mathcal{V}(\sigma)] \\ &\leq O((1 - 2^{1-k})^m) \sum_{\sigma: [n] \rightarrow \{0,1\}} \mathbb{P} [H_k(n, m, \sigma) \in \mathcal{A}_n] \\ &= O(2^n (1 - 2^{1-k})^m) \mathbb{P} [H_k(n, m, \sigma) \in \mathcal{A}_n] = o(2^n (1 - 2^{1-k})^m). \end{aligned} \quad (2.23)$$

By Theorem 1, for any  $\varepsilon > 0$  there is  $\delta > 0$  such that for all large enough  $n$  we have

$$\mathbb{P} [Z(H_k(n, m)) < \delta \mathbb{E} [Z(H_k(n, m))] ] < \varepsilon/2. \quad (2.24)$$

Now, let  $\mathcal{E}$  be the event that  $Z(H_k(n, m)) \geq \delta \mathbb{E} [Z(H_k(n, m))]$  and let  $q = \pi_{k,n,m}^{\text{rc}} [\mathcal{A}_n | \mathcal{E}]$ . Then

$$\begin{aligned} \mathbb{E} [Z(H_k(n, m))\mathbf{1}_{\mathcal{A}_n}] &\geq \delta \mathbb{E} [Z(H_k(n, m))] \mathbb{P} [((H_k(n, m), \sigma) \in \mathcal{A}_n, \mathcal{E})] \\ &\geq \delta q \mathbb{E} [Z(H_k(n, m))] \mathbb{P} [\mathcal{E}] \geq \delta q \mathbb{E} [Z(H_k(n, m))] / 2 \\ &= \frac{\delta q}{2} \cdot \Omega(2^n (1 - 2^{1-k})^m). \end{aligned} \quad (2.25)$$

Combining (2.23) and (2.25), we obtain  $q = o(1)$ . Hence, (2.24) implies that

$$\pi_{k,n,m}^{\text{rc}}[\mathcal{A}_n] = \pi_{k,n,m}^{\text{rc}}[\mathcal{A}_n | \neg \mathcal{E}] \cdot \mathbb{P}[\neg \mathcal{E}] + q \cdot \mathbb{P}[\mathcal{E}] \leq \mathbb{P}[\neg \mathcal{E}] + q \leq \varepsilon/2 + o(1),$$

in contradiction to (2.22). □

### 3 The first moment calculation

The aim in this section is to prove Proposition 6 and a result that we need for Proposition 9. For a hypergraph  $H$  let  $Z_\rho(H)$  be its number of 2-colourings with colour density  $\rho$ . We set  $\bar{\rho} = \frac{1}{2}$ . For  $\rho \in [0, 1]$  we define

$$f_1 : \rho \mapsto \mathcal{H}(\rho) + g_1(\rho) \quad \text{with} \quad g_1(\rho) = \frac{d}{k} \ln(1 - \rho^k - (1 - \rho)^k), \quad (3.1)$$

where  $\mathcal{H}(\rho)$  is the entropy introduced in (1.4). The next lemma shows that  $f_1(\rho)$  is the function we need to analyse in order to determine the expectation of  $Z_\rho$ .

**Lemma 13.** *Let  $d' \in (0, \infty)$ . There exist numbers  $C_1 = C_1(k, d)$ ,  $C_2 = C_2(k, d) > 0$  such that for any colour density  $\rho$ :*

$$C_1 n^{-1/2} \exp[nf_1(\rho)] \leq \mathbb{E}[Z_\rho(\mathcal{H}(n, m))] \leq C_2 \exp[nf_1(\rho)]. \quad (3.2)$$

Moreover, if  $|\rho - \bar{\rho}| = o(1)$ , then

$$\mathbb{E}[Z_\rho(\mathcal{H}(n, m))] \sim \sqrt{\frac{2}{\pi n}} \exp\left[\frac{d(k-1)}{2k-2}\right] \exp[nf_1(\rho)]. \quad (3.3)$$

*Proof.* The edges in the random hypergraph  $\mathcal{H}(n, m)$  are independent by construction, so the expected number of solutions with colour density  $\rho$  can be written as

$$\mathbb{E}[Z_\rho(\mathcal{H}(n, m))] = \binom{n}{\rho n} \left(1 - \frac{\binom{\rho n}{k} + \binom{(1-\rho)n}{k}}{N}\right)^m, \quad \text{where } N = \binom{n}{k}. \quad (3.4)$$

Further, the number of “forbidden” edges is given by

$$\begin{aligned} & \binom{\rho n}{k} + \binom{(1-\rho)n}{k} \\ &= \frac{1}{k!} \left( n^k (\rho^k + (1-\rho)^k) - \frac{k(k-1)}{2} n^{k-1} (\rho^{k-1} + (1-\rho)^{k-1}) + \Theta(n^{k-2}) \right) \\ &= N (\rho^k + (1-\rho)^k) - \frac{k(k-1)}{2k!} n^{k-1} (\rho^{k-1}(1-\rho) + \rho(1-\rho)^{k-1}) + \Theta(n^{k-2}) \end{aligned}$$

yielding

$$\begin{aligned} & 1 - \frac{\binom{\rho n}{k} + \binom{(1-\rho)n}{k}}{N} \\ &= 1 - \rho^k - (1-\rho)^k + \frac{k(k-1)}{2n} (\rho^{k-1}(1-\rho) + \rho(1-\rho)^{k-1}) + \Theta(n^{-2}). \end{aligned}$$

To proceed we observe that  $\ln(x + \frac{y}{n}) = \ln(x) + \ln(1 + \frac{y}{xn})$  for  $x > 0, y < xn$  and consequently

$$\begin{aligned} & m \ln \left( 1 - \frac{\binom{\rho n}{k} + \binom{(1-\rho)n}{k}}{N} \right) \\ &= \frac{dn}{k} \left( \ln(1 - \rho^k - (1-\rho)^k) + \ln \left( 1 - \frac{k(k-1)\rho^{k-1}(1-\rho) + \rho(1-\rho)^{k-1}}{2n} + \Theta(n^{-2}) \right) \right) \\ &\sim \frac{dn}{k} \ln(1 - \rho^k - (1-\rho)^k) + \frac{d(k-1)}{2} \left( \frac{\rho^{k-1}(1-\rho) + \rho(1-\rho)^{k-1}}{1 - \rho^k - (1-\rho)^k} \right) + \Theta(n^{-1}). \end{aligned} \quad (3.5)$$

Equation (3.2) follows from (3.4), (3.5) and Stirling's formula applied to  $\binom{n}{\rho n}$ . Moreover, equation (3.3) follows from (3.4) and (3.5) because  $|\rho - \bar{\rho}| = o(1)$  implies that

$$\binom{n}{\rho n} \sim \sqrt{\frac{2}{\pi n}} \exp[n\mathcal{H}(\rho)] \quad \text{and} \quad \frac{\rho^{k-1}(1-\rho) + \rho(1-\rho)^{k-1}}{1 - \rho^k - (1-\rho)^k} \sim \frac{1}{2^{k-1} - 1}. \quad \square$$

The following corollary states an expression for  $\mathbb{E}[Z(\mathcal{H}(n, m))]$ . Additionally, it shows that when  $\omega \rightarrow \infty$ , this value can be approximated by  $\mathbb{E}[Z_\omega(\mathcal{H}(n, m))]$ .

**Corollary 14.** *Let  $d' \in (0, \infty)$ . Then*

$$\mathbb{E}[Z(\mathcal{H}(n, m))] \sim \exp \left[ \frac{d(k-1)}{2^k - 2} + n f_1(\bar{\rho}) \right] \left( 1 + \frac{d(k-1)}{2^{k-1} - 1} \right)^{-\frac{1}{2}}. \quad (3.6)$$

Furthermore, for  $\omega > 0$  we have

$$\lim_{\omega \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{\mathbb{E}[Z_\omega(\mathcal{H}(n, m))]}{\mathbb{E}[Z(\mathcal{H}(n, m))]} = 1. \quad (3.7)$$

*Proof.* The functions  $\rho \mapsto \mathcal{H}(\rho)$  and  $\rho \mapsto g_1(\rho)$  are both concave and attain their maximum at  $\rho = \bar{\rho}$ . Consequently, setting  $B(d, k) = 4 \left( 1 + \frac{d(k-1)}{2^{k-1} - 1} \right)$  and expanding around  $\bar{\rho}$ , we obtain

$$f_1(\bar{\rho}) - \frac{B(d, k)}{2} (\rho - \bar{\rho})^2 - O((\rho - \bar{\rho})^3) \leq f_1(\rho) \leq f_1(\bar{\rho}) - \frac{B(d, k)}{2} (\rho - \bar{\rho})^2. \quad (3.8)$$

Plugging the upper bound from (3.8) into (3.2) and observing that the number of all colour densities for maps  $\sigma : [n] \rightarrow \{0, 1\}$  is bounded from above by  $n = \exp[o(n)]$ , we find

$$S_1 = \sum_{\rho: |\rho - \bar{\rho}| > n^{-3/8}} \mathbb{E}[Z_\rho(\mathcal{H}(n, m))] \leq C_2 \exp \left[ n f_1(\bar{\rho}) - \frac{B(d, k)}{2} n^{1/4} \right]. \quad (3.9)$$

On the other hand, equation (3.3) implies that

$$\begin{aligned} S_2 &= \sum_{\rho: |\rho - \bar{\rho}| \leq n^{-3/8}} \mathbb{E}[Z_\rho(\mathcal{H}(n, m))] \\ &\sim \sqrt{\frac{2}{\pi n}} \exp \left[ \frac{d(k-1)}{2^k - 2} \right] + \exp[n f_1(\bar{\rho})] \sum_{\rho} \exp \left[ -n \frac{B(d, k)}{2} (\rho - \bar{\rho})^2 \right]. \end{aligned} \quad (3.10)$$



The last sum is in the standard form of a Gaussian summation. Using the fact that  $\int_{-\infty}^{\infty} \exp[-a(x+b)^2] dx = \sqrt{\frac{\pi}{a}}$ , we get

$$\begin{aligned} \sum_{\rho \in \mathcal{A}(n)} \exp \left[ -n \frac{B(d,k)}{2} (\rho - \bar{\rho})^2 \right] &\sim n \int \exp \left[ -n \frac{B(d,k)}{2} (\rho - \bar{\rho})^2 \right] d\rho \\ &\sim n \sqrt{\frac{2\pi}{nB(d,k)}} = \sqrt{\frac{\pi n}{2}} \left( 1 + \frac{d(k-1)}{2^{k-1}-1} \right)^{-\frac{1}{2}} \end{aligned} \quad (3.11)$$

Plugging (3.11) into (3.10), we obtain

$$S_2 \sim \exp \left[ \frac{d(k-1)}{2^k-2} + nf_1(\bar{\rho}) \right] \left( 1 + \frac{d(k-1)}{2^{k-1}-1} \right)^{-\frac{1}{2}}. \quad (3.12)$$

Finally, comparing (3.9) and (3.12), we see that  $S_1 = o(S_2)$ . Thus,  $S_1 + S_2 \sim S_2$  and (3.6) follows from (3.12).

To prove (3.7), we find that analogously to (3.9), (3.10) and the calculation leading to (3.12), it holds that

$$S'_1 = \sum_{\rho: |\rho - \bar{\rho}| > \omega n^{-1/2}} \mathbb{E} [Z_\rho(\mathcal{H}(n, m))] \leq C_2 \exp \left[ nf_1(\bar{\rho}) - \frac{B(d,k)}{2} \omega^2 \right].$$

and

$$S'_2 = \sum_{\rho: |\rho - \bar{\rho}| \leq \omega n^{-1/2}} \mathbb{E} [Z_\rho(\mathcal{H}(n, m))] \sim \exp \left[ \frac{d(k-1)}{2^k-2} + nf_1(\bar{\rho}) \right] \left( 1 + \frac{d(k-1)}{2^{k-1}-1} \right)^{-\frac{1}{2}}.$$

Thus, we have  $\lim_{\omega \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{S'_1 + S'_2}{S'_2} = 1$ , yielding (3.7).  $\square$

*Proof of Proposition 6.* The statements are immediate by Corollary 14 and the fact that

$$f_1(\bar{\rho}) = \ln 2 + \frac{d}{k} \ln(1 - 2^{1-k}). \quad \square$$

Finally, we derive an expression for  $\mathbb{E} [Z_{\omega, \nu}^s(\mathcal{H}(n, m))]$  that we will need to prove Proposition 9.

**Lemma 15.** *Let  $d' \in (0, \infty)$ ,  $\omega, \nu \in \mathbb{N}$ ,  $s \in [\omega\nu]$  and  $\rho \in \mathcal{A}_{\omega, \nu}^s(n)$ . Then with  $\rho_{\omega, \nu}^s$  as defined in (2.2) we have*

$$\mathbb{E} [Z_{\omega, \nu}^s(\mathcal{H}(n, m))] \sim_\nu |\mathcal{A}_{\omega, \nu}^s(n)| \sqrt{\frac{2}{\pi n}} \exp \left[ \frac{d(k-1)}{2^k-2} \right] \exp [nf_1(\rho_{\omega, \nu}^s)].$$

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

$$f_1(\rho) = f_1(\rho_{\omega, \nu}^s) + \Theta \left( \frac{\omega}{\sqrt{n}} \right) |\rho - \rho_{\omega, \nu}^s| + \Theta \left( (\rho - \rho_{\omega, \nu}^s)^2 \right). \quad (3.13)$$

As  $|\rho - \rho_{\omega, \nu}^s| \leq \frac{1}{\nu\sqrt{n}}$  for  $\rho \in \mathcal{A}_{\omega, \nu}^s(n)$ , we conclude that  $f_1(\rho) = f_1(\rho_{\omega, \nu}^s) + O\left(\frac{\omega}{\nu n}\right)$  and as this is independent of  $\rho$  the assertion follows by inserting (3.13) in (3.3) and multiplying with  $|\mathcal{A}_{\omega, \nu}^s(n)|$ .  $\square$

## 4 Counting short cycles

We recall that for an integer  $L \geq 2$  and  $l \in \{2, \dots, L\}$  we denote by  $C_{l,n}$  the number of cycles of length  $l$  in  $\mathcal{H}(n, m)$ , where such a cycle consists of  $l$  vertices and  $l$  distinct edges, each of them connecting two consecutive vertices in the cycle. Further we let  $c_2, \dots, c_L$  be a sequence of non-negative integers and  $S$  be the event that  $C_{l,n} = c_l$  for  $l = 2, \dots, L$ . Additionally, for an assignment  $\sigma : [n] \rightarrow \{0, 1\}$  we let  $\mathcal{V}(\sigma)$  be the event that  $\sigma$  is a colouring of the random graph  $\mathcal{H}(n, m)$ . We also recall  $\lambda_l, \delta_l$  from (2.3).

*Proof of Proposition 8.* First observe that from the definition of  $\lambda_l$  and  $\delta_l$  in (2.3) and the fact that  $\sum_{n=1}^{\infty} \frac{x^n}{n} = -\ln(1-x)$  we get

$$\exp \left[ \sum_{l \geq 2} \lambda_l \delta_l^2 \right] = \exp \left[ -\frac{d(k-1)}{2} \frac{1}{(2^{k-1}-1)^2} \right] \left( 1 - \frac{d(k-1)}{(2^{k-1}-1)^2} \right)^{-1/2}. \quad (4.1)$$

Together with (4.1), Proposition 8 readily follows from the following lemma about the distribution of the random variables  $C_{l,n}$  given  $\mathcal{V}(\sigma)$ .

**Lemma 16.** *Let  $\mu_l = \frac{(d(k-1))^l}{2^l} \left[ 1 + \frac{(-1)^l}{(2^{k-1}-1)^l} \right]$ . Then  $\mathbb{P}[S|\mathcal{V}(\sigma)] \sim \prod_{l=2}^L \frac{\exp[-\mu_l]}{c_l!} \mu_l^{c_l}$  for any  $\sigma$  with  $\rho(\sigma) \in \mathcal{A}_\omega(n)$ .*

Before we establish Lemma 16, let us point out how it implies Proposition 8. By Bayes' rule, we have

$$\mathbb{E} [Z_{\omega, \nu}^s(\mathcal{H}(n, m)) | S] = \frac{1}{\mathbb{P}[S]} \sum_{\tau \in \mathcal{A}_{\omega, \nu}^s(n)} \mathbb{P}[\mathcal{V}(\tau)] \mathbb{P}[S|\mathcal{V}(\tau)]. \quad (4.2)$$

Inserting the result from Lemma 16 into (4.2) yields

$$\begin{aligned} \mathbb{E} [Z_{\omega, \nu}^s(\mathcal{H}(n, m)) | S] &\sim \frac{\prod_{l=2}^L \frac{\exp[-\mu_l]}{c_l!} \mu_l^{c_l}}{\mathbb{P}[S]} \sum_{\tau \in \mathcal{A}_{\omega, \nu}^s(n)} \mathbb{P}[\mathcal{V}(\tau)] \\ &\sim \frac{\prod_{l=2}^L \frac{\exp[-\mu_l]}{c_l!} \mu_l^{c_l}}{\mathbb{P}[S]} \mathbb{E} [Z_{\omega, \nu}^s(\mathcal{H}(n, m))]. \end{aligned}$$

From Lemma 16 and Fact 7 we get that

$$\frac{\prod_{l=2}^L \frac{\exp[-\mu_l]}{c_l!} \mu_l^{c_l}}{\mathbb{P}[S]} \sim \prod_{l=2}^L [1 + \delta_l]^{c_l} \exp[-\delta_l \lambda_l]$$

and Proposition 8 follows. □

*Proof of Lemma 16.* To prove this lemma, we are going to show that for any fixed sequence of integers  $m_1, \dots, m_L \geq 0$ , the joint factorial moments satisfy

$$\mathbb{E}[(C_{2,n})_{m_2} \cdots (C_{L,n})_{m_L} | \mathcal{V}(\sigma)] \sim \prod_{l=2}^L \mu_l^{m_l}. \quad (4.3)$$

Then Lemma 16 follows from [7, Theorem 1.23].

We consider the number of sequences of  $m_2 + \cdots + m_L$  distinct cycles such that  $m_2$  corresponds to the number of cycles of length 2, and so on. Clearly this number is equal to  $(C_{2,n})_{m_2} \cdots (C_{L,n})_{m_L}$ . We call a cycle *good*, if it does not contain edges that overlap on more than one vertex. We call a sequence of good cycles *good sequence* if for any two cycles  $C$  and  $C'$  in this sequence, there are no vertices  $v \in C$  and  $v' \in C'$  such that  $v$  and  $v'$  are contained in the same edge. Let  $Y$  be the number of good sequences and  $\bar{Y}$  be the number of sequences that are not good. Then it holds that

$$\mathbb{E}[(C_{2,n})_{m_2} \cdots (C_{L,n})_{m_L} | \mathcal{V}(\sigma)] = \mathbb{E}[Y | \mathcal{V}(\sigma)] + \mathbb{E}[\bar{Y} | \mathcal{V}(\sigma)]. \quad (4.4)$$

The following claim states that the contribution of  $\mathbb{E}[\bar{Y} | \mathcal{V}(\sigma)]$  is negligible. Its proof follows at the end of this section.

**Claim 17.** *We have  $\mathbb{E}[\bar{Y} | \mathcal{V}(\sigma)] = O(n^{-1})$ .*

Thus it remains to count good sequences given  $\mathcal{V}(\sigma)$ . We let  $\sigma \in \mathcal{A}_\omega(n)$  and first consider the number  $D_{l,n}$  of rooted, directed, good cycles of length  $l$ . This will introduce a factor of  $2l$  for the number of all good cycles of length  $l$ , thus  $D_{l,n} = 2lC_{l,n}$ . For a rooted, directed, good cycle of length  $l$  we need to pick  $l$  vertices  $(v_1, \dots, v_l)$  as roots, introducing a factor  $(1 + o(1)) \left(\frac{n}{2}\right)^l$ , and there have to exist edges between them which generates a factor  $(m / \binom{n}{k} (1 - 2^{1-k}))^l$ . To choose the remaining vertices in the participating edges we have to distinguish between pairs of vertices  $(v_i, v_{i+1})$  that are assigned the same colour and those that are not, because if  $\sigma(v_i) = \sigma(v_{i+1})$  we have to make sure that at least one of the other  $k - 2$  vertices participating in this edge is assigned the opposite colour. This gives rise to the third factor in the following calculation.

$$\begin{aligned} \mathbb{E}[D_{l,n} | \mathcal{V}(\sigma)] &\sim \left(\frac{n}{2}\right)^l \left[ \frac{m}{\binom{n}{k} (1 - 2^{1-k})} \right]^l \\ &\quad \cdot 2 \sum_{i=0}^l \left[ \binom{l}{i} \binom{n-2}{k-2}^i \left[ \binom{n-2}{k-2} - \binom{n/2}{k-2} \right]^{l-i} \mathbf{1}_{\{i \text{ is even}\}} \right] \\ &= \left(\frac{n}{2}\right)^l \left[ \frac{m}{\binom{n}{k} (1 - 2^{1-k})} \right]^l \cdot \left[ \left[ 2 \binom{n-2}{k-2} - \binom{n/2}{k-2} \right]^l + \left[ -\binom{n/2}{k-2} \right]^l \right] \\ &\sim \left(\frac{n}{2}\right)^l \left[ \frac{k!dn}{kn^k (1 - 2^{1-k})} \right]^l \cdot \left[ \frac{[(2^{k-1} - 1)n^{k-2}]^l + (-n^{k-2})^l}{[2^{k-2}(k-2)!]^l} \right] \\ &= [d(k-1)]^l \left( 1 + \frac{(-1)^l}{(2^{k-1} - 1)^l} \right) \end{aligned}$$

Hence, recalling that  $C_{l,n} = \frac{1}{2l}D_{l,n}$ , we get

$$\mathbb{E}[C_{l,n}|\mathcal{V}(\sigma)] \sim \frac{[d(k-1)]^l}{2l} \left( 1 + \frac{(-1)^l}{(2^{k-1}-1)^l} \right). \quad (4.5)$$

In fact, since  $Y$  considers only good sequences and  $l, m_2, \dots, m_L$  remain fixed as  $n \rightarrow \infty$ , (4.5) yields

$$\mathbb{E}[Y|\mathcal{V}(\sigma)] \sim \prod_{l=2}^L \left( \frac{[d(k-1)]^l}{2l} \left( 1 + \frac{(-1)^l}{(2^{k-1}-1)^l} \right) \right)^{m_l}.$$

Plugging the above relation and Claim 17 into (4.4) we get (4.3). The proposition follows.  $\square$

**Proof of Claim 17:** The idea of the proof is to find an event, namely the existence of an induced subgraph with too many edges, that always occurs if  $\bar{Y} > 0$  and whose probability can be bounded from above by  $O(n^{-1})$ . To this aim let  $A = \{i \in \mathbb{R} | i = (l-1)(k-1) + j \text{ for some } l \leq L, j \in \{0, \dots, k-2\}\}$ . For every subset  $R$  of  $(l-1)(k-1) + j$  vertices in the hypergraph, where  $l \leq L$  and  $j \in \{0, \dots, k-2\}$  let  $\mathbb{I}_R$  be the indicator that the number of edges only consisting of vertices in  $R$  is at least  $l$ , i.e.,  $\mathbb{I}_R$  evaluates to 1, if the induced subgraph with vertices in  $R$  contains too many edges. Let  $H_L$  be the event that  $\sum_{R:|R| \in A} \mathbb{I}_R > 0$ . It is direct to check that if  $\bar{Y} > 0$  then  $H_L$  occurs. This implies that

$$\mathbb{P}[\bar{Y} > 0 | \mathcal{V}(\sigma)] \leq \mathbb{P}[H_L | \mathcal{V}(\sigma)].$$

The claim follows by appropriately bounding  $\mathbb{P}[H_L | \mathcal{V}(\sigma)]$ . We are going to use Markov's inequality, i.e.

$$\mathbb{P}[H_L | \mathcal{V}(\sigma)] \leq \mathbb{E} \left[ \sum_{R:|R| \in A} \mathbb{I}_R | \mathcal{V}(\sigma) \right] = \sum_{l=2}^L \sum_{j=0}^{k-2} \sum_{R:|R|=(l-1)(k-1)+j} \mathbb{E}[\mathbb{I}_R | \mathcal{V}(\sigma)].$$

For any set  $R$  such that  $|R| = (l-1)(k-1) + j$ , we can put  $l$  edges inside the set in at most  $\binom{(l-1)(k-1)+j}{l}$  ways, which obviously gets largest if  $j = k-2$  and thus  $(l-1)(k-1) + j = l(k-1) - 1$ . Clearly conditioning on  $\mathcal{V}(\sigma)$  can only reduce the number of different placings of the edges.

We observe that for a colouring  $\sigma$  and two fixed vertices  $v$  and  $v'$  with  $\sigma(v) \neq \sigma(v')$  the probability that  $e(v, v')$  does not exist is  $\left(1 - \frac{1}{N - \mathcal{F}(\sigma)}\right)^m$ . Using inclusion/exclusion and the binomial theorem, with  $N = \binom{n}{k}$  and  $\mathcal{F}(\sigma) \sim 2^{1-k}N$ , for a fixed set  $R$  of cardinality  $(l-1)(k-1) + j$  we get that

$$\begin{aligned} \mathbb{E}[\mathbb{I}_R | \mathcal{V}(\sigma)] &\leq \binom{(l(k-1)-1)}{k} \sum_{i=0}^l \binom{l}{i} (-1)^i \left(1 - \frac{i}{N - \mathcal{F}(\sigma)}\right)^m \\ &\leq \binom{(l(k-1)-1)}{k} \left(\frac{m}{N - \mathcal{F}(\sigma)}\right)^l \sim \binom{(l(k-1)-1)}{k} \left(\frac{m}{\binom{n}{k} (1 - 2^{1-k})}\right)^l. \end{aligned}$$

With  $m = \frac{dn}{k}$  and since  $\binom{i}{j} \leq (ie/j)^j$ , it holds that

$$\begin{aligned} & \mathbb{P}[H_L | \mathcal{V}(\sigma)] \\ & \leq (1 + o(1)) \sum_{l=2}^L \binom{n}{l(k-1)-1} \binom{l(k-1)-1}{l} \left( \frac{m}{\binom{n}{k} (1-2^{1-k})} \right)^l \\ & = (1 + o(1)) \sum_{l=2}^L \left( \frac{ne}{l(k-1)-1} \right)^{l(k-1)-1} \left( \frac{e^{k+1} (l(k-1)-1)^k}{k^k l} \right)^l \left( \frac{mk^k}{n^k e^k (1-2^{1-k})} \right)^l \\ & = (1 + o(1)) \sum_{l=2}^L \frac{m^l e^{kl-1} (l(k-1)-1)^{l+1}}{n^{l+1} l^l (1-2^{1-k})^l} \\ & = \frac{1 + o(1)}{n} \sum_{l=2}^L \left( \frac{e^k d (l(k-1)-1)}{l(1-2^{1-k})} \right)^l \frac{l(k-1)-1}{e} = O(n^{-1}), \end{aligned}$$

where the last equality follows since  $L$  is a fixed number. □

## 5 The second moment calculation

In this section we prove Proposition 9. To this end, we need to derive an expression for the second moment of the random variables  $Z_{\omega, \nu}^s$  for  $s \in [\omega \nu]$  that is asymptotically tight. As a consequence, we need to put more effort into the calculations than done in prior work on hypergraph-2-colouring (e.g. [10]), where the second moment of  $Z$  is only determined up to a constant factor. Part of the proof is based on ideas from [4], but as we aim for a stronger result, the arguments are extended and adapted to our situation.

### 5.1 The overlap

For two colour assignments  $\sigma, \tau : [n] \rightarrow \{0, 1\}$  we define the *overlap matrix*

$$\rho(\sigma, \tau) = \begin{pmatrix} \rho_{00}(\sigma, \tau) & \rho_{01}(\sigma, \tau) \\ \rho_{10}(\sigma, \tau) & \rho_{11}(\sigma, \tau) \end{pmatrix}$$

with entries

$$\rho_{ij}(\sigma, \tau) = \frac{1}{n} \cdot |\sigma^{-1}(i) \cap \tau^{-1}(j)| \quad \text{for } i, j \in \{0, 1\}.$$

Obviously, it holds that

$$\rho_{00}(\sigma, \tau) + \rho_{01}(\sigma, \tau) + \rho_{10}(\sigma, \tau) + \rho_{11}(\sigma, \tau) = 1.$$

If we further remember the definition from (2.1), we can alternatively represent  $\rho(\sigma, \tau)$  as

$$\rho(\sigma, \tau) = \begin{pmatrix} \rho_{00}(\sigma, \tau) & \rho(\sigma) - \rho_{00}(\sigma, \tau) \\ \rho(\tau) - \rho_{00}(\sigma, \tau) & 1 - \rho(\sigma) - \rho(\tau) + \rho_{00}(\sigma, \tau) \end{pmatrix}.$$

To simplify the notation, for a  $2 \times 2$ -matrix  $\rho = (\rho_{ij})$  we introduce the shorthands

$$\rho_{i\star} = \rho_{i0} + \rho_{i1}, \quad \rho_{\cdot\star} = (\rho_{0\star}, \rho_{1\star}), \quad \rho_{\star j} = \rho_{0j} + \rho_{1j}, \quad \rho_{\star\cdot} = (\rho_{\star 0}, \rho_{\star 1}).$$

We let  $\mathcal{B}(n)$  be the set of all overlap matrices  $\rho(\sigma, \tau)$  for  $\sigma, \tau : [n] \rightarrow \{0, 1\}$  and  $\mathcal{B}$  denote the set of all probability distributions  $\rho = (\rho_{ij})_{i,j \in \{0,1\}}$  on  $\{0, 1\}^2$ . Further, we let  $\bar{\rho}$  be the  $2 \times 2$ -matrix with all entries equal to  $1/4$ .

For a given hypergraph  $H$  on  $[n]$ , let  $Z_\rho^{(2)}(H)$  be the number of pairs  $(\sigma, \tau)$  of 2-colourings of  $H$  whose overlap matrix is  $\rho$ . Analogously to (3.1), we define the functions  $f_2, g_2 : \mathcal{B} \mapsto \mathbb{R}$  as

$$f_2 : \rho \mapsto \mathcal{H}(\rho) + g_2(\rho) \quad \text{with} \quad g_2(\rho) = \frac{d}{k} \ln \left( 1 - \sum \rho_{i\star}^k - \sum \rho_{\star j}^k + \sum \rho_{ij}^k \right).$$

The following lemma states a formula for  $\mathbb{E} \left[ Z_\rho^{(2)}(\mathcal{H}(n, m)) \right]$  for  $\rho \in \mathcal{B}(n)$  in terms of  $f_2(\rho)$ .

**Lemma 18.** *Let  $d' \in (0, \infty)$  and set*

$$C_n(d, k) = \sqrt{\frac{32}{(\pi n)^3}} \exp \left[ \frac{d(k-1)}{2} \frac{2^k - 3}{(2^{k-1} - 1)^2} \right]. \quad (5.1)$$

Then for  $\rho \in \mathcal{B}(n)$  we have

$$\begin{aligned} \mathbb{E} \left[ Z_\rho^{(2)}(\mathcal{H}(n, m)) \right] &\sim \sqrt{\frac{2\pi}{n^3}} \prod_{i,j=1}^2 (2\pi\rho_{ij})^{-1/2} \exp [nf_2(\rho)] \\ &\exp \left[ \frac{d(k-1)}{2} \frac{\sum \rho_{i\star}^{k-1} - \sum \rho_{i\star}^k + \sum \rho_{\star j}^{k-1} - \sum \rho_{\star j}^k - \sum \rho_{ij}^{k-1} + \sum \rho_{ij}^k}{1 - \sum \rho_{i\star}^k - \sum \rho_{\star j}^k + \sum \rho_{ij}^k} \right]. \end{aligned} \quad (5.2)$$

Moreover, if  $\rho \in \mathcal{B}(n)$  satisfies  $\|\rho - \bar{\rho}\|_2^2 = o(1)$ , then

$$\mathbb{E} \left[ Z_\rho^{(2)}(\mathcal{H}(n, m)) \right] \sim C_n(d, k) \exp [nf_2(\rho)]. \quad (5.3)$$

*Proof.* Let  $\rho = \begin{pmatrix} \rho_{00} & \rho_{10} \\ \rho_{01} & \rho_{11} \end{pmatrix} \in \mathcal{B}(n)$ . Then

$$\begin{aligned} \mathbb{E} \left[ Z_\rho^{(2)}(\mathcal{H}(n, m)) \right] &= \sum_{\sigma, \tau : \rho(\sigma, \tau) = \rho} \mathbb{P} [\sigma, \tau \text{ are colourings of } \mathcal{H}(n, m)] \\ &= \sum_{\sigma, \tau : \rho(\sigma, \tau) = \rho} \left( 1 - \frac{\mathcal{F}(\sigma, \tau)}{N} \right)^m \\ &= \binom{n}{\rho_{00}n, \rho_{01}n, \rho_{10}n, \rho_{11}n} \left( 1 - \frac{\mathcal{F}(\sigma, \tau)}{N} \right)^m. \end{aligned} \quad (5.4)$$

where  $N = \binom{n}{k}$  and  $\mathcal{F}(\sigma, \tau)$  is the total number of possible monochromatic edges under either  $\sigma$  or  $\tau$ . In the last line,  $\sigma$  and  $\tau$  are just two arbitrary fixed 2-colourings with overlap  $\rho$  and the equation is valid because the following computation shows that  $\mathcal{F}(\sigma, \tau)$  only depends on  $\rho$ :

$$\begin{aligned} \mathcal{F}(\sigma, \tau) &= \sum_{i=0}^1 \binom{\rho_{i\star} n}{k} + \sum_{j=0}^1 \binom{\rho_{\star j} n}{k} - \sum_{i,j=0}^1 \binom{\rho_{ij} n}{k} \\ &= N \left[ \sum_{i=0}^1 \rho_{i\star}^k + \sum_{j=0}^1 \rho_{\star j}^k - \sum_{i,j=0}^1 \rho_{ij}^k \right] + \frac{k(k-1)}{2k!} n^{k-1}. \\ &\quad \left[ \sum_{i=0}^1 \rho_{i\star}^k - \sum_{i=0}^1 \rho_{i\star}^{k-1} + \sum_{j=0}^1 \rho_{\star j}^k - \sum_{j=0}^1 \rho_{\star j}^{k-1} - \sum_{i,j=0}^1 \rho_{ij}^k + \sum_{i,j=0}^1 \rho_{ij}^{k-1} \right] + \Theta(n^{k-2}), \end{aligned}$$

yielding

$$\begin{aligned} 1 - \frac{\mathcal{F}(\sigma, \tau)}{N} &= 1 - \sum_{i=0}^1 \rho_{i\star}^k - \sum_{j=0}^1 \rho_{\star j}^k + \sum_{i,j=0}^1 \rho_{ij}^k \\ &\quad - \frac{k(k-1)}{2n} \left[ \sum_{i=0}^1 \rho_{i\star}^k - \sum_{i=0}^1 \rho_{i\star}^{k-1} + \sum_{j=0}^1 \rho_{\star j}^k - \sum_{j=0}^1 \rho_{\star j}^{k-1} - \sum_{i,j=0}^1 \rho_{ij}^k + \sum_{i,j=0}^1 \rho_{ij}^{k-1} \right] + \Theta(n^{-2}). \end{aligned}$$

We proceed as in the proof of Lemma 13 by using that  $\ln(x - \frac{y}{n}) = \ln(x) + \ln(1 - \frac{y}{xn})$  for  $x > 0, \frac{y}{n} < x$  and consequently

$$\begin{aligned} m \ln \left( 1 - \frac{\mathcal{F}(\sigma, \tau)}{N} \right) &= \frac{dn}{k} \left[ \ln \left( 1 - \sum_{i=0}^1 \rho_{i\star}^k - \sum_{j=0}^1 \rho_{\star j}^k + \sum_{i,j=0}^1 \rho_{ij}^k \right) \right. \\ &\quad \left. + \ln \left( 1 - \frac{k(k-1) \sum_{i=0}^1 \rho_{i\star}^k - \sum_{i=0}^1 \rho_{i\star}^{k-1} + \sum_{j=0}^1 \rho_{\star j}^k - \sum_{j=0}^1 \rho_{\star j}^{k-1} - \sum_{i,j=0}^1 \rho_{ij}^k + \sum_{i,j=0}^1 \rho_{ij}^{k-1}}{2n} + \Theta(n^{-2}) \right) \right] \\ &\sim \frac{dn}{k} \ln \left( 1 - \sum_{i=0}^1 \rho_{i\star}^k - \sum_{j=0}^1 \rho_{\star j}^k + \sum_{i,j=0}^1 \rho_{ij}^k \right) \\ &\quad + \frac{d(k-1)}{2} \frac{\sum_{i=0}^1 \rho_{i\star}^{k-1} - \sum_{i=0}^1 \rho_{i\star}^k + \sum_{j=0}^1 \rho_{\star j}^{k-1} - \sum_{j=0}^1 \rho_{\star j}^k - \sum_{i,j=0}^1 \rho_{ij}^{k-1} + \sum_{i,j=0}^1 \rho_{ij}^k}{1 - \sum_{i=0}^1 \rho_{i\star}^k - \sum_{j=0}^1 \rho_{\star j}^k + \sum_{i,j=0}^1 \rho_{ij}^k} + \Theta(n^{-1}). \end{aligned} \quad (5.5)$$

As  $\mathcal{F}(\sigma, \tau)$  does only depend on  $\rho$ , (5.4) becomes Using Stirling's formula, we get the following approximation for the number of colour assignments with overlap  $\rho$ :

$$\binom{n}{\rho_{00}n, \rho_{01}n, \rho_{10}n, \rho_{11}n} \sim \sqrt{2\pi n}^{-3/2} \prod_{i,j=0}^1 (2\pi \rho_{ij})^{-1/2} \exp[n\mathcal{H}(\rho)]. \quad (5.6)$$

Inserting (5.5) and (5.6) into (5.4) completes the proof of (5.2).

Equation (5.3) follows from (5.2) because if  $\|\rho - \bar{\rho}\|_2^2 = o(1)$  then

$$\prod_{i,j=1}^2 (2\pi\rho_{ij})^{-1/2} \sim \frac{4}{\pi^2}$$

and

$$\frac{\sum \rho_{i\star}^{k-1} - \sum \rho_{i\star}^k + \sum \rho_{\star j}^{k-1} - \sum \rho_{\star j}^k - \sum \rho_{ij}^{k-1} + \sum \rho_{ij}^k}{1 - \sum \rho_{i\star}^k - \sum \rho_{\star j}^k + \sum \rho_{ij}^k} \sim \frac{2^k - 3}{(2^{k-1} - 1)^2}. \quad \square$$

## 5.2 Dividing up the interval

Let  $\omega, \nu \in \mathbb{N}$  and  $s \in [\omega\nu]$  and recall  $\rho_{\omega,\nu}^s$  from (2.2). Analogously to the notation in Section 2 we introduce the sets

$$\mathcal{B}_\omega(n) = \left\{ \rho \in \mathcal{B}(n) : \rho_{i\star}, \rho_{\star i} \in \left[ \frac{1}{2} - \frac{\omega}{\sqrt{n}}, \frac{1}{2} + \frac{\omega}{\sqrt{n}} \right) \text{ for } i \in \{0, 1\} \right\}$$

and

$$\mathcal{B}_{\omega,\nu}^s(n) = \left\{ \rho \in \mathcal{B}_\omega(n) : \rho_{i\star}, \rho_{\star i} \in \left[ \rho_{\omega,\nu}^s - \frac{1}{\nu\sqrt{n}}, \rho_{\omega,\nu}^s + \frac{1}{\nu\sqrt{n}} \right) \text{ for } i \in \{0, 1\} \right\},$$

imposing constraints on the overlap matrix  $\rho$  insofar as the colour densities resulting from its projection on each colouring must not deviate too much from  $1/2$  in the set  $\mathcal{B}_\omega(n)$  and from  $\rho_{\omega,\nu}^s$  in the set  $\mathcal{B}_{\omega,\nu}^s(n)$ . By the linearity of expectation, for any  $s \in [\omega\nu]$  we have

$$\mathbb{E} [Z_{\omega,\nu}^s(\mathcal{H}(n, m))^2] = \sum_{\rho \in \mathcal{B}_{\omega,\nu}^s(n)} \mathbb{E} [Z_\rho^{(2)}(\mathcal{H}(n, m))].$$

We are going to show that the expression on the right hand side of this equation is dominated by the contributions with  $\rho$  “close to”  $\bar{\rho}$  in terms of the euclidian norm. More precisely, for  $\eta > 0$  we introduce the set

$$\mathcal{B}_{\omega,\nu,\eta}^s(n) = \{ \rho \in \mathcal{B}_{\omega,\nu}^s(n) : \|\rho - \bar{\rho}\|_2 \leq \eta \}$$

and define

$$Z_{\omega,\nu,\eta}^s(\mathcal{H}(n, m)) = \sum_{\rho \in \mathcal{B}_{\omega,\nu,\eta}^s(n)} Z_\rho^{(2)}(\mathcal{H}(n, m)).$$

The following proposition reveals that it is sufficient to consider overlap matrices  $\rho$  such that  $\|\rho - \bar{\rho}\|_2 \leq n^{-3/8}$ . Here, the number  $3/8$  is somewhat arbitrary, any number smaller than  $1/2$  would do.

**Proposition 19.** *Let  $k \geq 3$  and  $\omega, \nu \in \mathbb{N}$ . If  $d'/k < 2^{k-1} \ln 2 - 2$ , then for every  $s \in [\omega\nu]$  we have*

$$\mathbb{E} [Z_{\omega,\nu}^s(\mathcal{H}(n, m))^2] \sim \mathbb{E} [Z_{\omega,\nu,n^{-3/8}}^s(\mathcal{H}(n, m))].$$



To prove this proposition, we need the following lemma.

**Lemma 20.** *Let  $d/k < 2^{k-1} \ln 2 - 2$  and  $C_n(d, k)$  as defined in Lemma 18. Set*

$$B(d, k) = 4 \left( 1 - \frac{d(k-1)}{2^{k-1} - 1} \right).$$

1. *If  $\rho \in \mathcal{B}_\omega(n)$  satisfies  $\|\rho - \bar{\rho}\|_2 \leq n^{-3/8}$  then*

$$\mathbb{E} [Z_\rho^{(2)}(\mathcal{H}(n, m))] \sim C_n(d, k) \exp \left[ n f_2(\bar{\rho}) - n \frac{B(d, k)}{2} \|\rho - \bar{\rho}\|_2^2 \right]. \quad (5.7)$$

2. *There exists  $A = A(d, k) > 0$  such that if  $\rho \in \mathcal{B}_\omega(n)$  satisfies  $\|\rho - \bar{\rho}\|_2 > n^{-3/8}$ , then*

$$\mathbb{E} [Z_\rho^{(2)}(\mathcal{H}(n, m))] = O \left( \exp [n f_2(\bar{\rho}) - A n^{1/4}] \right). \quad (5.8)$$

*Proof.* To prove (5.7), we observe that if  $\rho \in \mathcal{B}_\omega(n)$  satisfies  $\|\rho - \bar{\rho}\|_2 \leq n^{-3/8}$ , by Taylor expansion around  $\bar{\rho}$  (where  $\mathcal{H}$  and  $g_2$  are maximized) we obtain

$$\mathcal{H}(\rho) = \mathcal{H}(\bar{\rho}) - 2\|\rho - \bar{\rho}\|_2^2 + o(n^{-1}) \quad \text{and} \quad (5.9)$$

$$g_2(\rho) = g_2(\bar{\rho}) - \frac{2d(k-1)}{2^{k-1} - 1} \|\rho - \bar{\rho}\|_2^2 + o(n^{-1}). \quad (5.10)$$

Inserting this into (5.3) yields (5.7).

To prove (5.8), we distinguish two cases.

**Case 1:**  $\|\rho - \bar{\rho}\|_2 = o(1)$ : We observe that similarly to (5.9) and (5.10) there exists a constant  $A = A(d, k) > 0$  such that

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

Hence, if  $\|\rho - \bar{\rho}\|_2 > n^{-3/8}$  and  $\|\rho - \bar{\rho}\|_2 = o(1)$ , then

$$\mathbb{E} [Z_\rho^{(2)}(\mathcal{H}(n, m))] = O(n^{-3/2}) \exp [n f_2(\rho)] \leq \exp [n f_2(\bar{\rho}) - A n^{1/4}]. \quad (5.11)$$

**Case 2:**  $\|\rho - \bar{\rho}\|_2 = c$  where  $c > 0$  is a constant independent of  $n$ : We consider the function  $\bar{f}_2 : [0, \frac{1}{2}] \mapsto \mathbb{R}$  that results from  $f_2$  by setting  $\rho_{i\star} = \rho_{\star i} = 1/2$ . This function was introduced by Achlioptas and Moore [3] and has been studied at different places in the literature on random hypergraph 2-colouring. The following lemma quantifies the largest possible deviation of  $f_2$  and  $\bar{f}_2$ .

**Lemma 21.** *Let  $\bar{f}_2 : [0, 1] \rightarrow \mathbb{R}$  be defined as*

$$\bar{f}_2(\rho) = \ln 2 + \mathcal{H}(2\rho) + \frac{d}{k} \ln \left( 1 - 2^{2-k} + 2\rho^k + 2 \left( \frac{1}{2} - \rho \right)^k \right).$$

*Then for  $\rho = (\rho_{ij}) \in \mathcal{B}_\omega(n)$  we have*

$$\exp [n f_2(\rho)] \sim \exp [n \bar{f}_2(\rho_{00}) + O(\omega^2)].$$

*Proof.* For  $\rho \in \mathcal{B}_\omega(n)$  we consider the function

$$\zeta(\rho) = f_2(\rho) - \bar{f}_2(\rho_{00})$$

and approximate  $\zeta(\rho)$  by a Taylor expansion around  $\rho = \bar{\rho}$ . As  $f_2(\bar{\rho}) = \bar{f}_2(\bar{\rho}_{00})$  and  $\frac{\partial f_2}{\partial \rho_{ij}}(\bar{\rho}) = 0$  for  $i, j \in \{0, 1\}$  and  $\bar{f}'_2(\bar{\rho}_{00}) = 0$ , we have  $\zeta(\rho) = C \cdot \|\rho - \bar{\rho}\|_2^2 = O\left(\frac{\omega}{\sqrt{n}}\right)$  for some constant  $C$ . Thus,

$$\max_{\rho \in \mathcal{B}_\omega(n)} |\zeta(\rho)| = O\left(\frac{\omega^2}{n}\right),$$

yielding the assertion.  $\square$

In [6, Lemma 4.11] the function  $\bar{f}_2$  is analysed and it is shown that in the regime  $d/k \leq 2^{k-1} \ln 2 - 2$  it takes its *global* maximum at  $\rho = \bar{\rho}$  and  $\bar{f}_2(\rho) < \bar{f}_2(\bar{\rho})$  for all  $\rho \in [0, \frac{1}{2}]$  with  $\rho \neq \bar{\rho}$  independent of  $n$ . Combining this with Lemma 21 we find that there exists a constant  $A' = A'(d, k) > 0$  such that

$$f_2(\rho) = f_2(\bar{\rho}) - A' + O\left(\frac{\omega^2}{n}\right),$$

where we used that  $f_2(\bar{\rho}) = \bar{f}_2(\bar{\rho})$ .

Thus,

$$\mathbb{E} [Z_\rho^{(2)}(\mathcal{H}(n, m))] = O(n^{-3/2}) \exp[nf_2(\rho)] \leq \exp[nf_2(\bar{\rho}) - A'n + O(\omega^2)]. \quad (5.12)$$

As  $\exp[nf_2(\bar{\rho}) - A'n + O(\omega^2)] = o(\exp[nf_2(\bar{\rho}) - An^{1/4}])$ , equation (5.12) together with (5.11) completes the proof of (5.8).  $\square$

*Proof of Proposition 19.* We let  $s \in [\omega\nu]$ . For a  $\hat{\rho} \in \mathcal{B}_{\omega, \nu, n-3/8}^s(n)$  we have  $\|\hat{\rho} - \bar{\rho}\|_2 = O\left(\frac{\omega}{\sqrt{n}}\right)$  and obtain from the first part of Lemma 20 that

$$\mathbb{E} [Z_{\omega, \nu, n-3/8}^{s(2)}(\mathcal{H}(n, m))^2] \geq \mathbb{E} [Z_{\hat{\rho}}^{(2)}(\mathcal{H}(n, m))] \sim C_n(d, k) \exp[nf_2(\bar{\rho}) + O(\omega^2)]. \quad (5.13)$$

On the other hand, because  $|\mathcal{B}_{\omega, \nu}^s(n)|$  is bounded by a polynomial in  $n$ , the second part of Lemma 20 yields

$$\sum_{\rho \in \mathcal{B}_{\omega, \nu}^s(n): \|\rho - \bar{\rho}\|_2 > n^{-3/8}} \mathbb{E} [Z_\rho^{(2)}(\mathcal{H}(n, m))] = O(\exp[nf_2(\bar{\rho}) - An^{1/4} + O(\ln n)]). \quad (5.14)$$

Combining (5.13) and (5.14), we obtain

$$\mathbb{E} [Z_{\omega, \nu}^s(\mathcal{H}(n, m))^2] \sim \sum_{\rho \in \mathcal{B}_{\omega, \nu, n-3/8}^s(n)} \mathbb{E} [Z_\rho^{(2)}(\mathcal{H}(n, m))] = \mathbb{E} [Z_{\omega, \nu, n-3/8}^{s(2)}(\mathcal{H}(n, m))]$$

as claimed.  $\square$

### 5.3 The leading constant

In this section we compute the contribution of overlap matrices  $\rho \in \mathcal{B}_{\omega, \nu, n^{-3/8}}^s(n)$ . In a first step we show that for  $\rho \in \mathcal{B}_{\omega, \nu, n^{-3/8}}^s(n)$  we can approximate  $f_2$  by a function  $f_2^s$  that results from  $f_2$  by (approximately) fixing the marginals  $\rho_{i\star}, \rho_{\star j}$  for  $i, j \in \{0, 1\}$ .

**Lemma 22.** *Let  $k \geq 3, \omega, \nu \in \mathbb{N}$  and  $C_n(d, k)$  as in (5.1). For  $s \in [\omega\nu]$  remember  $\rho_{\omega, \nu}^s$  from (2.2). Let  $f_2^s : \mathcal{B} \rightarrow \mathbb{R}$  be defined as*

$$f_2^s : \rho \mapsto \mathcal{H}(\rho) + \frac{d}{k} \ln \left( 1 - 2\rho_{\omega, \nu}^s{}^k - 2(1 - \rho_{\omega, \nu}^s)^k + \sum_{i, j=0}^1 \rho_{ij} \right).$$

Then for  $\rho \in \mathcal{B}_{\omega, \nu, n^{-3/8}}^s(n)$  it holds that

$$\mathbb{E} [Z_\rho^{(2)}(\mathcal{H}(n, m))] \sim C_n(d, k) \exp \left[ n f_2^s(\rho) + O\left(\frac{\omega}{\nu}\right) \right].$$

*Proof.* Equation (5.3) of Lemma 18 yields that

$$\mathbb{E} [Z_\rho^{(2)}(\mathcal{H}(n, m))] \sim C_n(d, k) \exp [n f_2(\rho)]. \quad (5.15)$$

Analogously to the proof of Lemma 21 we define

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

To bound  $\zeta^s(\rho)$  from above for all  $\rho \in \mathcal{B}_{\omega, \nu, n^{-3/8}}^s(n)$ , we observe that we can express the function  $f_2$  by setting  $\rho_{0\star} = \rho_{\omega, \nu}^s + \alpha$  and  $\rho_{\star 0} = \rho_{\omega, \nu}^s + \beta$ , where  $|\alpha|, |\beta| \leq \frac{1}{\nu\sqrt{n}}$  and thus

$$f_2 : \rho \mapsto \mathcal{H}(\rho) + \frac{d}{k} \ln \left( 1 - (\rho_{\omega, \nu}^s + \alpha)^k - (\rho_{\omega, \nu}^s + \beta)^k - (1 - \rho_{\omega, \nu}^s - \alpha)^k - (1 - \rho_{\omega, \nu}^s - \beta)^k + \sum_{i, j=0}^1 \rho_{ij} \right).$$

As we are only interested in the difference between  $f_2$  and  $f_2^s$ , we can reparametrise  $\zeta^s$  as

$$\zeta^s(\alpha, \beta) = \frac{d}{k} \ln \left( \frac{1 - (\rho_{\omega, \nu}^s + \alpha)^k - (\rho_{\omega, \nu}^s + \beta)^k - (1 - \rho_{\omega, \nu}^s - \alpha)^k - (1 - \rho_{\omega, \nu}^s - \beta)^k + \sum_{i, j=0}^1 \rho_{ij}}{1 - 2\rho_{\omega, \nu}^s{}^k - 2(1 - \rho_{\omega, \nu}^s)^k + \sum_{i, j=0}^1 \rho_{ij}} \right).$$

Differentiating and simplifying the expression yields  $\frac{\partial \zeta^s}{\partial \alpha}(\alpha, \beta), \frac{\partial \zeta^s}{\partial \beta}(\alpha, \beta) = O\left(\frac{\omega}{\sqrt{n}}\right)$ . As we are interested in  $\rho \in \mathcal{B}_{\omega, \nu, n^{-3/8}}^s(n)$  and  $|\mathcal{B}_{\omega, \nu, n^{-3/8}}^s(n)| \leq \frac{2}{\nu\sqrt{n}}$  according to the fundamental theorem of calculus it follows for every  $s \in [\omega\nu]$  that

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

Combining this with (5.15) yields the assertion. □

**Proposition 23.** Let  $k \geq 3, \omega, \nu \in \mathbb{N}$  and  $d'(k-1) < (2^{k-1} - 1)^2$ . Then for all  $s \in [\omega\nu]$  we have

$$\mathbb{E} \left[ Z_{\omega, \nu, n^{-3/8}}^{s(2)}(\mathcal{H}(n, m)) \right] \sim_{\nu} \left( |\mathcal{A}_{\omega, \nu}^s(n)| \sqrt{\frac{2}{\pi n}} \exp \left[ n f_1(\rho_{\omega, \nu}^s) \right] \right)^2 \cdot \exp \left[ \frac{d(k-1)}{2} \frac{2^k - 3}{(2^{k-1} - 1)^2} \right] \left( 1 - \frac{d(k-1)}{(2^{k-1} - 1)^2} \right)^{-1/2}.$$

*Proof.* By Lemma 22 we know that for  $\rho \in \mathcal{B}_{\omega, \nu, n^{-3/8}}^s(n)$  we have

$$\mathbb{E} \left[ Z_{\rho}^{(2)}(\mathcal{H}(n, m)) \right] \sim C_n(d, k) \exp \left[ n f_2^s(\rho) + O\left(\frac{\omega}{\nu}\right) \right]. \quad (5.16)$$

A Taylor expansion of  $f_2^s(\rho)$  around

$$\rho^s = \begin{pmatrix} \rho_{\omega, \nu}^{s^2} & \rho_{\omega, \nu}^s(1 - \rho_{\omega, \nu}^s) \\ (1 - \rho_{\omega, \nu}^s) \rho_{\omega, \nu}^s & (1 - \rho_{\omega, \nu}^s)^2 \end{pmatrix}$$

while setting  $D(d, k) = 4 \left( 1 - \frac{d(k-1)}{(2^{k-1} - 1)^2} \right)$  yields

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

Combining this with (5.16) we find that

$$\mathbb{E} \left[ Z_{\rho}^{(2)}(\mathcal{H}(n, m)) \right] \sim C_n(d, k) \exp \left[ n f_2^s(\rho^s) + \Theta(\omega) \|\rho - \rho^s\|_2 - n \frac{D(d, k)}{2} \|\rho - \rho^s\|_2^2 + O\left(\frac{\omega}{\nu}\right) \right]. \quad (5.17)$$

For  $\rho^0, \rho^1 \in \mathcal{B}_{\omega, \nu}^s(n)$ , we introduce the set of overlap matrices

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

In particular,  $\mathcal{B}_{\omega, \nu, n^{-3/8}}^s(n, \rho^0, \rho^1)$  contains the “product” overlap  $\rho^0 \otimes \rho^1$ , which is defined by  $(\rho^0 \otimes \rho^1)_{ij} = \rho_i^0 \rho_j^1$ . With these definitions we see that

$$\mathbb{E} \left[ Z_{\omega, \nu, n^{-3/8}}^{s(2)}(\mathcal{H}(n, m)) \right] = \sum_{\rho^0, \rho^1 \in \mathcal{B}_{\omega, \nu}^s(n)} \sum_{\rho \in \mathcal{B}_{\omega, \nu, n^{-3/8}}^s(n, \rho^0, \rho^1)} \mathbb{E} \left[ Z_{\rho}^{(2)}(\mathcal{H}(n, m)) \right]. \quad (5.18)$$

Let us fix from now on two colour densities  $\rho^0, \rho^1 \in \mathcal{B}_{\omega, \nu}^s(n)$ . We simplify the notation by setting

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

Thus, we are going to evaluate

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

We define the set  $\mathcal{E}_n = \{\boldsymbol{\varepsilon} = (\varepsilon, -\varepsilon, -\varepsilon, \varepsilon), \varepsilon \in \frac{1}{n}\mathbb{Z}, 0 \leq \varepsilon \leq 1\}$ . Then for each  $\rho \in \widehat{\mathcal{B}}$  we can find  $\boldsymbol{\varepsilon} \in \mathcal{E}_n$  such that

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

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

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

As  $\|\widehat{\rho} - \rho^s\|_2 \leq \frac{1}{\nu\sqrt{n}}$  and for  $\nu \rightarrow \infty$  it holds that  $\frac{1}{\nu\sqrt{n}} = o(n^{-1/2})$ , in this case we have

$$\|\rho - \rho^s\|_2 = \|\boldsymbol{\varepsilon}\|_2 + o(n^{-1/2}). \quad (5.19)$$

Observing that  $f_2^s(\rho^s) = (f_1(\rho_{\omega, \nu}^s))^2$  and inserting (5.19) into (5.17), we find

$$\begin{aligned} \mathcal{S}_1 &\sim_\nu \sum_{\rho \in \widehat{\mathcal{B}}} C_n(d, k) \exp \left[ n f_2^s(\rho^s) - n \frac{D(d, k)}{2} \|\boldsymbol{\varepsilon}\|_2^2 + o(n^{1/2}) \|\boldsymbol{\varepsilon}\|_2 + o(1) \right] \\ &\sim_\nu C_n(d, k) \exp [2n f_1^s(\rho_{\omega, \nu}^s)] \sum_{\rho \in \widehat{\mathcal{B}}} \exp \left[ -n \frac{D(d, k)}{2} \|\boldsymbol{\varepsilon}\|_2^2 + o(n^{1/2}) \|\boldsymbol{\varepsilon}\|_2 \right]. \end{aligned} \quad (5.20)$$

It follows from the definition of  $\widehat{\mathcal{B}}$  that

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

As

$$\begin{aligned} \mathcal{S}_2 &\sim_\nu C_n(d, k) \exp [n f_2^s(\rho^s)] \sum_{\boldsymbol{\varepsilon} \in \mathcal{E}_n, \|\boldsymbol{\varepsilon}\|_2 > n^{-3/8}/2} \exp \left[ -n \frac{D(d, k)}{2} \|\boldsymbol{\varepsilon}\|_2^2 (1 + o(1)) \right] \\ &\leq C_n(d, k) \exp [n f_2^s(\rho^s)] O(n) \exp \left[ -\frac{D(d, k)}{8} n^{1/4} \right], \end{aligned}$$

equation (5.20) yields  $\lim_{\nu \rightarrow \infty} \lim_{n \rightarrow \infty} \mathcal{S}_2/\mathcal{S}_1 = 0$  and we see that  $\boldsymbol{\varepsilon} \in \mathcal{E}_n$  with  $\|\boldsymbol{\varepsilon}\|_2 > n^{-3/8}/2$  do only contribute negligibly. Thus, we conclude, using the formula of Euler-Maclaurin and a Gaussian integration, that

$$\begin{aligned} \mathcal{S}_1 &\sim_\nu C_n(d, k) \exp [2n f_1^s(\rho_{\omega, \nu}^s)] \sum_{\boldsymbol{\varepsilon} \in \mathcal{E}_n} \exp \left[ -n \frac{D(d, k)}{2} \|\boldsymbol{\varepsilon}\|_2^2 + o(n^{1/2}) \|\boldsymbol{\varepsilon}\|_2 \right] \\ &\sim_\nu C_n(d, k) \exp [2n f_1^s(\rho_{\omega, \nu}^s)] n \int \exp \left[ -n \frac{D(d, k)}{8} \varepsilon^2 + o(n^{1/2}) \varepsilon \right] d\varepsilon \\ &\sim_\nu C_n(d, k) \exp [2n f_1^s(\rho_{\omega, \nu}^s)] \sqrt{\frac{\pi n}{8}} \left( 1 - \frac{d(k-1)}{(2^{k-1}-1)^2} \right)^{-1/2}. \end{aligned} \quad (5.21)$$

In particular, the last expression is independent of the choice of the vectors  $\rho^0, \rho^1$  that defined  $\widehat{\mathcal{B}}$ . Therefore, substituting (5.21) in the decomposition (5.18) completes the proof.  $\square$

*Proof of Proposition 9.* From (4.1) we remember that

$$\exp \left[ \sum_{l \geq 2} \lambda_l \delta_l^2 \right] = \exp \left[ -\frac{d(k-1)}{2} \frac{1}{(2^{k-1}-1)^2} \right] \left( 1 - \frac{d(k-1)}{(2^{k-1}-1)^2} \right)^{-1/2}. \quad (5.22)$$

To prove Proposition 9 we combine Lemma 15 with Propositions 19 and 23 yielding

$$\begin{aligned} \frac{\mathbb{E} [Z_{\omega, \nu}^s(\mathcal{H}(n, m))^2]}{\mathbb{E} [Z_{\omega, \nu}^s(\mathcal{H}(n, m))]^2} &\sim_{\nu} \exp \left[ \frac{d(k-1)}{2} \left( \frac{2^k-3}{(2^{k-1}-1)^2} - \frac{2}{2^{k-1}-1} \right) \right] \left( 1 - \frac{d(k-1)}{(2^{k-1}-1)^2} \right)^{-1/2} \\ &= \exp \left[ -\frac{d(k-1)}{2} \frac{1}{(2^{k-1}-1)^2} \right] \left( 1 - \frac{d(k-1)}{(2^{k-1}-1)^2} \right)^{-1/2}. \end{aligned} \quad (5.23)$$

Combining equations (5.22) and (5.23) completes the proof.  $\square$

## Acknowledgements

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

## References

- [1] D. Achlioptas, A. Coja-Oghlan: Algorithmic barriers from phase transitions. *Proc. 49th FOCS* (2008) 793–802.
- [2] D. Achlioptas, J. H. Kim, M. Krivelevich, P. Tetali: Two-coloring random hypergraphs. *Random Structures and Algorithms* **20(2)** (2002) 249–259.
- [3] D. Achlioptas, C. Moore: Random  $k$ -SAT: two moments suffice to cross a sharp threshold. *SIAM Journal on Computing* **36(3)** (2006) 740–762.
- [4] V. Bapst, A. Coja-Oghlan, C. Efthymiou: Planting colourings silently. *Combinatorics, Probability and Computing* **26(3)** (2017) 338–366.
- [5] V. Bapst, A. Coja-Oghlan, S. Hetterich, F. Rassmann, Dan Vilenchik: The condensation phase transition in random graph coloring. *Communications in Mathematical Physics* **341(2)** (2016) 543–606.
- [6] V. Bapst, A. Coja-Oghlan, F. Rassmann: A positive temperature phase transition in random hypergraph 2-colouring. *Annals of Applied Probability* **26(3)** (2016), 1362–1406.
- [7] B. Bollobás: Random graphs. 2nd edition. Cambridge University Press (2001).
- [8] A. Coja-Oghlan, K. Panagiotou: Catching the  $k$ -NAESAT threshold. *Proc. 44th STOC* (2012), 899–908.

- [9] A. Coja-Oghlan, N. Wormald: The number of satisfying assignments of random regular  $k$ -SAT formulas. Manuscript in preparation.
- [10] A. Coja-Oghlan, L. Zdeborová: The condensation transition in random hypergraph 2-coloring. *Proc. 23rd SODA* (2012) 241–250.
- [11] L. Dall’Asta, A. Ramezanpour, R. Zecchina: Entropy landscape and non-Gibbs solution in constraint satisfaction problems. *Phys. Rev. E* **77** (2008) 031118.
- [12] S. Janson: Random regular graphs: asymptotic distributions and contiguity. *Combinatorics, Probability and Computing* **4**(4) (1995) 369–405.
- [13] F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, L. Zdeborová: Gibbs states and the set of solutions of random constraint satisfaction problems. *Proc. National Academy of Sciences* **104**(25) (2007) 10318–10323.
- [14] F. Krzakala, L. Zdeborová: Hiding quiet solutions in random constraint satisfaction problems. *Phys. Rev. Lett.* **102** (2009) 238701.
- [15] M. Molloy: The freezing threshold for  $k$ -colourings of a random graph. *Proc. 43rd STOC* (2012) 921–930.
- [16] M. Molloy, R. Restrepo: Frozen variables in random boolean constraint satisfaction problems. *Proc. 24th SODA* (2013) 1306–1318.
- [17] A. Montanari, R. Restrepo, P. Tetali: Reconstruction and clustering in random constraint satisfaction problems. *SIAM Journal on Discrete Mathematics* **25**(2) (2011) 771–808.
- [18] C. Moore: The phase transition in random regular exact cover. *Annales l’Institut Henri Poincaré D, Combinatorics, Physics and their Interactions* **3**(3) (2016) 349–362.
- [19] J. Banks, C. Moore, J Neeman, P. Netrapalli: Information-theoretic thresholds for community detection in sparse networks. *Proc. 29th COLT* (2016) 383–416.
- [20] R. Robinson, N. Wormald: Almost all regular graphs are Hamiltonian. *Random Structures and Algorithms* **5**(2) (1994) 363–374.