2013 IEEE 54th Annual Symposium on Foundations of Computer Science

Chasing the k-colorability threshold

Amin Coja-Oghlan
Goethe University, Mathematics Institute
10 Robert Mayer St, Frankfurt 60325, Germany
acoghlan@math.uni-frankfurt.de

Dan Vilenchik
Facutly of Mathematics & Computer Science
The Weizamnn Institute, Rehovot, Israel
dan.vilenchik@weizmann.ac.il

***Abstract*—In this paper we establish a substantially improved lower bound on the k-colorability threshold of the random graph $G(n,m)$ with n vertices and m edges. The new lower bound is ≈ 1.39 less than the $2k\ln k-\ln k$ first-moment upper bound (and ≈ 0.39 less than the $2k\ln k-\ln k-1$ physics conjecture). By comparison, the best previous bounds left a gap of about $2+\ln k$, unbounded in terms of the number of colors [Achlioptas, Naor: STOC 2004]. Furthermore, we prove that, in a precise sense, our lower bound marks the so-called *condensation phase transition* predicted on the basis of physics arguments [Krzkala et al.: PNAS 2007]. Our proof technique is a novel approach to the second moment method, inspired by physics conjectures on the geometry of the set of k-colorings of the random graph.**

***Keywords*-random structures, phase transitions, graph coloring.**

I. Introduction

Let $G(n,m)$ be the random graph on the vertex set $V=\{1,\ldots,n\}$ with m edges. Unless specified otherwise, we assume that $m=\lceil dn/2\rceil$ for a number $d>0$ that remains fixed as $n\to\infty$ and that $k\geq 3$ is an n-independent integer. We say that $G(n,m)$ has a property $\mathcal{E}$ with high probability *('w.h.p.') if $\lim_{n\to\infty}\mathrm{P}\left[G(n,m)\in\mathcal{E}\right]=1$.*

The theory of random graphs started with the famous 1960 article by Erdős and Rényi [18], in which the existence of a *phase transition* was established by proving the sudden emergence of a giant component at $d\sim 1$. Erdős and Rényi also set the agenda for future research by posing a number of questions on further phase transitions. To date, all but one of these questions have been answered. The last open one concerns the chromatic number of $G(n,m)$.[1] More precisely, to date it is widely conjecture that there is a sharp phase transition for k-colorability for any $k\geq 3$ (e.g., [1]).

Achlioptas and Friedgut [1] showed that for any fixed $k\geq 3$ there exists a *sharp threshold sequence* $d_{k-\mathrm{col}}=d_{k-\mathrm{col}}(n)$. This sequence is such that for any $\varepsilon>0$ the random graph $G(n,m)$ is k-colorable w.h.p. if the average degree is less than $(1-\varepsilon)d_{k-\mathrm{col}}(n)$, but there is no k-coloring w.h.p. for average degrees greater than $(1+\varepsilon)d_{k-\mathrm{col}}(n)$.[2] While this is a pure existence result, in a landmark paper Achlioptas and Naor [6] used the second moment method to prove that

$$d_{k-\mathrm{col}}\geq d_{k,\mathrm{AN}}=2(k-1)\ln(k-1)= \\ =2k\ln k-2\ln k-2+o_k(1), \tag{1}$$

where the $o_k(1)$ hides a term that tends to zero for large k. By comparison, a simple "first moment" calculation shows

$$d_{k-\mathrm{col}}\leq d_{k,\mathrm{first}}=2k\ln k-\ln k. \tag{2}$$

This leaves a gap of about $2+\ln k$, a function that diverges as k gets larger.

Independently of the rigorous work, the random graph coloring problem has been studied in statistical physics under the snappy title of "diluted mean-field Potts antiferromagnet". In fact, over the past decade physicists have developed a deep but mathematically non-rigorous formalism called the "cavity method" for locating phase transitions in discrete structures [26], [27]. According to the cavity method [22], [28], [29], [31],

$$d_{k-\mathrm{col}}=2k\ln k-\ln k-1+o_k(1). \tag{3}$$

In addition, the cavity method has inspired new message passing algorithms called *Belief/Survey Propagation Guided Decimation* [11], [26].

Recently, there has been progress in verifying the physicists' predictions on the phase transitions in *binary* problems. For instance, the current gap between the best lower and upper bounds in random k-SAT is ≈ 0.19 [13]. In random k-NAESAT the gap is as tiny as $2^{-(1-o_k(1))k}$ [12], [15]. This leaves graph k-coloring as the single most prominent example with a gap that is *unbounded* in terms of k.

This large gap remains because the techniques of [12], [13], [15] do not extend easily beyond binary problems. More specifically, the presence of k possible colors (in physics jargon, 'spins') per vertex *dramatically* complicates the use of the second moment method, the mainstay for proving lower bounds.

Here we develop a new approach to the second moment method in the presence of more than two spins. This

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.

[1] We owe this observation to Charilaos Efthymiou.

[2] In order to prove that there actually is a sharp k-colorability threshold one would have to show that $d_{k-\mathrm{col}}(n)$ converges.

0272-5428/13 $26.00 © 2013 IEEE
DOI 10.1109/FOCS.2013.48

approach, based on an analysis of the geometry of the set of k-colorings and a local variations argument, is directly inspired by physics ideas. We view this technique as an important step towards the long-term goal of providing a rigorous foundation for the 'cavity method'. Our main result is summarized in the following theorem.

Theorem 1.1: The k-colorability threshold satisfies $d_{k-\mathrm{col}} \geq d_{k,\mathrm{cond}} - o_k(1)$, with

$$d_{k,\mathrm{cond}} = 2k \ln k - \ln k - 2 \ln 2. \tag{4}$$

The gap between the new lower bound (4) and the elementary upper bound (2) is an additive $2 \ln 2 + o_k(1) \approx 1.39$, rather than a function that grows with k. Moreover, the gap between (4) and the physics prediction (3) is a mere $2 \ln 2 - 1 \approx 0.39$.

In fact, Theorem 1.1 determines the chromatic number of $G(n, m)$ exactly for "most" average degrees d. More precisely, let us say that a (measurable) set $A \subset \mathbf{R}_{\geq 0}$ has ***density*** α if $\lim_{z \to \infty} \frac{1}{z} \int_0^z \mathbf{1}_A = \alpha$, where $\mathbf{1}_A$ is the indicator of A.

Corollary 1.2: There exist a set $\mathcal{A} \subset \mathbf{R}_{\geq 0}$ of density 1 and a function $F : \mathcal{A} \to \mathbf{Z}_{\geq 0}$ such that for all average degrees $d \in \mathcal{A}$ we have $\chi(G(n, m)) = F(d)$ w.h.p.

To be specific, $\mathcal{A}$ is the union of the intervals $(d_{k-1,\mathrm{first}}, d_{k,\mathrm{cond}} - o_k(1))$ where Theorem 1.1 and (2) show that $G(n, m)$ is k-colorable but not $k-1$-colorable w.h.p.

Corollary 1.2 improves a result from [6], who used (1) and (2) to determine the chromatic number on a set $\mathcal{A}'$ of density $\frac{1}{2}$. Furthermore, Corollary 1.2 answers a question of Alon and Krivelevich whether the chromatic number of $G(n, m)$ is concentrated on a single integer for most d "in an appropriately defined sense" [8] in the case $m = O(n)$.[3]

Finally, why doesn't our second moment argument determine the threshold $d_{k-\mathrm{col}}$ precisely? According to the cavity method, the demise of the second moment method at $d_{k,\mathrm{cond}}$ is due to a phase transition called *condensation* that marks a change in the geometry of the set of k-colorings. According to the physics predictions, when the average degree is smaller than $d_{k,\mathrm{cond}} - o_k(1)$, the k-colorings are arranged in well-separated "clusters", each comprising only an exponentially small fraction of the total number of k-colorings. As the average degree crosses $d_{k,\mathrm{cond}} + o_k(1)$, this formation changes: the size of the largest cluster has the same order of magnitude as the total number of k-colorings w.h.p. In effect, a bounded number of clusters dominate the entire set of k-colorings. Hence the term "condensation".

[3] A proof that the threshold sequence $d_{k-\mathrm{col}}(n)$ converges would imply a one-point concentration result for the chromatic number outside a countable set of average degrees. However, the known non-uniform sharp threshold result does not.

Based on our techniques we can verify that indeed, in a precise sense, a phase transition occurs at $d_{k,\mathrm{cond}}$ (see Proposition 2.1 below). But before we come to that we need to discuss the second moment method and its relationship to the physics predictions.

II. Graph coloring and the second moment method

Most of the current results on phase transitions in random constraint satisfaction problems are based on the *second moment method*. Suppose that $Z = Z(G(n, m)) \geq 0$ is a random variable such that $Z(G) > 0$ only if G is k-colorable. To prove that $d_{k-\mathrm{col}} \geq d - o(1)$, it suffices to show that $\liminf \mathrm{P}[Z(G(n, m)) > 0] > 0$, and then use the sharp threshold result from [1]. To show that $\liminf \mathrm{P}[Z(G(n, m)) > 0] > 0$, we prove that there is a number $C = C(d, k) > 0$ that may depend on the average degree d such that

$$0 < \mathrm{E}\left[Z^2\right] \leq C \cdot \mathrm{E}\left[Z\right]^2, \tag{5}$$

and use the *Paley-Zygmund inequality*

$$\mathrm{P}\left[Z \geq \frac{1}{2}\mathrm{E}\left[Z\right]\right] \geq \frac{\mathrm{E}\left[Z\right]^2}{4\,\mathrm{E}\left[Z^2\right]}. \tag{6}$$

A. Balanced colorings and the Birkhoff polytope

Perhaps the most obvious choice of random variable is the total number $Z_{k-\mathrm{col}}$ of k-colorings of $G(n, m)$. However, following [6] we are going to work with a particular type of colorings to simplify our calculations: we call a map $\sigma : V \to [k]$ ***balanced*** if $|\sigma^{-1}(i)| = n/k$ for all $i \in [k]$.[4] Let $\mathcal{B}$ be the set of all such balanced maps and let $Z_{k,\mathrm{bal}}$ be the number of balanced k-colorings of $G(n, m)$. As it turns out, the second moment bound (5) does not hold for either $Z_{k-\mathrm{col}}$ or $Z_{k,\mathrm{bal}}$ in the entire range $0 < d < d_{k,\mathrm{cond}}$. To remedy this problem, we need to understand its origin. Thus, let us sketch the approach taken in [6] in the following paragraphs.

To get started, we compute the first moment. More precisely, since the first moment scales exponentially with n, we estimate its logarithm. By Stirling's formula the number of balanced $\sigma : V \to [k]$ is $|\mathcal{B}| = k^{n-o(n)}$. Furthermore, for any balanced σ a random edge is bichromatic with probability $1 - 1/k + O(1/n)$. Since $G(n, m)$ consists of $m \sim dn/2$ nearly independent random edges, we obtain

$$\frac{1}{n} \ln \mathrm{E}\left[Z_{k,\mathrm{bal}}\right] \quad \sim \quad \ln k + \frac{d}{2} \ln(1 - 1/k). \tag{7}$$

Working out the second moment is not quite so straightforward. Since $\mathrm{E}[Z_{k,\mathrm{bal}}^2]$ is nothing but the expected number of *pairs* of balanced k-colorings, we need to compute the probability that two balanced $\sigma, \tau \in \mathcal{B}$ *simultaneously* happen to be k-colorings of $G(n, m)$. Of course, this probability will depend on how "similar" σ, τ are.

[4] We assume for the sake of presentation that n is divisible by k. Otherwise one requires $||\sigma^{-1}(i)| - n/k| \leq 1$ instead.

In *binary* problems such as k-SAT similarity can be quantified just by the number of variables on which the two assignments coincide. However, for $\sigma, \tau \in \mathcal{B}$ knowing the number of vertices that receive the same color is insufficient. For instance, τ could be obtained from σ simply by permuting the color classes, in which case σ, τ are indistinguishable as far as the k-coloring problem goes without coloring a single vertex the same. Moreover, it is easy to construct examples where even applying the "obvious" permutation does not help. Therefore, we introduce the $k \times k$ ***overlap matrix*** $\rho(\sigma, \tau)$ whose entries

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

represent the proportion of vertices with color i under σ and color j under τ. The need for this high-dimensional overlap parameter is the root of our troubles.

The upshot is that $\rho(\sigma, \tau)$ contains all the information necessary to determine the probability that both σ, τ are k-colorings. In fact, let $Z_{\rho,\mathrm{bal}}$ be the number of pairs of balanced k-colorings with overlap ρ. Then

$$\frac{1}{n} \ln \mathrm{E}\left[Z_{\rho,\mathrm{bal}}\right] \sim f(\rho) = \ln k - \frac{1}{k}\left[\sum_{i,j=1}^{k} \rho_{ij} \ln \rho_{ij}\right] + \frac{d}{2} \ln\left[1 - \frac{2}{k} + \frac{1}{k^2} \|\rho\|_2^2\right], \tag{8}$$

with $\|\rho\|_2 = \left[\sum_{a,b=1}^{k} \rho_{ab}^2\right]^{1/2}$. (We use the convention that $0 \ln 0 = 0$.)[5]

Let $\mathcal{R}$ denote the set of all possible overlap matrices. Then $\mathrm{E}[Z_{k,\mathrm{bal}}^2] = \sum_{\rho \in \mathcal{R}} \mathrm{E}\left[Z_{\rho,\mathrm{bal}}\right]$. Furthermore, because we confined ourselves to balanced k-colorings, all the overlap matrices $\rho \in \mathcal{R}$ are doubly-stochastic, i.e., all rows and columns sum to one. In fact, as n grows $\mathcal{R}$ is dense in the set $\mathcal{D}$ of all doubly stochastic $k \times k$ matrices, the ***Birkhoff polytope***. Hence, we can express the second moment as an optimization problem over $\mathcal{D}$, namely

$$\tfrac{1}{n} \ln \mathrm{E}[Z_{k,\mathrm{bal}}^2] \sim \max_{\rho \in \mathcal{D}} f(\rho). \tag{9}$$

(Upon taking logarithms the sum $\sum_{\rho \in \mathcal{R}} \mathrm{E}\left[Z_{\rho,\mathrm{bal}}\right]$ turns into a $\max$ because the total number $|\mathcal{R}|$ of summands is easily bounded by n^{k^2}, a polynomial in n.)

Let $\bar\rho = \frac{1}{k}\mathbf{1}$ be the matrix with all entries equal to $\frac{1}{k}$, the barycenter of the Birkhoff polytope. A glimpse at (7) reveals that $f(\bar\rho) \sim \frac{2}{n} \ln \mathrm{E}\left[Z_{k,\mathrm{bal}}\right]$, which corresponds to the square of the first moment. Therefore, a *necessary* condition for the success of the second moment method is that the maximum (9) is attained at $\bar\rho$. Indeed, if $f(\rho) > f(\bar\rho)$ for some $\rho \in \mathcal{D}$, then $\mathrm{E}[Z_{k,\mathrm{good}}^2]$ exceeds $\mathrm{E}\left[Z_{k,\mathrm{good}}\right]^2$ by an *exponential* factor, because (9) is on a logarithmic scale.

[5]Equation (8) follows because by inclusion/exclusion a single random edge is bichormatic under both σ, τ with probability $1 - \frac{2}{k} + \frac{1}{k^2} \sum_{i,j=1}^{k} \rho_{ij}^2 + O(1/n)$. Moreover, the number of pairs (σ, τ) with overlap ρ is $k^{n-o(n)} \binom{n}{\rho_{11}\frac{n}{k}, \ldots, \rho_{kk}\frac{n}{k}}$ (cf. [6]).

This necessary condition turns out to be sufficient, i.e., the second moment method succeeds iff the dominant contribution to (9) comes from $\bar\rho$. Combinatorially, this means that pairs σ, τ that, judging by their overlap, look completely uncorrelated make up the lion's share of $\mathrm{E}[Z_{k,\mathrm{bal}}^2]$.

B. A first attempt: the singly-stochastic bound

Unfortunately, solving (9) proves seriously difficult. Achlioptas and Naor resort to a relaxation: letting $\mathcal{S}$ denote the set of all $k \times k$ *singly* stochastic matrices (with all row sums equal to one but no constraints on the column sums), they study $\max_{\rho \in \mathcal{S}} f(\rho)$. This optimization problem turns out to be much more amenable, In fact, while in (9) *all* matrix entries are tied together by the constraint that ρ be doubly stochastic, in $\max_{\rho \in \mathcal{S}} f(\rho)$ the constraints are confined to single rows. Thus, $\max_{\rho \in \mathcal{S}} f(\rho)$ decomposes into k separate optimization problems, each over a k-dimensional simplex.

Yet even solving this relaxation is quite non-trivial. Achlioptas and Naor perform a sophisticated "global" analysis based on chasing the zeros of the differentials of certain functions related to f, the signs of the second differentials at these points, etc. (up to the *sixth* derivative). They manage to solve the relaxed problem completely. The result is that its maximum and thus that of (9) is attained at the doubly-stochastic $\bar\rho$ for $d \le d_{k,\mathrm{AN}}$, about an additive $\ln k$ below $d_{k-\mathrm{col}}$ (cf. (1)).

But for larger densities the maximum of $f(\rho)$ over singly-stochastic ρ is attained at a matrix that fails to be doubly-stochastic. Indeed, the maximizer is very close to the matrix ρ_{half} whose first $k/2$ rows coincide with those of the identity matrix id (with ones on the diagonal and zeros elsewhere) and whose last $k/2$ rows have all entries equal to $1/k$. Of course, ρ_{half} fails to be doubly-stochastic. Hence, one might hope that $\bar\rho$ remains the maximizer of (9) for d up to $d_{k,\mathrm{cond}}$. That is, however, not the case. Indeed, consider the doubly-stochastic

$$\rho_{\mathrm{stable}} = (1 - 1/k)\mathrm{id} + k^{-2}\mathbf{1}, \tag{10}$$

where $\mathbf{1}$ denotes the matrix with all entries equal to one. A simple calculation reveals that $f(\rho_{\mathrm{stable}}) > f(\bar\rho)$, and thus that the second moment argument for $Z_{k,\mathrm{bal}}$ fails, for d strictly below $d_{k,\mathrm{cond}}$.

C. The new approach

Thus, to prove Theorem 1.1 we need to work with a different random variable. The key observation behind its definition is that the second moment (9) is driven up by certain "pathological" k-colorings σ. Their number behaves like a lottery: while the random graph typically has few such colorings, a tiny fraction of graphs have an abundance, boosting the second moment. To exclude these pathological cases, we define a notion of "good" colorings. This induces a decomposition $Z_{k,\mathrm{bal}} = Z_{k,\mathrm{good}} + Z_{k,\mathrm{bad}}$ such that

$\mathrm{E}[Z_{k,\mathrm{good}}] \sim \mathrm{E}[Z_{k,\mathrm{bal}}]$. The second moment bound (5) holds for $Z_{k,\mathrm{good}}$ so long as $d \le d_{k,\mathrm{cond}} - o_k(1)$.

The notion of "good" is inspired by statistical physics predictions on the geometry of the set of k-colorings. More precisely, according to the cavity method [21], [31], for $(1+o_k(1))k \ln k < d < d_{k,\mathrm{cond}}$ the set of all k-colorings, viewed as a subset of $[k]^n$, decomposes into tiny "clusters" that are well-separated from each other. To formalize this, we define the ***cluster*** of a balanced k-coloring σ of $G(n,m)$ as the set

$$\mathcal{C}(\sigma) = \{\tau \in [k]^n : \tau \text{ is a balanced } k\text{-coloring and } \rho_{ii}(\sigma,\tau) > 0.51 \text{ for all } i \in [k]\}. \tag{11}$$

In words, $\mathcal{C}(\sigma)$ contains all balanced k-colorings τ in which more than 51% of the vertices in each color class of σ retain their color. The definition of "good" imposes constraints on the cluster size and separation.

Computing the second moment of $Z_{k,\mathrm{good}}$ boils down to an optimization problem as well. However, in comparison to (9), this problem is over a *significantly* reduced domain $\mathcal{D}_{\mathrm{good}} \subset \mathcal{D}$, reflecting the physics predictions on the clustered geometry of k-colorings:

$$\frac{1}{n} \ln \mathrm{E}[Z_{k,\mathrm{good}}^2] \sim \max_{\rho \in \mathcal{D}_{\mathrm{good}}} f(\rho). \tag{12}$$

Thus, instead of relaxing (9) as in [6], our approach is to *add* constraints to the problem. In particular, $\rho_{\mathrm{stable}} \notin \mathcal{D}_{\mathrm{good}}$. Furthermore, to solve the maximization problem (12), we pursue a novel approach: instead of performing a global analysis as in [6], we use an argument based on *local variations*, somewhat reminiscent of a gradient method in mathematical programming. Sections IV and V fill in the details.

D. The condensation transition

Finally, why does the second moment method fail beyond $d_{k,\mathrm{cond}}$? According to the (again, non-rigorous) physics predictions, as d increases up to $d_{k,\mathrm{cond}}$, both the total number $Z_{k-\mathrm{col}}$ of k-colorings and the cluster sizes decrease. However, $Z_{k-\mathrm{col}}$ drops at a faster rate, and at $d_{k,\mathrm{cond}}+o_k(1)$ the size of the largest cluster $\mathcal{C}(\sigma)$ has the same order of magnitude as the total number of k-colorings w.h.p. In effect, a bounded number of clusters dominate the entire set of k-colorings.

This prediction explains the demise of the second moment method at $d_{k,\mathrm{cond}}$. Indeed, as we saw above, the second moment method succeeds iff two random colorings σ, τ of $G(n,m)$ "look uncorrelated" in the sense that their overlap is $\bar{\rho}$ w.h.p. Once there is condensation, this type of decorrelation does no longer occur because σ, τ belong to the same cluster (and thus are highly correlated) with a non-vanishing probability.

But can we *prove* the existence of a "phase transition" at $d_{k,\mathrm{cond}}$ in any sense? The second moment argument enables us to trace both the cluster size and the number of k-colorings for $d < d_{k-\mathrm{col}} - o_k(1)$. If one extrapolates these formulas to larger d, one finds that the formula for the cluster size exceeds the extrapolation of the *total* number of k-colorings by an exponential factor! Of course, in actuality $Z_{k-\mathrm{col}}$ cannot possibly be less than the size of a single cluster. Thus, under an appropriate scaling the limiting behavior of $Z_{k-\mathrm{col}}$ and/or the cluster size has to change at $d_{k,\mathrm{cond}}$. Indeed, in physics jargon a *phase transition* is a point d_0 where the function

$$\varphi(d) = \lim_{n\to\infty} \mathrm{E}[Z_{k-\mathrm{col}}^{1/n}] \tag{13}$$

is non-analytic.[6] We believe this to occur at $d_{k,\mathrm{cond}} + o_k(1)$. However, the limit (13) is not currently known to exist for all d. Therefore, we have to phrase the following result with a bit of care.

Proposition 2.1: There is $\varepsilon_k = o_k(1)$ such that the following is true.

1) The limit $\varphi(d)$ exists and is analytic for all $d < d_{k,\mathrm{cond}} - \varepsilon_k$. Indeed, $\varphi(d) = k(1-1/k)^{d/2}$.
2) By contrast, either $\varphi(d)$ does not exist for some $d \in (d_{k,\mathrm{cond}} - \varepsilon_k, d_{k,\mathrm{cond}} + \varepsilon_k)$ or, if it exists for all such d, the limiting function $\varphi(d)$ is non-analytic at some point in this interval.

While (13) is not known to exists for all d, Bayati, Gamarnik and Tetali [9] proved the existence of a closely related limit, the so-called "free energy". Emboldened by their result, we pose

Conjecture 2.2: For any $k \ge 3$ and any $d > 0$ the limit (13) exists.

III. Related work

Over the years the random graph coloring problem has attracted a lot of attention. Shamir and Spencer used martingale tail bounds to proved concetration results [30]. Their work was enhanced first by Łuczak [24] and then by Alon and Krivelevich [8], who proved that the chromatic number of $G(n,m)$ is concentrated on two consecutive integers if $m \ll n^{3/2}$. In a breakthrough contribution, Bollobás [10] determined the asymptotic value of the chromatic number of dense random graphs (with $m = \Omega(n^2)$). This result improved prior work by Matula [25], whose "merge-and-exposure" technique Łuczak built upon to approximate the chromatic number of sparse random graphs [23].

Due to the $o_k(1)$ error term in (4), Theorem 1.1 does not yield improved bounds on $d_{k-\mathrm{col}}$ for small values of k. For instance, the best current bound on the threshold for 3-colorability remains 4.03 [3]. This bound is constructive. It is obtained by tracing a certain linear time algorithm via the

[6]We use the term "analytic" in the sense of complex analysis (i.e., the function admits an expansion into a power series with a positive radius of absolute convergence). The physics tradition is to actually consider $\lim_{n\to\infty} \frac{1}{n}\mathrm{E}[\ln Z]$ (cf. [27]). We work with the nth root instead as $Z = Z_{k-\mathrm{col}}$ may be zero.

differential equations method. While we have not attempted to optimize the error term in Theorem 1.1, it would be interesting to see if our techniques render better results for, say, $k = 3, 4, 5$ as well.

The techniques of Achlioptas and Naor [6] have been used to prove several further important results. For instance, Achlioptas and Moore [4] identified three (and for some d just two) consecutive integers on which the chromatic number of the random d-regular is concentrated. This was reduced to two integers for all fixed of d (and one for about half of all d) by adding in the small subgraph conditioning technique [20]. We expect that our techniques can be combined with small subgraph conditioning as well to get improved results for random d-regular graphs.

Both [6] and our Theorem 1.1 deal with the case that the average degree d remains fixed as $n \to \infty$. In [14] the second moment method from [6] was combined with the concentration argument from [8] to determine three (and in some cases two) integers on which the chromatic number of $G(n, m)$ is concentrated for $m \ll n^{5/4}$. We expect that the present techniques allow for an improvement.

Recently Dyer, Frieze and Greenhill [17] generalized the second moment argument from [6] to the problem of k-coloring j-uniform random hypergraphs (with average degree d fixed as $n \to \infty$ and $k, j \geq 3$ fixed as well). As in [6], a key step in their proof is to relax an optimization problem over doubly-stochastic matrices to the singly-stochastic case. Thus, it would be interesting to see if the present techniques allow for improved results in the hypergraph case.

Dani, Moore and Olson [16] studied a "decorated" coloring problem in which each pair of (u, v) of vertices comes with a permutation $\pi_{u,v}$ of the k possible colors. These permutations are chosen independently and uniformly at random for each edge. This leads to a notion of decorated k-colorings that involves the permutations on the edges. They conjecture that the threshold for k-colorability in the "decorated" problem coincides with the common $d_{k-\mathrm{col}}$. It might be interesting to see if our approach yields better bounds for the decorated k-coloring problem, possibly matching its condensation transition.

The use of the second moment method in random constraint satisfaction problems was pioneered by Achlioptas and Moore [5] and Frieze and Wormald [19], who dealt with random k-SAT. Recently improved results on *binary* random constraint satisfaction problems have been obtained via enhanced second moment arguments [12], [13], [15]. As mentioned earlier, the crucial difference between the previous and the present work is that we deal with a problem in which each "variable" (i.e., vertex) has more than two "spins" (colors) to choose from. That said, we harness the idea, first suggested in [15], of combining the second moment method with physics predictions on the geometry of the solution space. To study these geometric properties we build upon and extend techniques from [2], [7].

IV. THE RANDOM VARIABLE

The goal in this section is to define the random variable $Z_{k,\mathrm{good}}$ on which our second moment argument is based and to compute its expectation. At the expense of the $o_k(1)$ error term in (4) we may assume throughout that $k \geq k_0$ for a big constant k_0. We may also assume that n is sufficiently large.

The definition of $Z_{k,\mathrm{good}}$ is guided by the statistical mechanics predictions on the geometry of the set of k-colorings, according to which for densities $(1 + o_k(1))k \ln k \leq d \leq d_{k,\mathrm{cond}}$ the k-colorings come in well-separated clusters; recall the formal definition (11) of the "cluster" $\mathcal{C}(\sigma)$.

To formalize the concept of "well-separated", we call a balanced k-coloring σ ***separable*** if for any other balanced k-coloring τ and any $i, j \in [k]$ such that $\rho_{ij}(\sigma, \tau) > 0.51$ we indeed have $\rho_{ij}(\sigma, \tau) \geq 1 - \kappa$, where $\kappa = \ln^9 k/k = o_k(1)$. In other words, the overlap matrix $\rho(\sigma, \tau)$ does not have entries in the interval $(0.51, 1 - \kappa)$. This definition ensures that the clusters of two separable colorings σ, τ are either disjoint or identical (to see this, apply the condition to the diagonal entries $\rho_{ii}(\sigma, \tau)$).

Furthermore, according to the physics calculations each cluster only contains a small fraction of all balanced k-colorings w.h.p. Since w.h.p. their *total* number does not exceed the expectation $\mathrm{E}[Z_{k,\mathrm{bal}}]$ by much (Markov's inequality), we definitely expect that each cluster has size at most $\mathrm{E}[Z_{k,\mathrm{bal}}]$ w.h.p. These considerations lead us to

Definition 4.1: A balanced k-coloring σ is ***good*** if it is separable and $|\mathcal{C}(\sigma)| \leq \mathrm{E}[Z_{k,\mathrm{bal}}]$.

Let $Z_{k,\mathrm{good}}$ be the number of good k-colorings. A key fact is that for $d \leq d_{k,\mathrm{cond}}$ the *expectation* of $Z_{k,\mathrm{good}}$ coincides with the expectation of $Z_{k-\mathrm{col}}$, the total number of k-colorings, up to a sub-exponential factor. Hence, we merely rule out a (for our purposes) negligible fraction of "bad" colorings.

Proposition 4.2: For $d_{k,\mathrm{AN}} \leq d \leq d_{k,\mathrm{cond}} - o_k(1)$ we have

$$\frac{1}{n} \ln \mathrm{E}[Z_{k,\mathrm{good}}] \sim \frac{1}{n} \ln \mathrm{E}[Z_k] \sim \ln k + \frac{d}{2} \ln(1 - 1/k) > 0. \tag{14}$$

The notion of "good" turns out to be sufficient to ensure the success of the second moment method. More precisely, the core of this work is to establish

Proposition 4.3: There is $C = C(k) > 0$ such that

$$\mathrm{E}[Z_{k,\mathrm{good}}^2] \leq C \cdot \mathrm{E}[Z_{k,\mathrm{good}}]^2$$

for all $d_{k,\mathrm{AN}} \leq d \leq d_{k,\mathrm{cond}} - o_k(1)$.

Propositions 4.2 and 4.3 together with Eq. (5) and (6) imply Theorem 1.1. We are going to sketch the second moment argument in Section V. But before we come to that, we deal with the first moment.

Proving Proposition 4.2. We compute the first moment by way of the "planted model". Let Λ be the set of all pairs (G,σ) such that G is a graph on $V=[n]$ with m edges and σ is a balanced k-coloring of G. Moreover, let Λ_{good} be the set of all $(G,\sigma)\in\Lambda$ such that σ is a good k-coloring of G. Letting $N=\binom{\binom{n}{2}}{m}$ equal the total number of graphs with m edges, we see that

$$\mathrm{E}\left[Z_{k,\text{bal}}\right]=|\Lambda|/N, \qquad \mathrm{E}\left[Z_{k,\text{good}}\right]=|\Lambda_{\text{good}}|/N.$$

Since we already know the expectation of $Z_{k,\text{bal}}$ (from (7)), we just need to show $|\Lambda_{\text{good}}|\sim|\Lambda|$.

The ***planted distribution*** provides a simple way to draw a pair $(\boldsymbol{G},\boldsymbol{\sigma})\in\Lambda$ uniformly at random:

P1. First, draw a balanced map $\boldsymbol{\sigma}:V\to[k]$ uniformly at random.

P2. Then, draw a graph $\boldsymbol{G}$ with m edges that are bichromatic under $\boldsymbol{\sigma}$ uniformly at random.

This experiment induces the uniform distribution on Λ because each balanced σ is a proper k-coloring for an equal number of graphs. (This is not generally true for non-balanced colorings.)

Hence, to show that $|\Lambda_{\text{good}}|\sim|\Lambda|$ it suffices to verify that $(\boldsymbol{G},\boldsymbol{\sigma})\in\Lambda_{\text{good}}$ w.h.p. Standard expansion arguments show that w.h.p. $\boldsymbol{\sigma}$ is separable in $\boldsymbol{G}$. Furthermore, with respect to the cluster size we find

Lemma 4.4: Suppose that $d_{k,\text{AN}}\leq d\leq 2k\ln k$. Then $\frac{1}{n}\ln|\mathcal{C}(\boldsymbol{\sigma})|=(1+o_k(1))\frac{\ln 2}{k}$ w.h.p.

The proof of Lemma 4.4 is fairly intricate. It draws on techniques developed in [2], [7]. Roughly speaking, we establish that w.h.p. $\boldsymbol{G}$ is dominated by a *core* $\hat{\boldsymbol{G}}$ comprising of vertices that each have at least, say, 100 neighbors in $\hat{\boldsymbol{G}}$ of each color other than their own. Due to expansion properties, no vertex in $\hat{\boldsymbol{G}}$ can be recolored without leaving the cluster $\mathcal{C}(\boldsymbol{\sigma})$. Furthermore, w.h.p. most vertices $v\notin\hat{\boldsymbol{G}}$ that have at least one neighbor in each color class other than their own are "attached" to the core. This means that switching the color of v necessitates recoloring a vertex in $\hat{\boldsymbol{G}}$, which is impossible inside $\mathcal{C}(\boldsymbol{\sigma})$.

Thus, the volume of the cluster (mostly) stems from vertices v that fail to have a neighbor of some color $i\neq\boldsymbol{\sigma}(v)$. Standard calculations show that there are about $\frac{n}{k}$ such v w.h.p., and that for most of them there is only one "free" color $i\neq\boldsymbol{\sigma}(v)$. Hence, v has two colors to choose from. These choices turn out to be more or less independent for all v. In effect, the cluster size is $2^{(1+o_k(1))n/k}$ w.h.p., which is less than $\mathrm{E}[Z_{k,\text{bal}}]$ for $d<d_{k,\text{cond}}-o_k(1)$.

V. The second moment

As outlined in Section II, the "vanilla" second moment argument is based on optimizing the function $f(\rho)$ over the entire set $\mathcal{D}$ of doubly-stochastic matrices. But the notion of "good" colorings enables us to restrict the domain over which we need to optimize significantly. More precisely, let us call $\rho\in\mathcal{D}$ ***separable*** if for any $i,j\in[k]$ such that $\rho_{ij}>0.51$ we have $\rho_{ij}\geq 1-\kappa$ (with $\kappa=\ln^9 k/k$). Furthermore, we say that ρ is ***s-stable*** if there are precisely s pairs $(i,j)\in[k]\times[k]$ such that $\rho_{ij}\geq 1-\kappa$. Clearly, any doubly-stochastic matrix is s-stable for some $0\leq s\leq k$, and in each row (and column) at most one entry is $\geq 1-\kappa$. Let

$$\mathcal{D}_{\text{good}}=\{\rho\in\mathcal{D}:\rho \text{ is separable and } s\text{-stable for some } 0\leq s\leq k-1\}.$$

In other words, $\mathcal{D}_{\text{good}}$ consists of all $\rho\in\mathcal{D}$ with at most $k-1$ entries that are at least $1-\kappa$, while all other entries are at most 0.51. In particular, $\mathcal{D}_{\text{good}}$ does not contain k-stable matrices such as ρ_{stable} from (10).

Geometrically, the set $\mathcal{D}_{\text{good}}$ is obtained from the Birkhoff polytope $\mathcal{D}$ by cutting out "cylinders" consisting of matrices with an entry in $(0.51,1-\kappa)$. In effect, $\mathcal{D}_{\text{good}}$ is a disconnected set. It decomposes into the sets $\mathcal{D}_{s,\text{good}}$ of s-stable $\rho\in\mathcal{D}_{\text{good}}$ for $0\leq s\leq k-1$.

These sets $\mathcal{D}_{s,\text{good}}$ can be interpreted nicely in terms of the faces of the Birkhoff polytope. More precisely, as all $\rho\in\mathcal{D}_{s,\text{good}}$ are s-stable there are precisely s entries ρ_{ij} such that $\rho_{ij}\geq 1-\kappa$. By permuting the rows and columns suitably, we may assume that $\rho_{ii}\geq 1-\kappa$ for $i=1,\ldots,s$. Thus, ρ is close to the $k-s$-dimensional face of the Birkhoff polytope where the first s diagonal entries are equal to one. Furthermore, since all other entries of ρ are ≤ 0.51 (because ρ is separable), ρ is in fact close to a point "deep inside" this face. In fact, we are going to show that the maximum of $f(\rho)$ over $\mathcal{D}_{s,\text{good}}$ is attained at a point very close to the barycenter of the face. The result of this analysis is

Proposition 5.1: For any $0\leq s\leq k-1$ we have $\max_{\rho\in\mathcal{D}_{s,\text{good}}}f(\rho)\leq f(\bar{\rho})$, with equality only for $s=0$.

Before we come to the proof of Proposition 5.1, let us indicate how it implies the second moment bound.

Proof of Proposition 4.3 *(assuming Proposition 5.1).* Let $\mathcal{Z}_s$ be the number of pairs (σ,τ) of good k-colorings whose overlap matrix is s-stable. Then (by Cauchy-Schwarz)

$$Z_{k,\text{good}}^2=\left[\sum_{s=0}^{k}\mathcal{Z}_s\right]^2\leq(k+1)\sum_{s=0}^{k}\mathcal{Z}_s^2. \qquad (15)$$

By construction, the overlap matrix of any two good colorings is separable. Hence, Proposition 5.1 yields

$$\frac{1}{n}\ln\sum_{s=0}^{k-1}\mathrm{E}\left[\mathcal{Z}_s^2\right] \sim \max_{\rho\in\mathcal{D}_{\text{good}}}f(\rho) = f(\bar{\rho})\sim\frac{2}{n}\ln\mathrm{E}\left[Z_{k,\text{good}}\right]. \qquad (16)$$

With a bit of calculus ("Laplace method") we rid (16) of the logarithms to find $C'=C'(k)>0$ such that

$$\sum_{s=0}^{k-1}\mathrm{E}\left[\mathcal{Z}_s^2\right]\leq C'\cdot\mathrm{E}\left[Z_{k,\text{good}}\right]^2. \qquad (17)$$

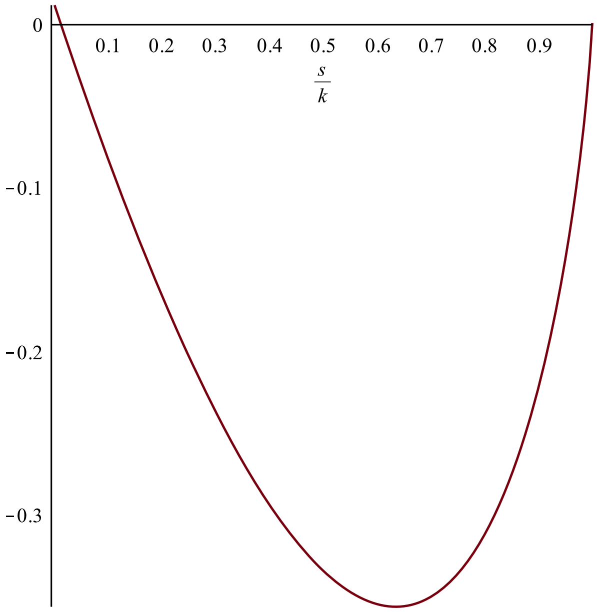

Figure 1. the function values $f(\mu_s)$ for $k = 1000$.

Finally, let σ, τ be two good colorings such that $\rho(\sigma, \tau)$ is k-stable. A suitable permutation of the color classes of τ yields a good $\hat{\tau} \in \mathcal{C}(\sigma)$. Since all good k-colorings satisfy $|\mathcal{C}(\sigma)| \leq \mathrm{E}\left[Z_{k,\mathrm{good}}\right]$, we obtain

$$\mathrm{E}\left[\mathcal{Z}_k^2\right] \leq \mathrm{E}\left[\sum_{\sigma \text{ good}} k!|\mathcal{C}(\sigma)|\right] \leq \tag{18}$$

$$\leq \mathrm{E}\left[k! \cdot \mathrm{E}\left[Z_{k,\mathrm{good}}\right] \cdot Z_{k,\mathrm{good}}\right] \ \leq k! \cdot \mathrm{E}\left[Z_{k,\mathrm{good}}\right]^2.$$

Combining (15), (17) and (18), we find that $\mathrm{E}[Z_{k,\mathrm{good}}^2] \leq C \cdot \mathrm{E}\left[Z_{k,\mathrm{good}}\right]^2$ for some $C = C(k) > 0$. $\square$

Proving Proposition 5.1. The basic idea behind the proof of Proposition 5.1 is to show that the function $f(\rho)$ is maximised on each "splinter" $\mathcal{D}_{s,\mathrm{good}}$ by a particular matrix μ_s whose value $f(\mu_s)$ can be estimated easily. Roughly speaking, the idea is to show that for each $\rho \in \mathcal{D}_{s,\mathrm{good}}$ there is a path from ρ to μ_s along which the function value increases monotonically. Although this path may leave the Birkhoff polytope temporarily, the target matrices μ_s are in $\mathcal{D}_{\mathrm{good}}$.

The matrices μ_s are "candidate local maxima" of a particularly simple form. The top-left $s \times s$ block of μ_s is of the form $(1-\alpha)\mathrm{id} + \beta\mathbf{1}$. The bottom $(k-s) \times (k-s)$ square is $\zeta\mathbf{1}$. The off-diagonal $s \times (k-s)$ and $(k-s) \times s$ blocks are of the form $\gamma \cdot \mathbf{1}$. Clearly, $\alpha, \beta, \gamma, \zeta$ must be chosen so that μ_s is doubly-stochastic, i.e., $1 - \alpha + s \cdot \beta + (k-s)\gamma = s \cdot \gamma + (k-s)\zeta = 1$, and, of course, $\alpha, \beta, \gamma, \zeta \geq 0$. Furthermore, to ensure that $\mu_s \in \mathcal{D}_{s,\mathrm{good}}$ we need that $\alpha - \beta \leq \kappa$. We let μ_s be the matrix that maximizes f subject to these constraints. The values $f(\mu_s)$ turn out to be negative for intermediate $\sqrt{k} \leq s < k$, and the overall maximum lies at $s = 0$ (see Figure 1 for an illustration). Note that $\mu_0 = \bar{\rho}$.

In fact, the parameters β, γ in the definition of μ_s tend to 0 rapidly as k gets larger. In effect, μ_s is close to the doubly-stochastic matrix $\hat{\mu}_s$ whose top-left $s \times s$ block is the identity matrix and whose bottom-right $(k-s) \times (k-s)$ block is the flat matrix $(k-s)^{-1}\mathbf{1}$. This matrix $\hat{\mu}_s$ is the barycenter of the $k-s$-dimensional face of $\mathcal{D}$ defined by the equations $\rho_{ii} = 1$ for $i = 1, \ldots, s$.

We are going to demonstrate the maximization of $f(\rho)$ over $\mathcal{D}_{s,\mathrm{good}}$ in two cases. First, for $s = 0$, where the overall maximum is attained; this turns out to be the simplest case technically.

Proposition 5.2: For any stochastic matrix ρ such that $\max_{i,j} \rho_{ij} \leq 0.51$ we have $f(\rho) \leq f(\bar{\rho})$.

In addition, we deal with one somewhat more intricate case.

Proposition 5.3: Suppose that $1 \leq s \leq k^{0.99}$. Then for any s-stable $\rho \in \mathcal{D}_{\mathrm{good}}$ we have $f(\rho) < f(\bar{\rho})$.

Proof of Proposition 5.2. We are going to argue that we can increase the function value by making the rows "flatter", eventually replacing each of them by the vector with all entries equal to $1/k$. Indeed, suppose that row i is not "flat", i.e., there exist j, l such that $\rho_{ij} < \rho_{il}$. A straight computation shows that in the extreme case $\rho_{ij} = 0$ we have $f(\rho) < f((1-\varepsilon)\rho + \varepsilon\bar{\rho})$ for a small enough $\varepsilon > 0$. In other words, the maximum of f does not occur on the boundary. Hence, we may assume that $\rho_{ij} > 0$. If we increase ρ_{ij} slightly at the expense of ρ_{il}, what will happen to the function value?

Lemma 5.4: Suppose that ρ is stochastic and that

$$0 < \rho_{ij} = \min_{q \in [k]} \rho_{iq} < \rho_{il} \leq 0.49.$$

Then

$$\frac{\partial f}{\partial \rho_{ij}} - \frac{\partial f}{\partial \rho_{il}} > 0.$$

Proof: A direct computation shows that

$$\frac{\partial f}{\partial \rho_{ij}} = -\frac{1 + \ln \rho_{ij}}{k} + \frac{d \cdot \rho_{ij}}{k^2 - 2k + \|\rho\|_2^2}.$$

Hence,

$$\frac{\partial f}{\partial \rho_{ij}} - \frac{\partial f}{\partial \rho_{il}} = \frac{1}{k} \ln\left(\frac{\rho_{il}}{\rho_{ij}}\right) + \frac{d \cdot (\rho_{ij} - \rho_{il})}{k^2 - 2k + \|\rho\|_2^2}.$$

Taking exponentials, we find that

$$\operatorname{sign}\left\{\frac{\partial f}{\partial \rho_{ij}} - \frac{\partial f}{\partial \rho_{il}}\right\} = \tag{19}$$

$$\operatorname{sign}\left\{1 + \frac{\rho_{il} - \rho_{ij}}{\rho_{ij}} - \exp\left(\frac{d \cdot (\rho_{il} - \rho_{ij})}{k - 2 + k^{-1}\|\rho\|_2^2}\right)\right\}$$

(with the convention that $\operatorname{sign}(z) = \pm 1$ if z is positive/negative, and $\operatorname{sign}(0) = 0$). Thus, we need to figure out where the linear function $z \mapsto 1 + z/\rho_{ij}$ intersects the exponential $z \mapsto \exp[z \cdot d/(k - 2 + \|\rho\|_2^2/k)]$.

We claim that for $0 < z \leq 0.49$,

$$1 + \frac{z}{\rho_{ij}} > \exp\left(\frac{d \cdot z}{k - 2 + k^{-1}\|\rho\|_2^2}\right). \tag{20}$$

Indeed, by convexity, the line and the exponential function intersect in at most one point $z^* > 0$, and for $0 < z < z^*$ the linear function is greater. Therefore, it suffices to verify that (20) holds at $z = 0.49$. On the one hand, because $d \leq 2k \ln k$, we have

$$\exp\left(\frac{0.49d}{k - 2 + \|\rho\|_2^2/k}\right) \leq \exp\left(\frac{0.98k \ln k}{k-2}\right) \leq k^{0.99},$$

provided that k is not too small. On the other hand, because ρ_{ij} is the smallest entry in row i and ρ is stochastic, we have $\rho_{ij} \leq 1/k$ and thus $1 + z_*/\rho_{ij} \geq 0.49k > k^{0.99}$. ■

Corollary 5.5: Suppose that ρ is stochastic and that

$$0 < \rho_{ij} = \min_{q \in [k]} \rho_{iq} < \rho_{il} = \max_{q \in [k]} \rho_{iq} \leq 0.49.$$

Let $\hat{\rho}$ be the matrix obtained from ρ by replacing the ith row by $\frac{1}{k}\mathbf{1}$. Then $f(\rho) < f(\hat{\rho})$.

Proof: Let Q be the set of all stochastic matrices $\tilde{\rho}$ that coincide with ρ outside row i, and that satisfy $\max_{q \in [k]} \tilde{\rho}_{iq} \leq 0.49$. Then Q is a compact set and thus f attains a maximum on Q. Assume for contradiction that the maximum is attained at ρ itself. Since $\rho_{il} < 0.49$, we clearly have $0 < z = \rho_{il} - \rho_{ij} \leq 0.49$. Hence, Lemma 5.4 and (19) show that increasing ρ_{ij} by a tiny $\varepsilon > 0$ and decreasing ρ_{il} by the same ε yields a stochastic matrix $\tilde{\rho} \in Q$ with a strictly greater function value. Since this argument applies whenever there are two distinct entries in row i, the maximum of f on Q is attained strictly at the matrix $\hat{\rho}$ where all entries in row i are equal. ■

Geometrically, the proof of Corollary 5.5 can be viewed as showing that there is a path from ρ to $\hat{\rho}$ along which the function value increases. We use a similar argument to show

Corollary 5.6: Suppose that ρ is stochastic and that

$$0.49 < \rho_{il} = \max_{q \in [k]} \rho_{iq} \leq 0.51.$$

Let $\hat{\rho}$ be the matrix obtained from ρ by replacing the ith row by $\frac{1}{k}\mathbf{1}$. Then $f(\rho) < f(\hat{\rho})$.

Proof: We may assume without loss that $i = 1$ and $\rho_{11} \geq \cdots \geq \rho_{1k} > 0$. There are two cases to consider, depending on the value of ρ_{12}. The first case is that $\rho_{12} < 0.49$. Let $\tilde{\rho}$ be the matrix obtained from ρ by replacing each of $\rho_{12}, \ldots, \rho_{1k}$ by $\frac{1-\rho_{11}}{k-1}$. Using (19) as in the proof of Corollary 5.5, we find that $f(\rho) \leq f(\tilde{\rho})$. Furthermore, direct calculations yield

$$H(\tilde{\rho}) - H(\hat{\rho}) \leq \frac{\ln 2 - 0.49 \ln k}{k},$$
$$E(\tilde{\rho}) - E(\hat{\rho}) \leq \frac{0.27 \ln k}{k}.$$

In particular, $f(\rho) \leq f(\tilde{\rho}) < f(\hat{\rho})$. An analogous argument applies in the case $\rho_{12} \geq 0.49$. ■

Corollaries 5.5 and 5.6 allow us to "flatten" the rows of a matrix ρ as in Proposition 5.2 one by one. Ultimately, this yields the desired bound $f(\rho) \leq f(\bar{\rho})$, and thus Proposition 5.2.

Proof of Proposition 5.3. Somewhat more delicate arguments are necessary to deal with $\rho \in \mathcal{D}_{s,\mathrm{good}}$ for $1 \leq s \leq k-1$. By permuting the rows and columns suitably, we may assume that $\rho_{ii} \geq 1 - \kappa$ for $i = 1, \ldots, s$ and that all other entries are less than 0.51. Think of the matrix ρ as consisting of four blocks: the upper-left $s \times s$ matrix, the off-diagonal $s \times (k-s)$ and $(k-s) \times s$ blocks, and the bottom-right $(k-s) \times (k-s)$ matrix. Roughly speaking, to estimate $f(\rho)$ we apply local variations arguments combined with estimates of their contributions to f to each of these four blocks.

We outline the proof of Proposition 5.3 to demonstrate this approach. Thus, suppose that $\rho \in \mathcal{D}_{s,\mathrm{good}}$ for some $1 \leq s \leq k^{0.99}$. Assume that $\rho_{ii} \geq 1 - \kappa$ for $1 \leq i \leq s$. We are going to compare $f(\rho)$ with $f(\hat{\mu}_s)$, where $\hat{\mu}_s$ is the doubly-stochastic matrix whose top-left $s \times s$ block is the identity matrix and whose bottom-right $(k-s) \times (k-s)$ is $\frac{1}{k-s}\mathbf{1}$. A direct calculation yields

$$f(\hat{\mu}_s) < f(\bar{\rho}) - \frac{1 + o_k(1)}{k}. \tag{21}$$

Local variations arguments akin to those in the proofs of Lemma 5.4 and Corollaries 5.5–5.6 yield the following estimate.

Lemma 5.7: Let $\hat{\rho}$ be the stochastic matrix with entries

$$\hat{\rho}_{ij} = \begin{cases} \rho_{ij} & \text{if } i \in [k], j \leq s, \\ \frac{1}{k-s}\sum_{l>s} \rho_{il} & \text{if } i \in [k], j > s. \end{cases}$$

Then $f(\rho) \leq f(\hat{\rho})$.

We are going to compare $f(\hat{\rho})$ with $f(\hat{\mu}_s)$. Because ρ is doubly stochastic, we have

$$\kappa s \geq \sum_{i=1}^{s} \sum_{j>s} \hat{\rho}_{ij} = \sum_{i=1}^{s} \sum_{j>s} \rho_{ij} = \tag{22}$$
$$= \sum_{i>s} \sum_{j=1}^{s} \rho_{ij} = \sum_{i>s} \sum_{j=1}^{s} \hat{\rho}_{ij}.$$

Let

$$q_i = \begin{cases} 1 - \rho_{ii} & \text{for } i \in [s], \\ \sum_{j=1}^{s} \hat{\rho}_{ij} & \text{for } i > s. \end{cases}$$

Because $\hat{\rho}$ is a stochastic matrix, we can view

$$\mathcal{H}(\hat{\rho}_i) = -\sum_{j=1}^{k} \hat{\rho}_{ij} \ln \hat{\rho}_{ij}$$

as the entropy of the probability distribution $(\hat{\rho}_{ij})_{j\in[k]}$ on $[k]$. Since the uniform probability distribution maximizes the entropy and as $\rho_{ii} \geq 1-\kappa$ for $i \in [s]$, we find for $i \in [s]$

$$\mathcal{H}(\hat{\rho}_i) \leq -(1-q_i)\ln(1-q_i) + q_i \ln(k/q_i) \leq \\ \leq -(1-\kappa)\ln(1-\kappa) + \kappa \ln(k/\kappa). \tag{23}$$

Similar estimates show that for $i > s$

$$\mathcal{H}(\hat{\rho}_i) \leq (1-q_i)\ln\frac{k-s}{1-q_i} \\ + q_i \ln(s/q_i) + (1-q_i)\ln(k-s) \\ \leq -(1-q_i)\ln(1-q_i) + q_i\ln(s/q_i) + \ln(k-s). \tag{24}$$

Further, because the function $z \mapsto -z\ln(z) - (1-z)\ln(1-z)$ is concave, we obtain from (22) and (24)

$$\frac{1}{k}\sum_{i>s}\mathcal{H}(\hat{\rho}_i) \leq \frac{k-s}{k}\ln(k-s) - (1-\kappa s/k)\ln(1-\kappa s/k) \\ + \frac{\kappa s}{k}\ln(k/\kappa). \tag{25}$$

Combining (23) and (25) and simplifying yields

$$\ln k - \frac{1}{k}\sum_{i,j=1}^{k}\hat{\rho}_{ij}\ln\rho_{ij} \leq \ln k + \frac{k-s}{k}\ln(k-s) + o(1/k). \tag{26}$$

Because $\hat{\rho}$ is stochastic we have $\|\hat{\rho}_i\|_2^2 \leq 1$ for $i = 1, \ldots, s$, and by construction $\hat{\rho}$ satisfies for every $i > s$

$$\sum_{j>s}\hat{\rho}_{ij}^2 = (k-s)\left(\frac{\sum_{j>s}\rho_{ij}}{k-s}\right)^2 \leq 1/(k-s).$$

Further, as $\sum_{i>s}\sum_{j=1}^{s}\hat{\rho}_{ij} \leq \kappa s$ by (22), we see that

$$\sum_{i>s}\sum_{j=1}^{s}\hat{\rho}_{ij}^2 \leq (\kappa s)^2.$$

Combining these bounds, and the fact that $s \leq k^{0.99}$, we obtain

$$\|\hat{\rho}\|_2^2 \leq s + 1 + (\kappa s)^2 \leq s + 1 + \ln^{-2} k. \tag{27}$$

By estimating the derivative of the function

$$z \mapsto \frac{d}{2}\ln(1 - 2/k + k^{-2}z^2),$$

we obtain from (27)

$$\frac{d}{2}\ln(1 - 2/k + k^{-2}\|\hat{\rho}\|_2^2) \leq \\ \leq \frac{d}{2}\ln(1 - 2/k + k^{-2}\|\hat{\mu}_s\|_2^2) + o(1/k). \tag{28}$$

Finally, combining (21), (26) and (28), we obtain $f(\rho) \leq f(\hat{\rho}) \leq f(\hat{\mu}_s) + o(1/k) < f(\bar{\rho})$, as desired.

REFERENCES

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

[2] D. Achlioptas, A. Coja-Oghlan: Algorithmic barriers from phase transitions. Proc. 49th FOCS (2008) 793–802.

[3] D. Achlioptas, C. Moore: Almost all graphs with average degree 4 are 3-colorable. Journal of Computer and System Sciences **67** (2003) 441–471.

[4] D. Achlioptas, C. Moore: The chromatic number of random regular graphs. Proc. 8th RANDOM (2004) 219–228

[5] D. Achlioptas, C. Moore: Random k-SAT: two moments suffice to cross a sharp threshold. SIAM Journal on Computing **36** (2006) 740–762.

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

[7] N. Alon, N. A. Kahale, A spectral technique for coloring random 3-colorable graphs. SIAM J. Comput. **26** (1997) 1733–1748.

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

[9] M. Bayati, D. Gamarnik, P. Tetali: Combinatorial approach to the interpolation method and scaling limits in sparse random graphs. Proc. 42nd STOC (2010) 105–114.

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

[11] A. Braunstein, R. Mulet, A. Pagnani, M. Weigt, R. Zecchina: Polynomial iterative algorithms for coloring and analyzing random graphs. Phys. Rev. E **68** (2003) 036702.

[12] A. Coja-Oghlan, K. Panagiotou: Catching the k-NAESAT threshold. Proc. 44th STOC (2012) 899–908.

[13] A. Coja-Oghlan, K. Panagiotou: Going after the k-SAT threshold. Proc. 45th STOC (2013), to appear.

[14] A. Coja-Oghlan, K. Panagiotou, A. Steger: On the chromatic number of random graphs. Journal of Combinatorial Theory, Series B, **98** (2008) 980–993.

[15] A. Coja-Oghlan, L. Zdeborová: The condensation transition in random hypergraph 2-coloring. Proc. 23rd SODA (2012) 241–250.

[16] V. Dani, C. Moore, A. Olson: Tight bounds on the threshold for permuted k-colorability. Proc. 16th RANDOM (2012) 505–516.

[17] M. Dyer, A. Frieze, C. Greenhill: On the chromatic number of a random hypergraph. Preprint (2012).

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

[19] A. Frieze, N. Wormald: Random k-Sat: a tight threshold for moderately growing k. Combinatorica **25** (2005) 297–305.

[20] G. Kemkes, X. Pérez-Giménez, N. Wormald: On the chromatic number of random d-regular graphs. Advances in Mathematics **223** (2010) 300–328.

[21] F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, L. Zdeborova: Gibbs states and the set of solutions of random constraint satisfaction problems. Proc. National Academy of Sciences **104** (2007) 10318–10323.

[22] F. Krzakala, A. Pagnani, M. Weigt: Threshold values, stability analysis and high-q asymptotics for the coloring problem on random graphs. Phys. Rev. E **70** (2004) 046705.

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

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

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

[26] M. Mézard, G. Parisi, R. Zecchina: Analytic and algorithmic solution of random satisfiability problems. Science **297** (2002) 812–815.

[27] M. Mézard, A. Montanari: Information, physics and computation. Oxford University Press 2009.

[28] R. Mulet, A. Pagnani, M. Weigt, R. Zecchina: Coloring random graphs. Phys. Rev. Lett. **89** (2002) 268701

[29] J. van Mourik, D. Saad: Random Graph Coloring - a Statistical Physics Approach. Phys. Rev. E **66** (2002) 056120.

[30] E. Shamir, J. Spencer: Sharp concentration of the chromatic number of random graphs $G(n, p)$. Combinatorica **7** (1987) 121–129

[31] L. Zdeborová, F. Krzakala: Phase transition in the coloring of random graphs. Phys. Rev. E **76** (2007) 031131.