Combinatorics, Probability and Computing (2009) 18, 881–912. © Cambridge University Press 2009
doi:10.1017/S096354830900981X First published online 24 March 2009

A Spectral Approach to Analysing Belief Propagation for 3-Colouring

AMIN COJA-OGHLAN[1*], ELCHANAN MOSSEL[2†]
and DAN VILENCHIK[3]

[1]Laboratory for Foundations of Computer Science, University of Edinburgh,
10 Crichton Street, Edinburgh EH8 9AB, UK
(e-mail: acoghlan@inf.ed.ac.uk)

[2]Department of Statistics, University of California at Berkeley, 367 Evans Hall, Berkeley, CA 94720-3860, USA
and
Faculty of Mathematics and Computer Science, The Weizmann Institute of Science, Rehovot 76100, Israel
(e-mail: mossel@stat.berkeley.edu)

[3]Computer Science Division, University of California Berkeley, CA 94720, USA
(e-mail: vilenchi@post.tau.ac.il)

Received 4 December 2007; revised 5 January 2009

Belief propagation (BP) is a message-passing algorithm that computes the exact marginal distributions at every vertex of a graphical model without cycles. While BP is designed to work correctly on trees, it is routinely applied to general graphical models that may contain cycles, in which case neither convergence, nor correctness in the case of convergence is guaranteed. Nonetheless, BP has gained popularity as it seems to remain effective in many cases of interest, even when the underlying graph is 'far' from being a tree. However, the theoretical understanding of BP (and its new relative survey propagation) when applied to CSPs is poor.

Contributing to the rigorous understanding of BP, in this paper we relate the convergence of BP to spectral properties of the graph. This encompasses a result for random graphs with a 'planted' solution; thus, we obtain the first rigorous result on BP for graph colouring in the case of a complex graphical structure (as opposed to trees). In particular, the analysis shows how belief propagation breaks the symmetry between the 3! possible permutations of the colour classes.

* Supported by grant DFG CO 646. Research done while visiting Carnegie Mellon University.
† Supported by a Sloan fellowship in Mathematics, by NSF Career award DMS-0548249, NSF grant DMS-0528488 and ONR grant N0014-07-1-05-06.

1. Introduction and results

1.1. Message-passing algorithms

This paper deals with a rigorous analysis of the *belief propagation* ('BP' for short) algorithm on certain instances of the 3-colouring problem. Originally BP was introduced by Pearl [16] as a message-passing algorithm to compute the marginals at the vertices of a probability distribution described by an acyclic 'graphical model', *i.e.*, a representation of the distribution's dependency structure as an acyclic graph. Although in the worst case BP will fail if the graphical representation features cycles, various versions of BP are in common use as heuristics in artificial intelligence and statistics, where they frequently perform well empirically as long as the underlying model does at least not contain (many) 'short' cycles. However, there is currently no general theory that could explain the empirical success of BP (with the notable exception of the use of BP in LDPC decoding [13, 14, 17]).

A striking recent application of BP is to instances of NP-hard constraint satisfaction problems such as 3-SAT or 3-colouring; this is the type of problem that we are dealing with in the present work. In this case the primary objective is *not* to compute the marginals of some distribution, but to construct a solution to the constraint satisfaction problem. For example, BP can be used to (attempt to) compute a proper 3-colouring of a given graph. Indeed, empirically BP (and its sibling *survey propagation*, 'SP') seems to perform well on problem instances that are notoriously 'hard' for other current algorithmic approaches, including the case of sparse *random graphs*.

For instance, let $G(n,p)$ be the random graph with vertex set $V = \{1,\ldots,n\}$ that is obtained by including each possible edge with probability $0 < p = p(n) < 1$ independently. Thus, the expected degree of any vertex in $G(n,p)$ is $(n-1)p \sim np$. Then there exists a threshold $\tau = \tau(n)$ such that, for any $\epsilon > 0$, the random graph $G(n,p)$ is 3-colourable with probability $1 - o(1)$ if $np < (1-\epsilon)\tau$, whereas $G(n,p)$ is not 3-colourable if $np > (1+\epsilon)\tau$ [1]. In fact, random graphs $G(n,p)$ with average degree np just below τ were considered prime examples of 'hard' instances of the 3-colouring problem, until statistical physicists discovered that BP/SP can solve these graph problems efficiently in a regime considered 'hard' for any previously known algorithms (possibly right up to the threshold density) [4, 6]. While there are exciting and deep arguments from statistical physics that provide a plausible explanation of why these message-passing algorithms succeed, these arguments are non-rigorous, and indeed no mathematically rigorous analysis is currently known.

The difficulty in understanding the performance of BP/SP on $G(n,p)$ actually lies in two aspects. The first aspect is the combinatorial structure of the random graph $G(n,p)$ with respect to the 3-colouring problem, which is not very well understood. In fact, even the basic problem of obtaining the precise value of the threshold τ is one of the current challenges in the theory of random graphs. Furthermore, we lack a rigorous understanding of the 'solution space geometry', *i.e.*, the structure of the set of all proper 3-colourings of a typical random graph $G(n,p)$ (*e.g.*, how many proper 3-colourings are there typically, and what is the typical Hamming distance between any two). But according to the statistical physics analysis, the solution space geometry affects the behaviour of BP significantly.

The second aspect, which we focus on in the present work, is the actual BP algorithm: given a graph G, how/why does the BP algorithm 'construct' a 3-colouring? Thus far there has been no rigorous analysis of BP that applies to graph colouring instances except for graphs that are *globally* tree-like (such as trees or forests). However, it seems empirically that BP performs well on many graphs that are just *locally* tree-like (*i.e.*, do not contain 'short' cycles). Therefore, in the present paper our goal is to analyse BP rigorously on a class of graphs that may have a complex combinatorial structure globally, but that have a very simple solution space geometry. More precisely, we shall relate the success of BP to spectral properties of the adjacency matrix of the input graph. In addition, we point out that the analysis comprises a natural random graph model (namely, a 'planted solution' model).

1.2. Belief propagation and spectral techniques

The main contribution of this paper is a rigorous analysis of BP for 3-colouring. We basically show that if a certain (simple) spectral heuristic for 3-colouring succeeds, then so does BP. Thus, the result does not refer to a specific random graph model, but to a special class of graphs – namely graphs that satisfy a certain spectral condition. More precisely, we say that a graph $G = (V, E)$ on n vertices is (d, ϵ)-*regular* if there exists a 3-colouring of G with colour classes V_1, V_2, V_3 such that the following is true. Let $\vec{1}_{V_i} \in \mathbb{R}^V$ be the vector whose entries equal 1 on coordinates $v \in V_i$, and 0 on all other coordinates. Then:

R1 for all $1 \leqslant i < j \leqslant 3$, the vector $\vec{1}_{V_i} - \vec{1}_{V_j}$ is an eigenvector of the adjacency matrix $A(G)$ with eigenvalue $-d$, and
R2 if $\xi \perp \vec{1}_{V_i}$ for all $i = 1, 2, 3$, then $\|A(G)\xi\| \leqslant \epsilon d \|\xi\|$.

We shall state a few elementary properties of (d, ϵ)-regular graphs in Proposition 3.3 below (assuming that ϵ is sufficiently small, say $\epsilon < 0.01$). For instance, we shall see that (d, ϵ)-regularity implies that each vertex $v \in V_i$ has precisely d neighbours in each other colour class V_j $(i \neq j)$. Moreover, (V_1, V_2, V_3) is the only 3-colouring of G (up to permutations of the colour classes, of course), and for each pair $i \neq j$ the bipartite graph consisting of the V_i–V_j-edges is an expander.

Furthermore, if a graph G is (d, ϵ)-regular for some $\epsilon < 0.01$, say, then the following spectral heuristic is easily seen to produce a 3-colouring.

(1) Compute a pair of perpendicular eigenvectors $\chi^1, \chi^2 \in \mathbb{R}^V$ of $A(G)$ with eigenvalue $-d$.
(2) Define an equivalence relation $\approx$ on V by letting $v \approx w$ if and only if $\chi^i_v = \chi^i_w$ for $i = 1, 2$. Output the equivalence classes of $\approx$ as a 3-colouring of G.

The equivalence classes of $\approx$ are precisely the three colour classes V_1, V_2, V_3, for if v, w belong to the same colour class, then their entries in all three vectors $\vec{1}_{V_i} - \vec{1}_{V_j}$ $(i < j)$ coincide; hence, as the space spanned by these vectors contains χ^1, χ^2, we have $v \approx w$. Conversely, if $v \approx w$, then the entries of v and w in all the vectors $\vec{1}_{V_i} - \vec{1}_{V_j}$ coincide, because these vectors lie in the space spanned by χ^1, χ^2; consequently, v, w belong to the same colour class V_k.

The main result of this paper is that BP can 3-colour $(d, 0.01)$-regular graphs in polynomial time, provided that d is not too small and the number of vertices is sufficiently large. We defer the description of the actual (randomized, polynomial time) BP colouring algorithm `BPCol`, to which the following theorem refers, to Section 2.

Theorem 1.1. *There exist constants* $d_0, \kappa > 0$ *such that for each* $d \geqslant d_0$ *there is a number* $n_0 = n_0(d)$ *so that the following holds. If* $G = (V, E)$ *is a* $(d, 0.01)$*-regular graph on* $n = |V| \geqslant n_0$ *vertices, then with probability* $\geqslant \kappa n^{-1}$ *over the coin tosses of the algorithm,* `BPCol`(G) *outputs a proper* 3*-colouring of* G.

Observe that Theorem 1.1 deals with 'sparse' graphs, since the lower bound n_0 on the number of vertices depends on d. The proof yields an exponential dependence, *i.e.*, $n_0 = \exp(\Theta(d))$. Conversely, this means that the average degree of G is at most logarithmic in n, which is arguably the most relevant regime to analyse BP (see Section 2). Moreover, by applying `BPCol` $O(n)$ times independently, the success probability can be boosted to $1 - \alpha$ for any $\alpha > 0$. Besides, there is an easy way to modify the (initialization step of) `BPCol` so that the success probability of one iteration is at least κ (rather than κn^{-1}): see Section 2 for details.

Let us emphasize that the contribution of Theorem 1.1 is *not* that we can now 3-colour a class of graphs for which no efficient algorithms were previously known, as the aforementioned spectral heuristic 3-colours $(d, 0.01)$-regular graphs in polynomial time. Instead, the new aspect is that we can show that the *belief propagation algorithm* 3-colours $(d, 0.01)$-regular instances, thus shedding new light on this heuristic. Indeed, the proof of Theorem 1.1, which we present in Section 3, shows that in a sense `BPCol` 'emulates' the spectral heuristic (although no spectral techniques occur in the description of `BPCol`). Thus, we establish a connection between spectral methods and BP. Besides, we note that no 'purely combinatorial' algorithm (that avoids the use of advanced techniques such as semi-definite programming or spectral methods) is known to 3-colour $(d, 0.01)$-regular graphs.

To illustrate Theorem 1.1, and to provide an example of $(d, 0.01)$-regular graphs, we point out that the main result comprises a regular random graph model with a 'planted' 3-colouring. Let $G_{n,d,3}$ be the random graph with vertex set $V = \{1, \ldots, 3n\}$ obtained as follows.

(1) Let V_1, V_2, V_3 be a random partition of V into three pairwise disjoint sets of equal size.
(2) For any pair $1 \leqslant i < j \leqslant 3$, independently choose a d-regular bipartite graph with vertex set $V_i \dot\cup V_j$ uniformly at random.

For a fixed d we say that $G_{n,d,3}$ has a certain property $\mathcal{P}$ *with high probability* ('w.h.p.'), if the probability that $G_{n,d,3}$ enjoys $\mathcal{P}$ tends to 1 as $n \to \infty$. Concerning $G_{n,d,3}$, Theorem 1.1 implies the following.

Corollary 1.2. *Suppose that* $d \geqslant d_0$ *is fixed. With high probability a random graph* $G = G_{n,d,3}$ *has the following property: with probability* $\geqslant \kappa n^{-1}$ *over the coin tosses of the algorithm,* `BPCol`(G) *outputs a proper* 3*-colouring of* G.

To prove Corollary 1.2, we show that w.h.p. $G_{n,d,3}$ is $(d, 0.01)$-regular; see Section 4.

1.3. Related work

Alon and Kahale [2] were the first to employ spectral techniques for 3-colouring sparse random graphs. They present a spectral heuristic and show that this heuristic finds a 3-colouring in the so-called 'planted solution model'. This model is somewhat more difficult to deal with algorithmically than the $G_{n,d,3}$ model that we study in the present work. For while in the $G_{n,d,3}$-model each vertex $v \in V_i$ has *exactly* d neighbours in each of the other colour classes $V_j \neq V_i$, in the planted solution model of Alon and Kahale the number of neighbours of $v \in V_i$ in V_j has a Poisson distribution with mean d. In effect, the spectral algorithm in [2] is more sophisticated than the spectral heuristic from Section 1.2. In particular, the Alon–Kahale algorithm succeeds on $(d, 0.01)$-regular graphs (and hence on $G_{n,d,3}$ w.h.p.).

There are numerous papers on the performance of message-passing algorithms for constraint satisfaction problems (*e.g.*, belief propagation/survey propagation) by authors from the statistical physics community (see [4, 5, 12] and the references therein). While these papers provide rather plausible (and insightful) explanations of the success of message-passing algorithms on problem instances such as random graphs $G_{n,p}$ or random k-SAT formulae, the arguments (*e.g.*, the replica or the cavity method) are mathematically non-rigorous.

Feige, Mossel and Vilenchik [9] showed that the *warning propagation* (WP) algorithm for 3-SAT converges in polynomial time to a satisfying assignment on a model of random 3-SAT instances with a planted solution. Since the messages in WP are additive in nature, and not multiplicative as in BP, the WP algorithm is conceptually much simpler. Moreover, on the model studied in [9], a fairly simple combinatorial algorithm (based on the 'majority vote' algorithm) is known to succeed. In contrast, no purely combinatorial algorithm (that does not rely on spectral methods or semi-definite programming) is known to 3-colour $G_{n,d,3}$ or even arbitrary $(d, 0.01)$-regular instances.

A very recent paper by Yamamoto and Watanabe [20] deals with a spectral approach to analysing BP for the Minimum Bisection problem. Their work is similar to ours in that they point out that a BP-related algorithm `pseudo-bp` emulates spectral methods. However, a significant difference is that `pseudo-bp` is a simplified version of BP that is easier to analyse, whereas in the present work we make a point of analysing the BP algorithm for colouring as it is stated in [4] (see Section 2 for more detailed comments).

The effectiveness of message-passing algorithms for amplifying local information, in order to decode codes close to channel capacity was recently established in a number of papers, *e.g.*, [13, 14, 17]. Our results are similar in flavour; however, the analysis provided here allows us to recover a proper 3-colouring of the entire graph, whereas in the random LDPC codes setting, message passing allows us to recover only a $1 - o(1)$ fraction of the codeword correctly. In [14] it is shown that for the erasure channel, all bits may be recovered correctly using a message-passing algorithm; however, in this case the message-passing algorithm is of combinatorial nature (all messages are either 0 or 1) and the LDPC code is designed so that message passing works for it.

It is important to note the difference between our work and the coding work and some interesting recent work analysing message-passing algorithms. In this work it has been shown that BP converges if the computation tree has strong correlation decay. Further, if the graph does not contain short cycles, then it converges to marginal probabilities that are close to correct. This was established in [18] and follow-up work. One way of formulating correlation decay on the computation tree is in terms of spectral properties of a recursion operator. These spectral properties are not related to those used here, as we use spectral properties of the underlying graph. Similar comments apply to the impressive work of [19] and follow-up work which applies also to graph with cycles; however, this work still requires correlation decay, and does not apply BP but a different related algorithm.

In our set-up there is no correlation decay, either for the computation tree or for the graph itself. For the graph itself this follows from the fact that there are only 6 legal colourings and that any two of them differ in at least $n/3$ of the vertices. In other words, fixing the colours of 3 of the vertices determines the colouring of the complete graph. Similarly, the computation tree is a $d-1$ regular tree, and it is easy to set up boundary conditions that fix the colouring of the inside of the tree: see [7, 11].

2. The belief propagation algorithm for 3-colouring

Following [4], in this section we will describe the basic ideas behind the BP algorithm. Since BP is a heuristic based on non-rigorous ideas (mainly from artificial intelligence and/or statistical physics), the discussion of its main ideas will somewhat lack mathematical rigour. Nonetheless, as we pointed out in the Introduction, BP makes up for this by being very successful empirically. At the end of this section, we will state precisely the version of BP that we are going to work with.

The basic strategy behind the BP algorithm for 3-colouring is to perform a fixed-point iteration for certain 'messages', starting from a suitable initial assignment. In the case of 3-colouring, the messages correspond to the edges of the graph and to the three available colours. More precisely, to each (undirected) edge $\{v, w\}$ of the graph $G = (V, E)$ and each colour $a \in \{1, 2, 3\}$, we associate two messages $\eta^a_{v\to w}$ from v to w about a, and $\eta^a_{w\to v}$ from w to v about a; in general, we will have $\eta^a_{v\to w} \neq \eta^a_{w\to v}$. Thus, the messages are *directed* objects. Each of these messages $\eta^a_{v\to w}$ is a number between 0 and 1, which we interpret as the 'probability' that vertex v takes the colour a in the graph obtained from G by removing w. Here 'probability' refers to the choice of a random (proper) 3-colouring of $G - w$, while the graph G is considered fixed. (There is an obvious symmetry issue with this definition, which we will discuss shortly.)

Having introduced the variables $\eta^a_{v\to w}$, we can set up the *belief propagation equations* for colouring, which are the basis of the BP algorithm. The BP equations reflect a relationship that the probabilities $\eta^a_{v\to w}$ should (approximately) satisfy under certain assumptions on the graph G, namely that

$$\eta^a_{v\to w} = \frac{\prod_{u\in N(v)\setminus w} 1 - \eta^a_{u\to v}}{\sum_{b=1}^{3} \prod_{u\in N(v)\setminus w} 1 - \eta^b_{u\to v}} \tag{2.1}$$

for all edges $\{v, w\}$ of G and all $a \in \{1, 2, 3\}$ (see Figure 1).

The idea behind (2.1) is that v takes colour a in the graph $G - w$ if and only if none of its neighbours $u \in N(v) \setminus w$ has colour a in $G - v$. Furthermore, the probability of

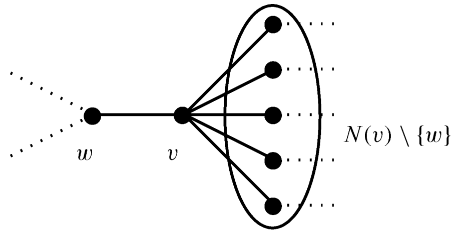

Figure 1. The BP equation.

this event ('no u has colour a') is assumed to be (asymptotically) equal to the *product* $\prod_{u\in N(v)\setminus w} 1-\eta^a_{u\to v}$ of the individual probabilities; that is, the neighbours $u \neq w$ of v are assumed to be *asymptotically independent*. Of course, this assumption does not hold for arbitrary graphs G. Finally, the numerator on the right-hand side of (2.1) is just a normalizing term, which ensures that $\sum_{a=1}^{3} \eta^a_{v\to w} = 1$.

The reason why, in the above discussion, we refer to the probability that v takes colour a in the graph $G-w$ *obtained by removing* w, rather than just to the probability that v takes colour a in G, is that in the latter case the neighbours $u \in N(v)$ would *never* be (asymptotically) independent – not even if G is a tree. For in this case the presence of v – more precisely, the existence of the short path (u, v, u') for any two neighbours $u, u' \in N(v)$ of v – would render the colours within the neighbourhood $N(v)$ heavily dependent. Similarly, if G contains triangles, so that for some vertices v the neighbourhood $N(v)$ is not an independent set, then the independence assumption that is implicit in (2.1) will be violated. Nonetheless, if G does not feature (many) short cycles – say, all the cycles are of length $\Omega(\log|V|)$ as $|V| \to \infty$ – then the BP equations (2.1) may at least be asymptotically valid. The random graph model $G_{n,d,3}$ provides an example of graphs (essentially) without such short cycles.

Now, the basic idea behind the BP algorithm is the following. We start with a 'reasonable' initial assignment $\eta^a_{v\to w}(0)$ and use (2.1) to perform a fixed-point iteration by letting

$$\eta^a_{v\to w}(l+1) = \frac{\prod_{u\in N(v)\setminus\{w\}} 1-\eta^a_{u\to v}(l)}{\sum_{b=1}^{3}\prod_{u\in N(v)\setminus\{w\}} 1-\eta^b_{u\to v}(l)} \tag{2.2}$$

for all $\{v, w\} \in E$ and $a \in \{1, 2, 3\}$. As soon as some of the values $\eta^a_{v\to w}(l+1)$ are strongly 'biased' toward either 0 or 1, we try to exploit this information to obtain a colouring.

Before we state the BP algorithm precisely, we need to discuss an important issue with the BP equations (2.1). Namely, in the case of 3-colouring the set of all 3-colourings is symmetric under permuting the colour classes. Therefore, if we actually define $\eta^a_{v\to w}$ to equal the probability w.r.t. a random 3-colouring of $G-w$, then trivially $\eta^a_{v\to w} = \frac{1}{3}$ for all a, v, w. In fact, this trivial solution is actually a fixed point of (2.2). Hence, we need to 'break symmetry'. In particular, it is not a good idea to choose the initial assignment $\eta^a_{v\to w}(0) = \frac{1}{3}$ for all a, v, w. Therefore, we do not start from $\eta^a_{v\to w}(0) = \frac{1}{3}$, but we assign to each $\eta^a_{v\to w}$ the value $\frac{1}{3}$ plus a small random error δ. The hope is that this random error will cause the fixed-point iterations (2.2) to converge to a non-trivial fixed point (other than $\eta^a_{v\to w}(0) = \frac{1}{3}$ for all a, v, w), and that this fixed point yields sufficient information to

Algorithm 2.1. BPCol(G)
Input: A graph $G = (V, E)$. *Output:* An assignment of colours to the vertices of G.

1. Let $\delta = \exp(-\log^3 n)$.
   For each $v \in V$ perform the following independently:
   choose $a \in \{1,2,3\}$ uniformly at random and assign $\eta^a_{v\to w}(0) = \frac{1}{3} + \delta$ and $\eta^b_{v\to w}(0) = \frac{1}{3} - \frac{\delta}{2}$ for all $b \in \{1,2,3\} \setminus \{a\}$ and $w \in N(v)$.
2. For $l = 1, \ldots, l^* = \lceil \log^4 n \rceil$
   compute $\eta^a_{v\to w}(l+1)$ using (2.2) for all a, v, and w.
3. For each $v \in V$ and each $a \in \{1,2,3\}$ compute $\beta^a_v = |N(v)|^{-1} \sum_{u\in N(v)} 1 - \eta^a_{u\to v}(l^*)$.
   Assign to each $v \in V$ a colour $a \in \{1,2,3\}$ such that $\beta^a_v = \max_{b\in\{1,2,3\}} \beta^b_v$.

Figure 2. The algorithm BPCol.

3-colour G. For instance, if $\chi : V \to \{1,2,3\}$ is a 3-colouring of G, then

$$\eta^a_{v\to w} = \begin{cases} 1 & \text{if } \chi(v) = a \\ 0 & \text{otherwise} \end{cases} \qquad (a = 1,2,3;\ \{v,w\} \in E)$$

is a fixed point of (2.2), and clearly the 3-colouring χ can be read out of the above messages easily. The algorithm BPCol is shown in Figure 2. Observe that step 1 ensures that

$$\sum_{a=1}^{3} \eta^a_{v\to w}(0) = 1, \quad \text{for all } \{v,w\} \in E. \tag{2.3}$$

Remarks. **(1)** Theorem 1.1 states that the probability (over the random decisions in step 1) that BPCol yields a proper 3-colouring of its $(d, 0.01)$-regular input graph is $\Omega(n^{-1})$. This can be boosted to $\Omega(1)$ by means of the following slightly more careful initialization. Instead of choosing a random a for each $v \in V$ independently, we choose a random permutation σ of V and let $W_a = \{\sigma((a-1)n/3 + 1), \ldots, \sigma(an/3)\}$ $(a = 1,2,3)$. Then, for each $v \in W_a$ we set $\eta^a_{v\to w}(0) = \frac{1}{3} + \delta$ and $\eta^b_{v\to w}(0) = \frac{1}{3} - \frac{\delta}{2}$ $(b \in \{1,2,3\} \setminus \{a\},\ w \in N(v))$. The proof of Proposition 3.5 below shows that this leads to a success probability of $\Omega(1)$. Nonetheless, we chose to state BPCol with independent decisions in its initialization, because this appears more natural (and generic) to us.

(2) Although in the above discussion of the BP equation (2.2) we referred to 'local' properties (such as the absence of short cycles), such local properties will not occur explicitly in our analysis of BPCol. Indeed, relating BPCol to spectral graph properties, the analysis has a 'global' character. Nonetheless, various local conditions (*e.g.*, a relatively small number of short cycles) are implicit in the 'global' assumption that the graph G is $(d, 0.01)$-regular (see Theorem 1.1). For more background on spectral versus combinatorial graph properties see Chung and Graham [8].

(3) BPCol updates the messages $\eta^a_{v\to w}$ 'in parallel', *i.e.*, the messages carry 'time stamps' (*cf.* (2.2)). An alternative, equally common option would be 'serial' updates, *e.g.*, by choosing each time a random pair v, w of adjacent vertices along with a colour $a \in \{1, 2, 3\}$ and updating $\eta^a_{v\to w}$ via (2.1).

(4) `BPCol` exploits the result of the fixed-point iteration (2.2) in a more straightforward fashion than the version of BP described in [4]. Namely, after performing a fixed-point iteration of (2.2), the algorithm in [4] does not assign colours to *all* vertices (as step 3 of `BPCol` does), but only to a small fraction (the most decisive ones with respect to the calculated values). Then, the algorithm performs another fixed-point iteration, *etc.* The reason is that in the random graph model considered in [4] typically the number of proper 3-colourings is exponential in the number of vertices, whereas $(d, 0.01)$-regular graphs have only one 3-colouring (up to permutations of the colours).

(5) Let us discuss the key differences between `BPCol` for $k = 2$ and the algorithm `pseudo-bp` analysed in [20].

(a) In `pseudo-bp` the products in (2.1) are taken over *all* neighbours of v, including w. This apparently minor modification has a major impact on the analysis, for including w causes the messages $\eta^a_{v\to w}$ to be independent of w. Consequently, in `pseudo-bp` the messages at time l are $2|V|$-dimensional objects, whereas in the present work the dimension is $2k|E|$.

(b) `pseudo-bp` actually works with the logarithms $\ln(\eta^a_{v\to w})$ of the messages instead of the original $\eta^a_{v\to w}$. Of course, the equation (2.1) can be phrased equivalently in terms of $\ln(\eta^a_{v\to w})$ as $\ln(\eta^a_{v\to w}) = F(\ln(\eta^a_{u\to v}))_{u\in N(v)}$ for some function F. However, in `pseudo-bp` this non-linear function F is replaced by a truncated linear function $\hat{F}$.

3. Proof of Theorem 1.1

3.1. Preliminaries and notation

Throughout this section, we let $\epsilon > 0$ be a sufficiently small constant (whose value will be determined implicitly in the course of the proof). Moreover, we keep the assumptions from Theorem 1.1. Thus, we let $d > d_0$ for a sufficiently large constant d_0; in particular, we assume that $d_0 > \exp(\epsilon^{-2})$. In addition, we assume that $n > n_0$ for some sufficiently large number $n_0 = n_0(d)$, and that $G = (V, E)$ is a $(d, 0.01)$-regular graph on $n = |V|$ vertices. This is reflected by the use of asymptotic notation in the analysis, which always refers to n being sufficiently large.

Furthermore, we let (V_1, V_2, V_3) be a 3-colouring of G with respect to which the conditions R1 and R2 from the definition of $(d, 0.01)$-regularity hold. (Actually a $(d, 0.01)$-regular graph has a unique 3-colouring up to permutations of the colour classes, but we will not use this fact.) The following easy observation will be used frequently.

Lemma 3.1. *Let $i, j \in \{1, 2, 3\}$, $i \neq j$. Then in G each vertex $v \in V_i$ has precisely d neighbours in V_j. Consequently, $|N(v)| = 2d$.*

Proof. Assume w.l.o.g. that $i = 1$ and $j = 2$. By condition R1, $\xi = \vec{1}_{V_i} - \vec{1}_{V_j}$ is an eigenvector of the adjacency matrix $A(G) = (a_{vw})_{v,w\in V}$ with eigenvalue $-d$. Hence, letting $\eta = -d\xi = A(G)\xi$, we have $-d = \eta_v = -\sum_{w\in N(v)\cap V_j} a_{vw} = -|N(v) \cap V_j|$. □

Let $\mathcal{A}$ be the set of all *ordered* pairs (v, w) such that $\{v, w\} \in E$. Following [4], we will denote the elements $(v, w) \in \mathcal{A}$ as $v \to w$. Furthermore, we shall frequently work with the vector space $\mathcal{R} = \mathbb{R}^3 \otimes \mathbb{R}^{\mathcal{A}}$. Each element $\Gamma \in \mathcal{R}$ has a unique representation,

$$\Gamma = \begin{pmatrix}1\\0\\0\end{pmatrix} \otimes \Gamma^1 + \begin{pmatrix}0\\1\\0\end{pmatrix} \otimes \Gamma^2 + \begin{pmatrix}0\\0\\1\end{pmatrix} \otimes \Gamma^3,$$

with $\Gamma^i = (\Gamma^i_{v\to w})_{v\to w\in\mathcal{A}} \in \mathbb{R}^{\mathcal{A}}$ $(i = 1, 2, 3)$. Hence, we shall denote such a vector as $\Gamma = (\Gamma^i_{v\to w})_{v\to w\in\mathcal{A}, i\in\{1,2,3\}}$. Semantically, one can think of $\Gamma^i_{v\to w}$ as the 'message' that v sends to w about colour i. Note that the messages $\eta^a_{v\to w}(l)$ defined from Section 2 constitute vectors $\eta(l) = (\eta^a_{v\to w}(l))_{v\to w\in\mathcal{A}, a\in\{1,2,3\}} \in \mathcal{R}$.

We will denote the scalar product of vectors ξ, η as $\langle \xi, \eta\rangle$. Moreover, $\|\xi\| = \sqrt{\langle \xi, \xi\rangle}$ denotes the ℓ_2-norm. In addition, if $M : \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$ is linear, then we let

$$\|M\| = \max_{\xi\in\mathbb{R}^{n_1}, \|\xi\|=1} \|M\xi\|$$

signify the operator norm of M. Further, M^T denotes the transpose of M, *i.e.*, the unique linear operator $\mathbb{R}^{n_2} \to \mathbb{R}^{n_1}$ such that $\langle M\xi, \eta\rangle = \langle \xi, M^T\eta\rangle$ for all $\xi \in \mathbb{R}^{n_1}$, $\eta \in \mathbb{R}^{n_2}$.

3.2. Outline of the analysis

In order to analyse `BPCol`, we shall relate the fixed-point iteration of (2.2) to the spectral colouring algorithm from Section 1.2. More precisely, we will approximate the fixed-point iteration of the non-linear operation (2.2) by a fixed-point iteration for a linear operator. One of the key ingredients in the analysis is to show how symmetry is broken (*i.e.*, convergence to the all-$\frac{1}{3}$ fixed point is avoided). Indeed, it may not be clear *a priori* that this will happen at all, because the random bias generated in step 1 of `BPCol` is uncorrelated to the planted colouring. The analysis is based on the following crucial observation (see Corollary 3.4 below): after a logarithmic number of iterations, for all $v \in V_i, w \in V_j, i \neq j$ the messages $\eta^a_{v\to w}$ are dominated by eigenvectors of the linear operator which we use to approximate (2.2). Furthermore, these eigenvectors mirror the colouring (V_1, V_2, V_3) and are (almost) constant on every colour class V_i (with basically $0, 1, -1$ values on the different colour classes). Hence, the (random) initial bias gets amplified so that the planted 3-colouring can eventually be read out of the messages.

To carry out this analysis precisely, we set

$$\Delta^a_{v\to w}(l) = \eta^a_{v\to w}(l) - \frac{1}{3}.$$

Moreover, we let $\mathcal{B} : \mathcal{R} \to \mathcal{R}$ denote the (non-linear) operator defined by

$$(\mathcal{B}\Gamma)^a_{v\to w} = -\frac{1}{3} + \frac{\prod_{u\in N(v)\setminus w} 1 - \frac{3}{2}\Gamma^a_{u\to v}}{\sum_{b=1}^{3}\prod_{u\in N(v)\setminus w} 1 - \frac{3}{2}\Gamma^b_{u\to v}} \quad (\Gamma \in \mathcal{R}).$$

Then (2.2) can be rephrased in terms of the vectors $\Delta(l) = (\Delta^a_{v\to w}(l))_{v\to w\in\mathcal{A}, a\in\{1,2,3\}} \in \mathcal{R}$ as

$$\Delta(l + 1) = \mathcal{B}\Delta(l). \tag{3.1}$$

We shall see that we can approximate the non-linear operator $\mathcal{B}$ in (3.1) by the following *linear* operator $\mathcal{B}'$ if $\|\Delta(l)\|_\infty$ is small; the operator $\mathcal{B}'$ maps a vector $\Gamma = (\Gamma^a_{v\to w})_{a\in\{1,2,3\},v\to w\in\mathcal{A}} \in \mathcal{R}$ to the vector $\mathcal{B}'(\Gamma) = (\mathcal{B}'(\Gamma)^a_{v\to w})_{a,v\to w} \in \mathcal{R}$ with entries

$$\mathcal{B}'(\Gamma)^a_{v\to w} = -\frac{1}{2}\sum_{u\in N(v)\setminus w} \Gamma^a_{u\to v} + \frac{1}{6}\sum_{b=1}^{3}\sum_{u\in N(v)\setminus w} \Gamma^b_{u\to v}. \tag{3.2}$$

Indeed, $\mathcal{B}' : \mathcal{R} \to \mathcal{R}$ is just the total derivative of $\mathcal{B}$ at 0.

We define a sequence $\Xi(l)$ by letting $\Xi(0) = \Delta(0)$ and $\Xi(l) = \mathcal{B}'^l\Xi(0)$ for $l \geqslant 1$, thinking of $\Xi(l)$ as a 'linear approximation' to $\Delta(l)$. As a first step, we shall simplify the operator $\mathcal{B}'$ a little.

Lemma 3.2. *We have* $(\mathcal{B}'(\Xi(l)))^a_{v\to w} = -\frac{1}{2}\sum_{u\in N(v)\setminus w} \Xi^a_{u\to v}(l)$ *for all* $l \geqslant 0$, $v \to w \in \mathcal{A}$, $a \in \{1,2,3\}$.

Proof. Step 1 of `BPCol` ensures that the initial vector satisfies

$$\sum_{b=1}^{3} \Xi^b_{u\to v}(0) = \sum_{b=1}^{3} \Delta^b_{u\to v}(0) = 0, \quad \text{for all } \{u,v\} \in E \quad (cf.\ (2.3)).$$

Therefore, by induction and by the definition (3.2) of $\mathcal{B}'$ we see that $\sum_{b=1}^{3} \Xi^b_{u\to v}(l) = 0$ for all $l \geqslant 0$. Consequently, $\sum_{b=1}^{3}\sum_{u\in N(v)\setminus w} \Xi^b_{u\to v}(l) = 0$ for all $l \geqslant 0$, *i.e.*, the second summand on the right-hand side of (3.2) vanishes. □

Due to Lemma 3.2, we may just replace $\mathcal{B}'$ by the simpler linear operator $\mathcal{L} : \mathcal{R} \to \mathcal{R}$ defined by

$$(\mathcal{L}\Gamma)^a_{v\to w} = -\frac{1}{2}\sum_{u\in N(v)\setminus w} \Gamma^a_{u\to v} \quad (v \to w \in \mathcal{A}, a \in \{1,2,3\}), \tag{3.3}$$

which satisfies

$$\Xi(l) = \mathcal{L}^l\Xi(0) = \mathcal{L}^l\Delta(0). \tag{3.4}$$

We also note for future reference that

$$\sum_{a=1}^{3} \Xi^a_{v\to w}(l) = 0, \quad \text{for all } v \to w \in \mathcal{A},\ l \geqslant 0, \tag{3.5}$$

because (2.3) entails that (3.5) is true for $l = 0$, whence the definition (3.3) of $\mathcal{L}$ shows that (3.5) holds for all $l > 0$.

In order to prove Theorem 1.1, we shall first analyse the sequence $\Xi(l)$ and then bound the error $\|\Xi(l) - \Delta(l)\|_\infty$ resulting from linearization. To study the sequence $\Xi(l)$, we investigate the dominant eigenvalues of $\mathcal{L}$ and their corresponding eigenvectors. More precisely, we shall see that our assumption on the spectrum of the adjacency matrix $A(G)$ implies that the dominant eigenvectors of $\mathcal{L}$ mirror a 3-colouring of G. We defer the proof of the following proposition to Section 3.3.

Proposition 3.3. *Let $e_{ij}^a \in \mathcal{R}$ be the vector with entries*

$$(e_{ij}^a)_{v\to w}^b = \begin{cases} 1 & \text{if } b = a,\ v \in V_i, \text{ and } w \in N(v) \cap V_j, \\ 0 & \text{otherwise,} \end{cases}$$

for $v \to w \in \mathcal{A}$, $a, b, i, j \in \{1,2,3\}$, and $i \neq j$. Let $\mathcal{E}$ be the space spanned by these 18 vectors. Then $\mathcal{L}$ operates on $\mathcal{E}$ as follows.

S1 *There are precisely six linearly independent eigenvectors $\{\zeta_2^a, \zeta_3^a : a = 1,2,3\}$ with eigenvalue $\lambda = \frac{d}{4} + \frac{1}{4}\sqrt{d^2 - 8d + 4}$, which satisfy*

$$\|\zeta_2^a - e_{12}^a - e_{13}^a + e_{21}^a + e_{23}^a\|_\infty \leqslant 100d^{-1}, \quad \|\zeta_3^a - e_{12}^a - e_{13}^a + e_{31}^a + e_{32}^a\|_\infty \leqslant 100d^{-1}. \tag{3.6}$$

These eigenvectors are symmetric with respect to the colours $a = 1,2,3$, i.e., for any two distinct $a, b \in \{1,2,3\}$ and all $v \to w \in \mathcal{A}$, we have

$$(\zeta_j^a)_{v\to w}^a = (\zeta_j^b)_{v\to w}^b \quad \text{and} \quad (\zeta_j^a)_{v\to w}^b = 0. \tag{3.7}$$

In addition,

$$\|\zeta_2^1\| = \|\zeta_j^a\|, \quad \text{for all } j \in \{2,3\}, a \in \{1,2,3\}. \tag{3.8}$$

S2 *The three vectors $e^a = \sum_{i\neq j} e_{ij}^a$ with $a = 1,2,3$ are eigenvectors with eigenvalue $\frac{1}{2} - d$.*
S3 *For all $\xi \in \mathcal{E}$ such that $\xi \perp \{e^a, \zeta_j^a : a = 1,2,3, j = 2,3\}$, we have $\|\mathcal{L}\xi\| \leqslant \frac{1}{2}\|\xi\|$.*
S4 *Furthermore, $\mathcal{L}\mathcal{E} \subset \mathcal{E}$ and $\mathcal{L}^T\mathcal{E} \subset \mathcal{E}$.*

Finally, we have

S5 $\|\mathcal{L}^2\xi\| \leqslant 0.01d^2\|\xi\|$, *for all $\xi \perp \mathcal{E}$.*

The eigenvectors that we are mostly interested in are ζ_2^a, ζ_3^a ($a = 1,2,3$), as (3.6) shows that these vectors represent the colouring (V_1, V_2, V_3) completely. As a next step, we show that $\Xi(l)$ can be approximated well by a linear combination of the vectors ζ_2^a, ζ_3^a, provided that l is sufficiently large. To this end, let

$$x_i^a = \sqrt{n} \cdot \frac{\langle \Delta(0), \zeta_i^a \rangle}{\|\Delta(0)\| \cdot \|\zeta_i^a\|} \quad (i = 2,3,\ a = 1,2,3) \tag{3.9}$$

be the projection of the initial vector $\Delta(0) = \Xi(0)$ onto the eigenvector ζ_i^a; we will see below that the normalization in (3.9) ensures that x_i^a is bounded away from 0. Furthermore, recalling from (3.8) that $\|\zeta_i^a\| = \|\zeta_2^1\|$ for all i, a, we set

$$\nu = \frac{\|\Delta(0)\|}{\sqrt{n}\|\zeta_2^1\|}. \tag{3.10}$$

Corollary 3.4. *Suppose that $l \geqslant L_1 = 2\lceil \log n \rceil$, and that $\Xi(0) \perp e^a$ for $a = 1,2,3$. Then*

$$\Xi_{v\to w}^a(l) = \nu\lambda^l \sum_{a=1}^{3}\sum_{i=2}^{3}(x_i^a + o(1))\zeta_{i\ v\to w}^a, \quad \text{for all } a \in \{1,2,3\} \text{ and } \{v,w\} \in E.$$

Proof. Since by assumption the initial vector $\Xi(0)$ is perpendicular to e^a for $a = 1, 2, 3$, and because e^1, e^2, e^3 are eigenvectors of $\mathcal{L}$ by S2, we have $\Xi(l) \perp e^a$. Therefore, we can decompose $\Xi(l)$ as

$$\Xi(l) = \xi(l) + \sum_{a=1}^{3}\sum_{i=2}^{3} z_i^a(l)\zeta_i^a, \quad \text{where } \xi(l) \perp \{e^a, \zeta_i^a : i \in \{2,3\}, a \in \{1,2,3\}\}. \tag{3.11}$$

Thus, to prove the corollary we need to compute the numbers $z_i^a(l)$ and bound $\|\xi(l)\|_\infty$.

With respect to the coefficients $z_i^a(l)$, note that $z_i^a(l) = \lambda^l z_i^a(0)$, because by S1, ζ_i^a is an eigenvector with eigenvalue λ. Moreover, $z_i^a(0) = \|\zeta_i^a\|^{-2}\langle \Xi(0), \zeta_i^a\rangle$. Hence, (3.9) and (3.10) yield $z_i^a(0) = x_i^a \cdot v$. Thus,

$$z_i^a(l) = \lambda^l v \cdot x_i^a. \tag{3.12}$$

To bound the 'error term' $\|\xi(l)\|_\infty$, we note that S3–S5 entail

$$\|\mathcal{L}^2\gamma\| \leqslant 0.01 d^2\|\gamma\| \leqslant (0.3\lambda)^2\|\gamma\|, \quad \text{for all } \gamma \perp \{e^a, \zeta_i^a : i \in \{2,3\}, a \in \{1,2,3\}\}, \tag{3.13}$$

provided that $d \geqslant d_0$ for a large enough constant $d_0 > 0$. Let $k = \lfloor l/2 \rfloor$. Since $\xi(2k) = \mathcal{L}^{2k}\xi(0)$, (3.13) implies that

$$\|\xi(2k)\| = \|\mathcal{L}^{2k}\xi(0)\| \leqslant (0.3\lambda)^{2k}\|\xi(0)\| \leqslant (0.3\lambda)^{2k}\|\Xi(0)\|. \tag{3.14}$$

Moreover, as $l \leqslant 2k + 1$ and $\|\mathcal{L}\| \leqslant d - \frac{1}{2}$ by Proposition 3.3, (3.14) yields

$$\|\xi(l)\|_\infty \leqslant \|\xi(l)\| \leqslant d\|\xi(2k)\| \leqslant d(0.3\lambda)^l\|\Xi(0)\|. \tag{3.15}$$

Finally, if $l \geqslant L_1$, then $d(0.3\lambda)^l\|\Xi(0)\| = o(\lambda^l v)$. Thus, the assertion follows from (3.11), (3.12) and (3.15). □

While in the initial vector $\Delta(0) = \Xi(0)$ the messages are completely uncorrelated with the colouring (V_1, V_2, V_3), Corollary 3.4 entails that the dominant contribution to $\Xi(L_1)$ comes from the eigenvectors ζ_i^a, which represent that colouring. This implies that all vertices v in each class V_a send essentially the same messages to all other vertices $w \in V_b$ about each of the colours 1, 2, 3, and these messages are solely determined by the initial projections x_i^a of $\Delta(0)$ onto ζ_i^a. Hence, after L_1 iterations the messages are essentially coherent and strongly correlated to the planted colouring. Thus, as a next step we analyse the distribution of the projections x_i^a. To simplify the expression resulting from Corollary 3.4, let

$$y_1^a = x_2^a + x_3^a, \quad y_2^a = -x_2^a \quad \text{and} \quad y_3^a = -x_3^a. \tag{3.16}$$

Then (3.6) and Corollary 3.4 entail that for all $v \in V_i$, all $w \in N(v)$, and $l \geqslant L_1$ we have

$$\Xi_{v\to w}^a(l) = (y_i^a + o(1)) \cdot v\lambda^l.$$

Of course, the numbers y_i^a only depend on the initial vector $\Delta(0)$. Therefore, we say that $\Delta(0)$ *is feasible* if:

F1 $\Delta(0) \perp e^a$ for $a = 1, 2, 3$, and

F2 for any pair $a, b \in \{1,2,3\}$, $a \neq b$ we have

$$|y_a^a - 1| < \exp(-1/\epsilon) \quad \text{and} \quad |y_a^b + 0.5| < \exp(-1/\epsilon). \tag{3.17}$$

Proposition 3.5. *With probability $\Omega(n^{-1})$ over the random bits used in step 1 of* `BPCol`, *$\Delta(0)$ is feasible.*

The elementary (though tedious) proof of Proposition 3.5 can be found in Section 3.4. Combining Corollary 3.4 and Proposition 3.5, we conclude that with probability $\Omega(n^{-1})$ (namely, if $\Delta(0)$ is feasible) we have

$$0.49\nu\lambda^l \leqslant \|\Xi(l)\|_\infty \leqslant 1.1\nu\lambda^l \quad (l \geqslant L_1). \tag{3.18}$$

Having obtained a sufficient understanding of the sequence $\Xi(l)$, we will now show that these vectors provide a good approximation to the vectors $\Delta(l)$, which we are actually interested in. The proof of the following proposition can be found in Section 3.5.

Proposition 3.6. *Suppose that $\Delta(0)$ is feasible. Let $L_2 > 0$ be the maximum integer such that $\|\Xi(L_2)\|_\infty \leqslant \epsilon$. Then $\|\Xi(L_2) - \Delta(L_2)\|_\infty \leqslant -\log(\epsilon) \cdot \|\Xi(L_2)\|_\infty^2$.*

Combining the information on the sequence $\Xi(l)$ provided by Corollary 3.4 and Proposition 3.5 with the bound on $\|\Xi(L_2) - \Delta(L_2)\|_\infty$ from Proposition 3.6, we can show that the messages obtained in the next one or two steps of the algorithm already represent the colouring rather well. To be precise, let us call the vector $\eta(l)$ *proper* if

$$\forall a \in \{1,2,3\},\ b \in \{1,2,3\} \setminus \{a\},\ v \in V_a,\ w \in N(v) : \eta_{v\to w}^a(l) \geqslant 0.99 \wedge \eta_{v\to w}^b(l) \leqslant 0.01.$$

Proposition 3.7. *If $\Delta(0)$ is feasible, then for either $L_3 = L_2 + 1$ or $L_3 = L_2 + 2$ the vector $\eta(L_3)$ is proper.*

The proof of Proposition 3.7 is the content of Section 3.6.

Proposition 3.7 shows that the 'rounding procedure' in step 3 of `BPCol` applied to the messages $\eta(L_3)$ would yield the colouring (V_1, V_2, V_3). However, `BPCol` actually applies that rounding procedure to $\eta(l^*)$, where $l^* > L_3$. Therefore, in order to show that `BPCol` outputs a proper 3-colouring, we need to show that these messages $\eta(l^*)$ are proper, too.

Lemma 3.8. *If $\eta(l)$ is proper, then so is $\eta(l+1)$.*

Proof. Let $v \in V_a$ for some $1 \leqslant a \leqslant 3$, $w \in N(v)$, and $\{b, c\} = \{1,2,3\} \setminus \{a\}$. Since $\eta(l)$ is proper, we have

$$\prod_{u \in V_c \cap N(v) \setminus w} \frac{1 - \eta_{u\to v}^a(l)}{1 - \eta_{u\to v}^b(l)} \geqslant \prod_{u \in V_c \cap N(v) \setminus w} 1 - \eta_{u\to v}^a(l) \geqslant 0.99^{2d}, \tag{3.19}$$

$$\prod_{u \in V_b \cap N(v) \setminus w} \frac{1 - \eta_{u\to v}^a(l)}{1 - \eta_{u\to v}^b(l)} \geqslant \left(\frac{0.99}{0.01}\right)^{2d-1} = 99^{2d-1}. \tag{3.20}$$

Consequently, the definition (2.2) of the sequence $\eta(l)$ shows that

$$\frac{\eta^a_{v\to w}(l+1)}{\eta^b_{v\to w}(l+1)} = \prod_{u\in V_c\cap N(v)\setminus w} \frac{1-\eta^a_{u\to v}(l)}{1-\eta^b_{u\to v}(l)} \cdot \prod_{u\in V_b\cap N(v)\setminus w} \frac{1-\eta^a_{u\to v}(l)}{1-\eta^b_{u\to v}(l)}$$

$$\geqslant 0.01 \cdot \left(\frac{(0.99)^2}{0.01}\right)^{2d} \geqslant 0.01 \cdot 90^{2d} \geqslant 1000. \tag{3.21}$$

As the construction (2.2) of $\eta(l+1)$ ensures that $\eta^1_{v\to w}(l+1) + \eta^2_{v\to w}(l+1) + \eta^3_{v\to w}(l+1) = 1$, (3.21) entails that $\eta^a_{v\to w}(l+1) \geqslant 0.99$ and $\eta^b_{v\to w}(l+1) \leqslant 0.01$, whence $\eta(l+1)$ is proper. □

Proof of Theorem 1.1. Proposition 3.5 states that $\Delta(0)$ is feasible with probability $\Omega(n^{-1})$. Therefore, to establish Theorem 1.1, we show that `BPCol` outputs the colouring (V_1, V_2, V_3) if $\Delta(0)$ is feasible.

Thus, assume that $\Delta(0)$ is feasible and let L_2 be the maximum integer such that $\|\Xi(L_2)\|_\infty \leqslant \epsilon$. Then Corollary 3.4 implies that $L_2 = \Theta(\log^3 n)$, because $\|\Xi(0)\|_\infty = \delta = \exp(-\log^3 n)$, and the ℓ_∞-norm of $\Xi(l)$ grows by a factor of λ in each iteration. Therefore, Proposition 3.7 entails that $\eta(L_3)$ is proper for some $L_3 = \Theta(\log^3 n)$. Thus, by Lemma 3.8 the final $\eta(\ell^*)$ generated in step 2 is proper, whence step 3 of `BPCol` outputs the colouring V_1, V_2, V_3.

3.3. Proof of Proposition 3.3

The operation (3.3) of $\mathcal{L}$ is symmetric with respect to the three colours $a = 1, 2, 3$. Therefore, we shall represent $\mathcal{L}$ as a tensor product of a 3×3 matrix and an operator that represents the graph G. To this end, we define operators $\mathcal{M} : \mathbb{R}^{\mathcal{A}} \to \mathbb{R}^{\mathcal{A}}$ and $\mathcal{K} : \mathbb{R}^{\mathcal{A}} \to \mathbb{R}^{\mathcal{A}}$ by

$$(\mathcal{M}\Xi)_{v\to w} = \sum_{u\in N(v)} \Xi_{u\to v}, \quad (\mathcal{K}\Xi)_{v\to w} = \Xi_{w\to v} \quad (\Xi \in \mathbb{R}^{\mathcal{A}}). \tag{3.22}$$

Thus,

$$-\frac{1}{2}((\mathcal{M}-\mathcal{K})\Xi)_{v\to w} = -\frac{1}{2}\sum_{u\in N(v)\setminus w} \Xi_{u\to v},$$

i.e., $-\frac{1}{2}(\mathcal{M}-\mathcal{K})$ represents the operation of $\mathcal{L}$ with respect to a single colour $a \in \{1, 2, 3\}$. Therefore, we can rephrase the definition (3.3) of $\mathcal{L}$ on the space $\mathcal{R} = \mathbb{R}^3 \otimes \mathbb{R}^{\mathcal{A}}$ as

$$\mathcal{L} = -\frac{1}{2}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \otimes (\mathcal{M}-\mathcal{K}). \tag{3.23}$$

Hence, in order to understand $\mathcal{L}$, we basically need to analyse $\mathcal{M}-\mathcal{K}$.

For $i, j \in \{1, 2, 3\}$ we define vectors $e_{ij} \in \mathbb{R}^{\mathcal{A}}$ by letting

$$(e_{ij})_{v\to w} = \begin{cases} 1 & \text{if } v \in V_i,\ w \in V_j, \text{ and } w \in N(v), \\ 0 & \text{otherwise.} \end{cases}$$

The following lemma shows that it makes sense to split the analysis of $\mathcal{M}-\mathcal{K}$ into two parts: first we shall analyse how $\mathcal{M}-\mathcal{K}$ operates on the space $\mathcal{E}_0$ spanned by the vectors e_{ij} ($1 \leqslant i, j \leqslant 3$, $i \neq j$); then, we will study the operation of $\mathcal{M}-\mathcal{K}$ on $\mathcal{E}_0^{\perp}$.

Lemma 3.9. *If $\xi \in \mathcal{E}_0$, then $\mathcal{M}\xi, \mathcal{M}^T\xi, \mathcal{K}\xi, \mathcal{K}^T\xi \in \mathcal{E}_0$.*

Proof. Let $i, j, k \in \{1, 2, 3\}$ be pairwise distinct. Since $\mathcal{K}e_{ij} = e_{ji}$, we have $\mathcal{K}\mathcal{E}_0 \subset \mathcal{E}_0$. Moreover, $\mathcal{K}^T = \mathcal{K}$. Furthermore, by Lemma 3.1,

$$(\mathcal{M}e_{ij})_{v\to w} = \sum_{u\in N(v)} (e_{ij})_{u\to v} = \begin{cases} d & \text{if } v \in V_j, \\ 0 & \text{otherwise.} \end{cases} \tag{3.24}$$

Hence, $\mathcal{M}e_{ij} = d(e_{jk} + e_{ji})$, and thus $\mathcal{M}\mathcal{E}_0 \subset \mathcal{E}_0$. In addition, the transpose of $\mathcal{M}$ is given by

$$(\mathcal{M}^T\Xi)_{v\to w} = \sum_{u\in N(w)} \Xi_{w\to u}.$$

Therefore,

$$(\mathcal{M}^T e_{ij})_{v\to w} = \sum_{u\in N(w)} (e_{ij})_{w\to u} = \begin{cases} d & \text{if } v \in V_i, \\ 0 & \text{otherwise.} \end{cases}$$

Consequently, $\mathcal{M}^T e_{ij} = d(e_{ij} + e_{ik})$, whence $\mathcal{M}^T\mathcal{E}_0 \subset \mathcal{E}_0$. □

To study the operation of $\mathcal{M}-\mathcal{K}$ on $\mathcal{E}_0$, note that (3.24) implies that $(\mathcal{M}-\mathcal{K})e_{ij} = de_{jk} + (d-1)e_{ji}$, if $i, j, k \in \{1, 2, 3\}$ are pairwise distinct. Therefore, with respect to the basis e_{12}, e_{23}, e_{21}, e_{23}, e_{31}, e_{32} of $\mathcal{E}_0$, we can represent the operation of $\mathcal{M}-\mathcal{K}$ on $\mathcal{E}_0$ by the 6×6 matrix

$$M = \begin{pmatrix} 0 & 0 & d-1 & 0 & d & 0 \\ 0 & 0 & d & 0 & d-1 & 0 \\ d-1 & 0 & 0 & 0 & 0 & d \\ d & 0 & 0 & 0 & 0 & d-1 \\ 0 & d-1 & 0 & d & 0 & 0 \\ 0 & d & 0 & d-1 & 0 & 0 \end{pmatrix}.$$

Observe that M is not symmetric. However, the columns of M can be permuted to obtain a symmetric matrix. The following lemma follows from a tedious direct computation.

Lemma 3.10. *The 6×6 matrix M is diagonalizable and has the non-zero eigenvalues 1, $2d-1$,*

$$\Lambda = -\frac{d}{2} - \frac{\sqrt{d^2 - 8d + 4}}{2}, \quad \Lambda' = -\frac{d}{2} + \frac{\sqrt{d^2 - 8d + 4}}{2}. \tag{3.25}$$

The eigenspace with eigenvalue $2d-1$ *is spanned by* $\vec{1}$. *Moreover, there are two mutually perpendicular eigenvectors* ζ_2', ζ_3' *with eigenvalue* Λ, *which satisfy*

$$\|\zeta_2' - (1,1,-1,-1,0,0)^T\|_\infty \leqslant \frac{10}{d}, \quad \|\zeta_3' - (1,1,0,0,-1,-1)^T\|_\infty \leqslant \frac{10}{d}$$

and $\|\zeta_2'\| = \|\zeta_3'\|$.

Since M describes the operation of $\mathcal{M} - \mathcal{K}$ on the subspace $\mathcal{E}_0$, Lemma 3.10 implies the following.

Corollary 3.11. *Restricted to the subspace* $\mathcal{E}_0$, *the operator* $\mathcal{M} - \mathcal{K}$ *is diagonalizable with non-zero eigenvalues* 1, $2d-1$, *and* Λ, Λ' *as in* (3.25). *The vector* $e^* = \sum_{i \neq j} e_{ij}$ *spans the eigenspace of* $2d-1$. *Furthermore, there are two mutually perpendicular eigenvectors* ζ_2, ζ_3 *with eigenvalue* Λ, *which satisfy*

$$\|\zeta_2 - (e_{12} + e_{13} - e_{21} - e_{23})\|_\infty \leqslant \frac{10}{d}, \quad \|\zeta_3 - (e_{12} + e_{13} - e_{31} - e_{32})\|_\infty \leqslant \frac{10}{d}.$$

Corollary 3.11 describes the operation of $\mathcal{M} - \mathcal{K}$ on $\mathcal{E}_0$ completely. Therefore, as a next step we shall analyse how $\mathcal{M} - \mathcal{K}$ operates on $\mathcal{E}_0^\perp$. More precisely, our goal is to show that, restricted to $\mathcal{E}_0^\perp$, the norm of $\mathcal{M} - \mathcal{K}$ is significantly smaller than Λ. To this end, we observe that the operator $\mathcal{K}$ merely permutes the coordinates. Consequently,

$$\|\mathcal{K}\| \leqslant 1. \tag{3.26}$$

To bound the norm of $\mathcal{M}$ on $\mathcal{E}_0^\perp$, we consider three subspaces of $\mathcal{E}_0^\perp$. The first subspace S consists of all vectors $\xi \in \mathcal{E}_0^\perp$ such that the value $\xi_{v \to w}$ only depends on the 'start vertex' v; in symbols,

$$S = \{\xi \in \mathcal{E}_0^\perp : \forall v \to w, v \to u \in \mathcal{A} : \xi_{v \to w} = \xi_{v \to u}\}.$$

If $\xi \in S$ and $v \in V$, then we let $\xi_{v \to} = \xi_{v \to w}$ for any $w \in N(v)$, *i.e.*, $\xi_{v \to}$ is the 'outgoing value' of v.

The second subspace T consists of all $\xi \in \mathcal{E}_0^\perp$ such that $\xi_{u \to v}$ depends only on the 'target vertex' v, *i.e.*,

$$T = \{\xi \in \mathcal{E}_0^\perp : \forall u \to v, w \to v \in \mathcal{A} : \xi_{u \to v} = \xi_{w \to v}\}.$$

For $\xi \in T$ and $v \in V$ we let $\xi_{\to v} = \xi_{u \to v}$ for any $u \in N(v)$, *i.e.*, $\xi_{\to v}$ signifies the 'incoming value' of v.

Furthermore, the third subspace U consists of all ξ such that, for any vertex, the sum of the 'incoming' values equals 0:

$$U = \left\{\xi \in \mathcal{E}_0^\perp : \forall v \in V : \sum_{u \in N(v)} \xi_{u \to v} = 0\right\}.$$

Lemma 3.12.

(1) *We have* $U = \mathrm{Kern}(\mathcal{M}) \cap \mathcal{E}_0^\perp$.

(2) *Moreover, if* $\xi \in T$, *then* $(\mathcal{M}\xi)_{v\to w} = 2d\xi_{\to v}$ *for all* $v \to w \in \mathcal{A}$. *In particular,* $\mathcal{M}\xi \in S$.
(3) *Furthermore,* $T \perp U$, *and* $\mathcal{E}_0^\perp = T \oplus U$.

Proof. The first assertion follows immediately from the definition (3.22) of $\mathcal{M}$. Moreover, if $\xi \in T$, then $(\mathcal{M}\xi)_{v\to w} = \sum_{u\in N(v)} \xi_{u\to v} = |N(v)|\xi_{\to v} = 2d\xi_{\to v}$ due to Lemma 3.1, whence (2) follows. Consequently, if $\xi \in T$ and $\eta \in U$, then

$$\langle \xi, \eta \rangle = \sum_{u\to v\in\mathcal{A}} \xi_{u\to v}\eta_{u\to v} = 2d \sum_{v\in V} \xi_{\to v} \sum_{u\in N(v)} \eta_{u\to v} = 0,$$

whence $T \perp U$. Furthermore, for any $\gamma \in \mathcal{E}_0^\perp$ the vector η with entries

$$\eta_{v\to w} = \frac{1}{2d} \sum_{u\in N(w)} \xi_{u\to w}$$

lies in T, because the sum on the right-hand side is independent of v. In addition, $\xi = \gamma - \eta$ satisfies

$$\sum_{u\in N(v)} \xi_{u\to v} = \left[\sum_{u\in N(v)} \gamma_{u\to v}\right] - 2d\eta_{\to v} = 0, \quad \text{for any } v \in V,$$

so that $\xi \in U$. Hence, any $\gamma \in \mathcal{E}_0^\perp$ can be written as $\gamma = \eta + \xi$ with $\eta \in T$ and $\xi \in U$, *i.e.*, $\mathcal{E}_0^\perp = T \oplus U$. □

By now we have all the prerequisites to analyse the operation of $\mathcal{M}$ on $\mathcal{E}_0^\perp$.

Lemma 3.13. *If* $\xi \in \mathcal{E}_0^\perp$, *then* $\|\mathcal{M}^2\xi\| \leqslant 0.01d^2\|\xi\|$.

Proof. Let $\xi \in \mathcal{E}_0^\perp$. By the third part of Lemma 3.12 there is a decomposition $\xi = \xi_T + \xi_U$ such that $\xi_T \in T$ and $\xi_U \in U$. Furthermore, the first part of Lemma 3.12 entails that $\mathcal{M}\xi = \mathcal{M}\xi_T$. Therefore, we may assume without loss of generality that $\xi = \xi_T \in T$. Hence, the second part of of Lemma 3.12 implies that

$$\|\xi'\| = 2d\|\xi\| \tag{3.27}$$

and $\xi' = \mathcal{M}\xi \in S$. Consequently, letting $\xi'' = \mathcal{M}\xi' = \mathcal{M}^2\xi$, we obtain

$$\xi''_{v\to w} = \sum_{u\in N(v)} \xi'_{u\to v} = \sum_{u\in N(v)} \xi'_{u\to}. \tag{3.28}$$

Since the right-hand side of (3.28) is independent of w, we conclude $\xi'' \in S$.

In order to bound $\|\xi''\| = \|\mathcal{M}^2\xi\|$, we shall express the sum on the right-hand side of (3.28) in terms of the adjacency matrix $A(G)$. To this end, consider the two vectors

$$\eta' = (\eta'_v)_{v\in V} \in \mathbb{R}^V \quad \text{with} \quad \eta'_v = \xi'_{v\to},$$
$$\eta'' = (\eta''_v)_{v\in V} \in \mathbb{R}^V \quad \text{with} \quad \eta''_v = \xi''_{v\to},$$

for all $v \in V$. Then

$$\|\xi'\|^2 = \sum_{v \to w \in \mathcal{A}} \xi'_{v\to w}{}^2 = 2d \sum_{v \in V} \xi'_{v\to}{}^2 = 2d\|\eta'\|^2, \quad \text{and analogously} \tag{3.29}$$

$$\|\xi''\|^2 = 2d\|\eta''\|^2. \tag{3.30}$$

Furthermore, (3.28) implies that $\eta''_v = \sum_{u \in N(v)} \eta'_u$ for all $v \in V$, *i.e.*,

$$\eta'' = A(G)\eta'. \tag{3.31}$$

Combining (3.27), (3.29), (3.30), and (3.31), we obtain

$$\|\mathcal{M}^2\xi\| = \|\xi''\| = \frac{2d\|A(G)\eta'\|}{\|\eta'\|} \cdot \|\xi\|. \tag{3.32}$$

Hence, we finally need to bound $\|A(G)\eta'\|$. To this end, we employ our assumption that G is $(d, 0.01)$-regular; namely, condition R2 from the definition of $(d, 0.01)$-regularity entails that $\|A(G)\zeta\| \leqslant 0.01d\|\zeta\|$ for all $\zeta \perp \vec{1}_{V_1}, \vec{1}_{V_2}, \vec{1}_{V_3}$. Thus, we need to show that $\eta' \perp \vec{1}_{V_i}$ for $i = 1, 2, 3$. Assuming w.l.o.g. that $i = 1$, we have

$$\begin{aligned}\langle \eta', \vec{1}_{V_1}\rangle &= \sum_{v \in V_1} \xi'_{v\to} = (2d)^{-1} \sum_{v \to w \in \mathcal{A}: v \in V_1} \xi'_{v\to w} = (2d)^{-1}\langle \xi', e_{12} + e_{13}\rangle \\ &= (2d)^{-1}\langle \mathcal{M}\xi, e_{12} + e_{13}\rangle = (2d)^{-1}\langle \xi, \mathcal{M}^T(e_{12} + e_{13})\rangle.\end{aligned} \tag{3.33}$$

Further, as $\mathcal{M}^T(e_{12} + e_{13}) \in \mathcal{E}_0$ by Lemma 3.9, while $\xi \in \mathcal{E}_0^\perp$ by our assumption, (3.33) implies that $\langle \eta', \vec{1}_{V_1}\rangle = 0$. Consequently, we obtain that $\|A(G)\eta'\| \leqslant 0.01d\|\eta'\|$, whence (3.32) yields the assertion. □

Proof of Proposition 3.3. Combining Corollary 3.11 with the tensor product representation (3.23) of $\mathcal{L}$, we conclude that the six vectors

$$\zeta_j^1 = \begin{pmatrix}1\\0\\0\end{pmatrix} \otimes \zeta_j, \quad \zeta_j^2 = \begin{pmatrix}0\\1\\0\end{pmatrix} \otimes \zeta_j, \quad \zeta_j^3 = \begin{pmatrix}0\\0\\1\end{pmatrix} \otimes \zeta_j \quad (j = 2, 3) \tag{3.34}$$

are eigenvectors of $\mathcal{L}$ with eigenvalue $\lambda = -\frac{1}{2}\Lambda$. In addition, the tensor representation (3.34) of the vectors ζ_j^a immediately implies the symmetry statement (3.7), while (3.8) follows from Corollary 3.11. Moreover, once more by Corollary 3.11 the three vectors

$$e^1 = \begin{pmatrix}1\\0\\0\end{pmatrix} \otimes e^*, \quad e^2 = \begin{pmatrix}0\\1\\0\end{pmatrix} \otimes e^*, \quad e^3 = \begin{pmatrix}0\\0\\1\end{pmatrix} \otimes e^*$$

are eigenvalues with eigenvector $-\frac{1}{2}(2d - 1) = \frac{1}{2} - d$, and all other eigenvalues of $\mathcal{L}$ restricted to $\mathcal{E}$ are $\leqslant \frac{1}{2}$ in absolute value. In addition, Lemma 3.9 shows in combination with (3.23) that $\mathcal{L}\mathcal{E}, \mathcal{L}^T\mathcal{E} \subset \mathcal{E}$. Finally, Lemma 3.13 implies in combination with (3.23) that $\|\mathcal{L}^2\xi\| \leqslant 0.01d^2\|\xi\|$ for all $\xi \perp \mathcal{E}$.

3.4. Proof of Proposition 3.5

Before we get to the proof, let us briefly discuss why the assertion (*i.e.*, Proposition 3.5) is plausible. In fact, let us point out that the vector $\Delta(0)$ is easily seen to satisfy F2 with probability $\Omega(1)$. For each of the inner products $\langle \Delta(0), \zeta_i^a \rangle$ is a sum of n independent random variables, whence the central limit theorem implies that

$$\sqrt{n}\|\Delta(0)\|^{-1}\|\zeta_i^a\|^{-1}\langle \Delta(0), \zeta_i^a \rangle$$

is asymptotically normal (the factor $\sqrt{n}\|\Delta(0)\|^{-1}\|\zeta_i^a\|^{-1}$, which is independent of the random vector $\Delta(0)$, is needed to ensure that mean and variance are of order $\Theta(1)$). In fact, since the vectors $(\zeta_i^a)_{a=1,2,3;\,i=2,3}$ are mutually perpendicular, the *joint* distribution of the random variables

$$\left(\sqrt{n}\|\Delta(0)\|^{-1}\|\zeta_i^a\|^{-1}\langle \Delta(0), \zeta_i^a \rangle\right)_{i=2,3;a=1,2,3}$$

is asymptotically a (multivariate) Gaussian. Therefore, the probability that $\Delta(0)$ satisfies F2 is $\Omega(1)$.

However, once we condition on $\Delta(0)$ satisfying F1, the entries of $\Delta(0)$ are no longer independent, whence the above argument does not yield a bound on the probability that $\Delta(0)$ satisfies both F1 and F2. Nonetheless, the dependence of the entries of $\Delta(0)$ is weak enough to allow for an elementary direct analysis. We begin with bounding the probability that $\Delta(0)$ satisfies F1. To this end, we define a partition (W_1, W_2, W_3) of V by letting

$$W_i = \{v \in V : \Delta^i_{v\to w} = \delta, \quad \text{for all } w \in N(v)\}.$$

In other words, W_i consists of all vertices for which the random number a chosen in step 1 of BPCol was equal to i.

Lemma 3.14. *The probability that $\Delta(0)$ satisfies F1 is $\Omega(n^{-1})$.*

Proof. A sufficient condition for $\Delta(0)$ to satisfy F1 is that $|W_1| = |W_2| = |W_3| = \frac{n}{3}$. Moreover, the total number of vectors that can be generated by step 1 of BPCol equals 3^n, out of which $\binom{n}{n/3\; n/3\; n/3}$ yield $W_1 = W_2 = W_3 = \frac{n}{3}$. Therefore, the assertion follows from Stirling's formula. □

In the remainder of this section we condition on the event that $\Delta(0)$ is such that $|W_1| = |W_2| = |W_3|$. Thus, (W_1, W_2, W_3) is just a random partition of V into three classes of equal size, and for all $v \in W_i$, all $j \in \{1, 2, 3\} \setminus \{i\}$, and all $w \in N(v)$ we have

$$\Delta^i_{v\to w} = \delta, \quad \Delta^j_{v\to w} = -\frac{\delta}{2}.$$

Lemma 3.15. *For any constant $c_1 > 0$ there exists a constant $c_2 > 0$ such that the following holds. If $(s_i^a)_{i,a=1,2,3}$ are integers of absolute value $|s_i^a| \leqslant c_1\sqrt{n}$ such that $\sum_{a=1}^3 s_j^a = \sum_{i=1}^3 s_i^b = 0$ for all $1 \leqslant b, j \leqslant 3$, then*

$$\mathrm{P}\left[\forall 1 \leqslant a, i \leqslant 3 : |V_a \cap W_i| = \frac{n}{9} + s_i^a\right] \geqslant c_2 n^{-2}.$$

Proof. The sets W_1, W_2, W_3 are randomly chosen mutually disjoint subsets of V of cardinality $n/3$ each, whereas V_1, V_2, V_3 are fixed subsets of V. Therefore, the total number of ways to choose W_1, W_2, W_3 is given by the multinomial coefficient $\binom{n}{n/3,n/3,n/3}$; by Stirling's formula,

$$\binom{n}{n/3, n/3, n/3} \leqslant 10n^{-1}3^n. \tag{3.35}$$

Moreover, the number of ways to choose W_1, W_2, W_3 such that $|V_a \cap W_i| = s_i^a$ equals

$$\prod_{a=1}^{3} \binom{n/3}{n/9 + s_1^a, n/9 + s_2^a, n/9 + s_3^a} \tag{3.36}$$

(because the ath factor on the right-hand side equals the number of ways to partition V_a into three pieces $V_a \cap W_1$, $V_a \cap W_2$, $V_a \cap W_3$ of the desired sizes). Combining (3.35) and (3.36) with Stirling's formula, we get

$$\begin{aligned} \mathrm{P}\left[\forall 1 \leqslant a, i \leqslant 3 : |V_a \cap W_i| = \frac{n}{9} + s_i^a\right] &\geqslant \frac{n(n/3)!^3}{10 \cdot 3^n \prod_{1 \leqslant i,a \leqslant 3}(n/9 + s_i^a)!} \\ &\geqslant \frac{n^{5/2+n}}{10 \cdot (9\mathrm{e})^n \prod_{1 \leqslant i,a \leqslant 3}(n/9 + s_i^a)!}. \end{aligned} \tag{3.37}$$

Furthermore, once again due to Stirling's formula,

$$\begin{aligned} (n/9 + s_i^a)! &\leqslant \exp(-n/9 - s_i^a)(n/9 + s_i^a)^{n/9+s_i^a}\sqrt{n} \\ &= \exp(-n/9 - s_i^a)(n/9)^{n/9+s_i^a}(1 + 9s_i^a/n)^{n/9+s_i^a}\sqrt{n} \\ &\leqslant \exp(-n/9 + 9s_i^{a2}/n)(n/9)^{n/9+s_i^a}. \end{aligned} \tag{3.38}$$

Since we are assuming that $s_i^a \leqslant c_1\sqrt{n}$ and $\sum_{i=1}^{3} s_i^a = 0$, (3.38) entails that

$$\prod_{1 \leqslant i,a \leqslant 3} (n/9 + s_i^a)! \leqslant (n/9\mathrm{e})^n n^{9/2} \exp\left(9\sum_{a,i} s_i^{a2}/n\right) \leqslant c_2'(n/9\mathrm{e})^n n^{9/2} \tag{3.39}$$

for a bounded number c_2' that depends only on c_1. Finally, plugging (3.39) into (3.37) and cancelling, we obtain the assertion. □

Corollary 3.16. *For any two constants $c_3, \beta > 0$ there exists a constant $c_4 > 0$ such that the following holds. If $(t_i^a)_{i,a=1,2,3}$ are numbers of absolute value $|t_i^a| \leqslant c_3$ such that $\sum_{a=1}^{3} t_j^a = \sum_{i=1}^{3} t_i^b = 0$ for all $1 \leqslant b, j \leqslant 3$, then*

$$\mathrm{P}\left[\forall 1 \leqslant a, i \leqslant 3 : \left|n^{-\frac{1}{2}}\left(|V_a \cap W_i| - \frac{n}{9}\right) - t_i^a\right| \leqslant \beta\right] \geqslant c_4.$$

Proof. Let S be the set of all tuples $(s_i^a)_{a,i=1,2,3}$ of integers such that $|n^{-\frac{1}{2}}s_j^b - t_j^b| \leqslant \beta$, and $\sum_{a=1}^{3} s_j^a = \sum_{i=1}^{3} s_i^b = 0$ for all $1 \leqslant b, j \leqslant 3$. Then $|S| \geqslant \beta^4 n^2/32$. Moreover, all $(s_i^a)_{a,i=1,2,3} \in S$ satisfy $|s_j^b| \leqslant (c_3 + 1)\sqrt{n}$ $(1 \leqslant b, j \leqslant 3)$. Therefore, Lemma 3.15 (applied with $c_1 = c_3 + 1$)

shows that

$$\mathrm{P}\left[\forall a,i:\left|n^{-\frac{1}{2}}\left(|V_a\cap W_i|-\frac{n}{9}\right)-t_i^a\right|\leqslant\beta\right]\geqslant\sum_{(s_i^a)\in S}\mathrm{P}\left[\forall a,i:|V_a\cap W_i|=\frac{n}{9}+s_i^a\right]$$
$$\geqslant c_2|S|n^{-2}\geqslant\beta^4c_2/32,$$

as desired. □

Since the vector $\Delta(0)$ just represents the partition W_1,W_2,W_3, and the vectors $\sum_{j\neq i}e_{ij}^a$ just represent the colouring V_1,V_2,V_3, Corollary 3.16 easily implies a result on the joint distribution of the inner products $\langle\Delta(0),\sum_{j\neq i}e_{ij}^a\rangle$.

Corollary 3.17. *For any two constants $c_5,\gamma>0$ there exists a constant $c_6>0$ such that the following is true. Suppose that $(z_i^a)_{1\leqslant a,i\leqslant 3}$ are numbers such that $|z_j^b|\leqslant c_5$ and $\sum_{i=1}^3 z_i^b=\sum_{a=1}^3 z_j^a=0$ for all $1\leqslant b,j\leqslant 3$. Then*

$$\mathrm{P}\left[\forall a,i:|z_i^a-\frac{\langle\Delta(0),\sum_{j\neq i}e_{ij}^a\rangle}{\|\Delta(0)\|\,\|\zeta_i^a\|}\cdot\sqrt{n}|\leqslant\gamma\right]\geqslant c_6.$$

Proof. The definition of $\eta(0)$ in step 1 of `BPCol` shows that

$$\Delta_{v\to w}^a(0)=\eta_{v\to w}^a(0)-\frac{1}{3}=\begin{cases}\delta & \text{if } v\in W_a,\\ -\delta/2 & \text{otherwise,}\end{cases}\quad\text{for all } v\to w\in\mathcal{A}.\tag{3.40}$$

Therefore,

$$\|\Delta(0)\|=\sqrt{3dn/2}\cdot\delta.\tag{3.41}$$

Moreover, by Proposition 3.3 there is a number $0.99\leqslant c_7\leqslant 1.01$ such that

$$\|\zeta_i^a\|=c_7\|e_{12}^a+e_{13}^a-e_{21}^a-e_{23}^a\|=2c_7\sqrt{dn}.\tag{3.42}$$

Furthermore, using (3.40), we can easily compute the scalar product $\langle\Delta(0),\sum_{j\neq i}e_{ij}^a\rangle$ $(1\leqslant a,i\leqslant 3)$:

$$\left\langle\Delta(0),\sum_{j\neq i}e_{ij}^a\right\rangle=\sum_{v\to w\in\mathcal{A}:v\in V_i}\Delta_{v\to w}^a(0)=|V_i\cap W_a|\cdot d\delta-|V_i\setminus W_a|\cdot\frac{d\delta}{2}$$
$$=\frac{3d\delta}{2}(|V_i\cap W_a|-n/9)\quad(\text{because }|V_i|=|W_a|=n/3).\tag{3.43}$$

Combining (3.41), (3.42), and (3.43), we conclude that for a certain constant $c_8>0$

$$\frac{\langle\Delta(0),\sum_{j\neq i}e_{ij}^a\rangle}{\|\Delta(0)\|\,\|\zeta_i^a\|}\cdot\sqrt{n}=\frac{c_8}{\sqrt{n}}\cdot(|V_i\cap W_a|-n/9).$$

Therefore, the assertion follows from Corollary 3.16 by setting $s_i^a=c_4^{-1}\sqrt{n}\cdot z_i^a$ and $\beta=\gamma/c_8$. □

Proof of Proposition 3.5. Let $\alpha = \exp(-1/\epsilon)$ and

$$\hat{x}_i^a = \begin{cases} -1 & \text{if } a = i \\ 1/2 & \text{otherwise} \end{cases} \quad (i = 2, 3;\ a = 1, 2, 3). \tag{3.44}$$

Then the definitions (3.9) and (3.16) of the variables x_i^a and y_i^a entail that

$$\begin{aligned} &\mathrm{P}\big[\forall a, i \in \{1,2,3\}, i \neq a : |y_a^a - 1| < \alpha \wedge |y_i^a - 1/2| < \alpha\big] \\ &\qquad \geqslant \mathrm{P}\big[\forall a, i \in \{1,2,3\} : |x_i^a - \hat{x}_i^a| < \alpha/2\big]. \end{aligned} \tag{3.45}$$

Therefore, we shall derive a lower bound on $\mathrm{P}\big[\forall a, i : |x_i^a - \hat{x}_i^a| < \alpha/2\big]$.

To this end, let

$$e_i^a = \sum_{j \in \{1,2,3\} \setminus \{i\}} e_{ij}^a \quad (1 \leqslant a, i \leqslant 3),$$

and let $\mathcal{V} \subset \mathbb{R}^{\mathcal{A}}$ be the space spanned by these nine vectors. In addition, let $q : \mathbb{R}^{\mathcal{A}} \to \mathcal{V}$ be the orthogonal projection onto $\mathcal{V}$. Since the construction of the initial vector $\Delta(0)$ in step 1 of `BPCol` ensures that $\Delta(0) \in \mathcal{V}$, we have

$$\frac{\|\Delta(0)\| \cdot \|\zeta_i^a\|}{\sqrt{n}} \cdot x_i^a = \langle \Delta(0), \zeta_i^a \rangle = \langle q\Delta(0), \zeta_i^a \rangle = \langle \Delta(0), q\zeta_i^a \rangle.$$

Hence, instead of the vectors ζ_i^a we may work with their projections $q\zeta_i^a$ onto $\mathcal{V}$. Thus, let $q_{ij}^a \in \mathbb{R}$ be the coefficients such that

$$q\zeta_i^a = \sum_{j=1}^{3} q_{ij}^a e_j^a \quad (i = 2, 3,\ a = 1, 2, 3).$$

Then by symmetry we have $q_{ij}^a = q_{ij}^b$ for all $1 \leqslant a, b \leqslant 3$; therefore, we will briefly write q_{ij} instead of q_{ij}^a. Furthermore, (3.6) implies the bounds

$$0.99 \leqslant q_{21} \leqslant 1.01,\ -1.01 \leqslant q_{22} \leqslant -0.99,\ -0.01 \leqslant q_{23} \leqslant 0.01, \tag{3.46}$$

$$0.99 \leqslant q_{31} \leqslant 1.01,\ -0.01 \leqslant q_{32} \leqslant -0.01,\ -1.01 \leqslant q_{33} \leqslant -0.99. \tag{3.47}$$

As a consequence, the matrix

$$Q = \begin{pmatrix} q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \\ 1 & 1 & 1 \end{pmatrix}$$

is regular, and there is a constant $c_9 > 0$ such that $\|Q^{-1}\| \leqslant c_9$.

Let

$$\begin{pmatrix} z_1^a \\ z_2^a \\ z_3^a \end{pmatrix} = Q^{-1} \begin{pmatrix} \hat{x}_2^a \\ \hat{x}_3^a \\ 0 \end{pmatrix} \quad (a = 1, 2, 3). \tag{3.48}$$

Since $\|Q^{-1}\| \leqslant c_9$ and $|x_i^a| \leqslant 1$ for all a, i, we have

$$|z_i^a| \leqslant 5c_9 \quad (1 \leqslant a, i \leqslant 3). \tag{3.49}$$

In addition, (3.44) and (3.48) imply that

$$\sum_{a=1}^{3} \begin{pmatrix} z_1^a \\ z_2^a \\ z_3^a \end{pmatrix} = Q^{-1} \left[\sum_{a=1}^{3} \begin{pmatrix} x_2^a \\ x_3^a \\ 0 \end{pmatrix} \right] = 0, \quad \text{and} \tag{3.50}$$

$$\sum_{i=1}^{3} z_i^b = 0 \quad (1 \leqslant b \leqslant 3). \tag{3.51}$$

Combining (3.49)–(3.51), we see that $(z_i^a)_{1 \leqslant a,i \leqslant 3}$ satisfies the assumptions of Corollary 3.17, whence

$$\mathrm{P}\left[\forall a, i : \left| z_i^a - \frac{\langle \Delta(0), e_i^a \rangle}{\|\Delta(0)\| \, \|\zeta_i^a\|} \cdot \sqrt{n} \right| \leqslant \alpha^2 \right] \geqslant c_6 \tag{3.52}$$

for some constant $c_2 > 0$. Furthermore, if $\Delta(0) \in \mathbb{R}^{\mathcal{A}}$ satisfies

$$\left| z_i^a - \frac{\langle \Delta(0), e_i^a \rangle}{\|\Delta(0)\| \, \|\zeta_i^a\|} \cdot \sqrt{n} \right| \leqslant \alpha^2,$$

then (3.48) and the bounds (3.46)–(3.47) imply that

$$|\hat{x}_j^a - x_j^a| = \left| \hat{x}_j^a - \frac{\langle \Delta(0), \zeta_j^a \rangle}{\|\Delta(0)\| \, \|\zeta_j^a\|} \cdot \sqrt{n} \right| = \left| \sum_{i=1}^{3} q_{ji} \left(z_i^a - \frac{\langle \Delta(0), e_i^a \rangle}{\|\Delta(0)\| \, \|\zeta_j^a\|} \cdot \sqrt{n} \right) \right|$$
$$\leqslant \alpha^2 \sum_{i=1}^{3} |q_{ji}| \leqslant 3\alpha^2 < \alpha/2 \quad (j = 2, 3,\ a = 1, 2, 3).$$

Therefore, (3.52) yields

$$\mathrm{P}\left[\forall a, i : |x_i^a - \hat{x}_i^a| < \alpha/2\right] \geqslant c_6.$$

Thus, the assertion follows from (3.45) and Lemma 3.14.

3.5. Proof of Proposition 3.6

Our goal in this section is to bound the error $\|\Delta(l) - \Xi(l)\|_\infty$ resulting from replacing the non-linear operator $\mathcal{B}$ by the linear operator $\mathcal{L}$. Since $\Delta(l) = \mathcal{B}^l \Delta(0)$ and $\Xi(l) = \mathcal{L}^l \Xi(0) = \mathcal{L}^l \Delta(0)$ by (3.4), the main difficulty of this analysis is to bound how errors that were made early on in the sequence (*i.e.*, for 'small' l) amplify in the subsequent iterations. To control this phenomenon, we proceed by induction on l. We begin with a simple lemma that bounds the error occurring in a single iteration. Recall that the constructions of $\Xi(l)$ and $\Delta(l)$ ensure that $\sum_{a=1}^{3} \Xi_{v\to w}^a(l) = \sum_{a=1}^{3} \Delta_{v\to w}^a(l) = 0$ for all $v \to w \in \mathcal{A}$ (*cf.* (2.3) and (3.5)).

Lemma 3.18. *Suppose that* Γ *satisfies* $\sum_{a=1}^{3} \Gamma_{v\to w}^a = 0$ *for all* $v \to w \in \mathcal{A}$. *If* $\|\Gamma\|_\infty < 0.001 d^{-1}$, *then* $\|\mathcal{B}\Gamma - \mathcal{L}\Gamma\|_\infty \leqslant 100 d^2 \|\Gamma\|_\infty^2$.

Proof. We employ the elementary inequalities

$$\exp(-x - x^2) \leqslant 1 - x \leqslant \exp(-x) \leqslant 1 - x + x^2 \quad (|x| \leqslant 0.1). \tag{3.53}$$

Let $v \to w \in \mathcal{A}$, $a \in \{1,2,3\}$, and set

$$\Pi_b = \prod_{u \in N(v) \setminus w} 1 - \frac{3}{2}\Gamma^b_{u \to v} \quad (b \in \{1,2,3\}).$$

Moreover, let $\hat{\Gamma} = \mathcal{B}\Gamma$. Then we can rephrase the definition (3.1) of $\mathcal{B}$ as

$$\hat{\Gamma}^a_{v \to w} = -\frac{1}{3} + \frac{\Pi_a}{\sum_{b=1}^3 \Pi_b}. \tag{3.54}$$

In order to prove the lemma, we bound the error term $|\Pi_b - (1 - \frac{3}{2}\sum_{u \in N(v)\setminus w} \Gamma^b_{u \to v})|$. To this end, note that by (3.53) there exist numbers $0 \leqslant \alpha^b_{u \to v} \leqslant 9/4$ such that $1 - \frac{3}{2}\Gamma^b_{u \to v} = \exp(-\frac{3}{2}\Gamma^b_{u \to v} - \alpha^b_{u \to v}\Gamma^{b\,2}_{u \to v})$. Hence, once more by (3.53) there is a number $-1 \leqslant \beta^b \leqslant 1$ such that

$$\begin{aligned}
\Pi_b &= \exp\left[-\sum_{u \in N(v)\setminus w} \frac{3}{2}\Gamma^b_{u \to v} + \alpha^b_{u \to v}\Gamma^{b\,2}_{u \to v}\right] \\
&= 1 - \sum_{u \in N(v)} \left[\frac{3}{2}\Gamma^b_{u \to v} + \alpha^b_{u \to v}\Gamma^{b\,2}_{u \to v}\right] + \beta^b \left[\sum_{u \in N(v)} \frac{3}{2}\Gamma^b_{u \to v} + \alpha^b_{u \to v}\Gamma^{b\,2}_{u \to v}\right]^2 \\
&= L_b + E_b, \qquad \text{where we let} \qquad (3.55)\\
L_b &= 1 - \sum_{u \in N(v)} \frac{3}{2}\Gamma^b_{u \to v}, \quad \text{and} \\
E_b &= \sum_{u \in N(v)\setminus w} \alpha^b_{u \to v}\Gamma^{b\,2}_{u \to v} + \beta^b \left[\sum_{u \in N(v)} \frac{3}{2}\Gamma^b_{u \to v} + \alpha^b_{u \to v}\Gamma^{b\,2}_{u \to v}\right]^2.
\end{aligned}$$

Further, since $\|\Gamma\|_\infty \leqslant 0.001/d$ by assumption and $|N(v)| = 2d$ by Lemma 3.1, we obtain the bound

$$|E_b| \leqslant 10d^2\|\Gamma\|^2_\infty \leqslant 0.01. \tag{3.56}$$

As $\sum_{b=1}^3 L_b = 3$, due to our assumption that $\sum_{b=1}^3 \Gamma^b_{u \to v} = 0$ for all $u \to v \in \mathcal{A}$, plugging (3.55) into (3.54) yields

$$\hat{\Gamma}^a_{v \to w} + \frac{1}{3} = \frac{L_a + E_a}{3 + E_1 + E_2 + E_3} = \frac{L_a}{3} + \frac{3E_a + L_a(E_1 + E_2 + E_3)}{3(3 + E_1 + E_2 + E_3)}. \tag{3.57}$$

Since $|L_a| \leqslant 1 + 3d\|\Gamma\|_\infty \leqslant 2$, (3.56) and (3.57) yield that

$$\left|\frac{1}{3}(1 - L_a) - \hat{\Gamma}^a_{v \to w}\right| \leqslant 100d^2\|\Gamma\|^2_\infty. \tag{3.58}$$

Finally, a glance at (3.3) reveals that $(\mathcal{L}\Gamma)^a_{v \to w} = \frac{1}{3}(1 - L_a)$, and thus the assertion follows from (3.58). □

Lemma 3.18 allows us to bound the error $\|\Delta(l+1) - \Xi(l+1)\|_\infty$ resulting from iteration $l+1$ in terms of the error $\|\Delta(l) - \Xi(l)\|_\infty$ from the previous iteration. Hereafter, we let $C > 0$ denote a sufficiently large constant.

Lemma 3.19. *Suppose that* $\|\Delta(l)-\Xi(l)\|_\infty \leqslant (Cd)^{-1}$. *Then*

$$\|\Delta(l+1)-\Xi(l+1)\|_\infty \leqslant 2Cd^2\|\Xi(l)\|_\infty^2 + 4d\|\Delta(l)-\Xi(l)\|_\infty.$$

Proof. By Lemma 3.18 and the definition (3.3) of $\mathcal{L}$ we have

$$\begin{aligned}\|\Delta(l+1)-\Xi(l+1)\|_\infty &= \|\mathcal{B}\Delta(l)-\mathcal{L}\Xi(l)\|_\infty \\ &\leqslant \|\mathcal{B}\Delta(l)-\mathcal{L}\Delta(l)\|_\infty + \|\mathcal{L}\Delta(l)-\mathcal{L}\Xi(l)\|_\infty \\ &\leqslant Cd^2\|\Delta(l)\|_\infty^2 + 2d\|\Delta(l)-\Xi(l)\|_\infty. \end{aligned} \tag{3.59}$$

Moreover, $\|\Delta(l)\|_\infty \leqslant \|\Xi(l)\|_\infty + \|\Xi(l)-\Delta(l)\|_\infty$, whence (3.59) yields

$$\|\Delta(l+1)-\Xi(l+1)\|_\infty \leqslant 2Cd^2\left[\|\Xi(l)\|_\infty^2 + \|\Xi(l)-\Delta(l)\|_\infty^2\right] + 2d\|\Delta(l)-\Xi(l)\|_\infty.$$

This implies the assertion, because we are assuming that $\|\Delta(l)-\Xi(l)\|_\infty \leqslant (Cd)^{-1}$. □

Further, applying Lemma 3.19 L times recursively, we obtain the following bound.

Corollary 3.20. *Suppose that* $\|\Delta(l)-\Xi(l)\|_\infty \leqslant (Cd)^{-1}$ *for all* $l < L$. *Then*

$$\|\Delta(L)-\Xi(L)\|_\infty \leqslant 2Cd^2\sum_{j=1}^{L-1}(4d)^{j-1}\|\Xi(L-j)\|_\infty^2 + Cd^2(4d)^{L-1}\|\Delta(0)\|_\infty^2.$$

To proceed, we need the following (rough) absolute bound on the error $\|\Delta(L)-\Xi(L)\|_\infty$.

Lemma 3.21. *If* $L \leqslant \log^2 n$, *then* $\|\Delta(L)-\Xi(L)\|_\infty < (Cd)^{-1}$.

Proof. The proof is by induction on L. For $L=0$ the assertion is trivially true. Thus, assume that $\|\Delta(l)-\Xi(l)\|_\infty < (Cd)^{-1}$ for all $l < L \leqslant \log^2 n$. Then Corollary 3.20 entails that

$$\|\Delta(L)-\Xi(L)\|_\infty \leqslant 2Cd^2\sum_{j=1}^{L-1}(4d)^{j-1}\|\Xi(L-j)\|_\infty^2 + Cd^2(4d)^{L-1}\|\Delta(0)\|_\infty^2.$$

Further, the definition (3.3) of $\mathcal{L}$ shows that

$$\|\Xi(l)\|_\infty \leqslant (2d)^l\|\Delta(0)\|_\infty = (2d)^l\delta.$$

Hence,

$$\|\Delta(L)-\Xi(L)\|_\infty \leqslant 4Cd^2(2d)^{2L-2}\delta^2 + Cd^2(4d)^{L-1}\delta^2.$$

As $\delta \leqslant \exp(-\log^3 n)$ and $d = O(1)$, the right-hand side is $o(1)$ as $n\to\infty$, and thus $\|\Delta(L)-\Xi(L)\|_\infty < (Cd)^{-1}$, provided that n is sufficiently large. □

Lemma 3.22. *Let* L^* *be the maximum integer such that* $\|\Xi(L^*)\|_\infty < \epsilon$. *Then, for all* $\log^2 n \leqslant L \leqslant L^*$ *we have* $\|\Xi(L)-\Delta(L)\|_\infty \leqslant -\log(\epsilon)\cdot\|\Xi(L)\|_\infty^2$.

Proof. By the definition (3.3) of $\mathcal{L}$ there are constants $c_1, c_2 > 0$ such that

$$\|\Xi(l)\|_\infty \leqslant (2d)^l \delta \quad (\forall\, l \leqslant c_2 \log n), \tag{3.60}$$

$$\|\Xi(l)\|_\infty \in \left[c_1^{-1} \lambda^l \delta / \sqrt{dn}, c_1 \lambda^l \delta / \sqrt{dn}\right] \quad (\forall\, l \geqslant c_2 \log n). \tag{3.61}$$

We proceed inductively for $\log^2 n \leqslant L \leqslant L^*$. Thus, assume that $\|\Xi(l) - \Delta(l)\|_\infty \leqslant c_1 \|\Xi(l)\|_\infty^2$ for all $\log^2 n \leqslant l < L$. Since $\lambda \geqslant 0.1d$ and $\|\Xi(L)\|_\infty < \epsilon$, this implies that

$$\|\Xi(l) - \Delta(l)\|_\infty \leqslant (Cd)^{-1}, \quad \text{for all } \log^2 n \leqslant l < L.$$

Furthermore, $\|\Xi(l) - \Delta(l)\|_\infty < (Cd)^{-1}$ for all $l < \log^2 n$ by Lemma 3.21. Therefore, we can apply Corollary 3.20 to obtain

$$\|\Xi(L) - \Delta(L)\|_\infty \leqslant 2Cd^2 \sum_{j=1}^{L-1} (4d)^{j-1} \|\Xi(L-j)\|_\infty^2 + Cd^2 (4d)^{L-1} \|\Delta(0)\|_\infty^2. \tag{3.62}$$

Since $L \geqslant \log^2 n$ and $\lambda \geqslant 0.1d$, (3.60) and (3.61) imply that the sum on the right-hand side of (3.62) is dominated by the term for $j = L - 1$. Hence,

$$\begin{aligned} \|\Xi(L) - \Delta(L)\|_\infty &\leqslant 4Cd^2 \|\Xi(L-1)\|_\infty^2 + Cd^2 (4d)^{L-1} \delta^2 \\ &\leqslant c_3 d^2 \delta^2 \left[n^{-1} \lambda^{2L-2} + (4d)^{L-1}\right] \\ &\leqslant 2c_3 d^2 \delta^2 \lambda^{2L-2} n^{-1} \leqslant c_4 \delta^2 \lambda^{2L} n^{-1}. \end{aligned} \tag{3.63}$$

Combining (3.61) and (3.63), we conclude that $\|\Xi(L) - \Delta(L)\|_\infty < -\log(\epsilon) \cdot \|\Xi(L)\|_\infty^2$ (provided that ϵ is chosen small enough). □

Finally, Proposition 3.6 follows from Lemma 3.22 directly.

3.6. Proof of Proposition 3.7

Let $\mu = v\lambda^{L_2}$. Then Corollary 3.4 and Proposition 3.5 entail that

$$(1 - \epsilon^3)\mu \leqslant \Delta^a_{v\to w}(L_2) \leqslant (1 + \epsilon^3)\mu \qquad \text{if } v \in V_a \text{ and } w \in N(v), \quad \text{and} \tag{3.64}$$

$$\left(-\frac{1}{2} - \epsilon^3\right)\mu \leqslant \Delta^a_{v\to w}(L_2) \leqslant \left(-\frac{1}{2} + \epsilon^3\right)\mu \quad \text{if } v \notin V_a \text{ and } w \in N(v). \tag{3.65}$$

To prove Proposition 3.7, we consider two cases. The first case is that $\|\Delta(L_2)\|_\infty \leqslant (\epsilon d)^{-1}$ is 'small'. Then it will take two more steps for the messages to properly represent the colouring (V_1, V_2, V_3), *i.e.*, $L_3 = L_2 + 2$. In contrast, if $\|\Delta(L_2)\|_\infty > (\epsilon d)^{-1}$ is 'large', we will just need one more step ($L_3 = L_2 + 1$). In both cases the proof is based on a direct analysis of the BP equations (2.2).

Lemma 3.23. *If* $0.01\epsilon d^{-1} \leqslant \|\Delta(L_2)\|_\infty \leqslant (\epsilon d)^{-1}$, *then*

$$\eta^i_{u\to v}(L_2 + 1) = \begin{cases} \frac{1}{3} + (1 + \gamma(u, v, i))\beta & \textit{if } u \in V_i, \\ \frac{1}{3} - (1 + \gamma(u, v, i))\beta' & \textit{otherwise,} \end{cases} \tag{3.66}$$

where $|\gamma(u, v, i)| \leqslant \epsilon^3$ *and* $\beta, \beta' > \epsilon^2$.

Proof. We have

$$\begin{aligned}\eta^i_{v\to w}(L_2+1) &= \frac{\prod_{u\in N(v)\setminus w} 1-\frac{3}{2}\Delta^i_{u\to v}(L_2)}{\sum_{j=1}^3 \prod_{u\in N(v)\setminus w} 1-\frac{3}{2}\Delta^j_{u\to v}(L_2)} \\ &= \frac{\exp\left(-\frac{3}{2}\sum_{u\in N(v)\setminus w}\Delta^i_{u\to v}(L_2)+O(\Delta^i_{u\to v}(L_2))^2\right)}{\sum_{j=1}^3 \exp\left(-\frac{3}{2}\sum_{u\in N(v)\setminus w}\Delta^j_{u\to v}(L_2)+O(\Delta^j_{u\to v}(L_2))^2\right)} \\ &= \left[\sum_{j=1}^3 \exp\left[\frac{3}{2}\sum_{u\in N(v)\setminus w}\Delta^i_{u\to v}(L_2)-\Delta^j_{u\to v}(L_2)+O(\epsilon d)^{-2}\right]\right]^{-1}. \qquad (3.67)\end{aligned}$$

Since for any v we have $|N(v)|=2d$, we can essentially neglect the $O(\epsilon d)^{-2}$-term in (3.67). More precisely, for some $-\epsilon^2 \leqslant \gamma_2=\gamma_2(i,v,w)\leqslant \epsilon^2$, we have

$$\eta^i_{v\to w}(L_2+1)=(1+\gamma_2)\left[\sum_{j=1}^3 \exp\left[\frac{3}{2}\sum_{u\in N(v)\setminus w}\Delta^i_{u\to v}(L_2)-\Delta^j_{u\to v}(L_2)\right]\right]^{-1}. \qquad (3.68)$$

To analyse (3.68), assume without loss of generality that $v\in V_1$. Then (3.64) and (3.65) entail that there is a number $-\epsilon^2<\gamma_3<\epsilon^2$ such that

$$\sum_{u\in N(v)\setminus w}\Delta^1_{u\to v}(L_2)-\Delta^2_{u\to v}(L_2)=-\left(\frac{3}{2}+\gamma_3\right)d\mu.$$

Consequently, $\eta^1_{v\to w}(L_2+1)=(1+\gamma_2)\left[1+2\exp\left(-(3/2+\gamma_3)d\mu\right)\right]^{-1}$. Finally, since $\mu\leqslant 2(\epsilon d)^{-1}$, we obtain

$$\eta^1_{v\to w}(L_2+1)=(1+\gamma_4)\left[1+2\exp\left(-\frac{3}{2}d\mu\right)\right]^{-1} \qquad (3.69)$$

for some $-2\epsilon^2\leqslant\gamma_4=\gamma_4(1,v,w)\leqslant 2\epsilon^2$.

Now, assume that $v\in V_2$. Then (3.64) and (3.65) entail that there are numbers $-\epsilon^2<\gamma_5,\gamma_6<\epsilon^2$ such that

$$\sum_{u\in N(v)\setminus w}\Delta^1_{u\to v}(L_2)-\Delta^3_{u\to v}(L_2)=\gamma_5 d\mu,$$
$$\sum_{u\in N(v)\setminus w}\Delta^1_{u\to v}(L_2)-\Delta^2_{u\to v}(L_2)=(3/2+\gamma_6)d\mu.$$

Therefore,

$$\eta^2_{v\to w}(L_2+1)=(1+\gamma_4)\left[2+\exp\left(\frac{3}{2}d\mu\right)\right]^{-1} \qquad (3.70)$$

for some $-2\epsilon^2\leqslant\gamma_4=\gamma_4(2,v,w)\leqslant 2\epsilon^2$. Combining (3.69) and (3.70), we obtain the assertion. □

Corollary 3.24. *Suppose that* $0.01\epsilon d^{-1}\leqslant\|\Delta(L_2)\|_\infty\leqslant(\epsilon d)^{-1}$. *Then* $\eta^a_{v\to w}(L_2+2)\geqslant 0.99$ *if* $v\in V_a$, *and* $\eta^a_{v\to w}(L_2+2)\leqslant 0.01$ *if* $v\notin V_a$.

Proof. We assume without loss of generality that $a = 1$. Moreover, suppose that $v \in V_1$. We shall bound the quotient

$$\frac{\eta^1_{v\to w}(L_2+2)}{\eta^2_{v\to w}(L_2+2)} = Q_2 \cdot Q_3, \quad \text{where} \tag{3.71}$$
$$Q_j = \prod_{u \in V_j \cap N(v)\setminus w} \frac{1-\eta^1_{u\to v}(L_2+1)}{1-\eta^2_{u\to v}(L_2+1)}, \quad \text{for } j = 2, 3,$$

from below. Lemma 3.23 implies that, for $u \in V_3$,

$$\frac{1-\eta^1_{u\to v}(L_2+1)}{1-\eta^2_{u\to v}(L_2+1)} \geqslant \frac{2/3+(1-\epsilon^3)\beta'}{2/3+(1+\epsilon^3)\beta'} \geqslant 1+3\epsilon^3\beta' \geqslant 1-6\epsilon^3.$$

Hence,

$$Q_2 \geqslant (1-6\epsilon^3)^d. \tag{3.72}$$

Furthermore, for $u \in V_2$, Lemma 3.23 entails that

$$\frac{1-\eta^1_{u\to v}(L_2+1)}{1-\eta^2_{u\to v}(L_2+1)} \geqslant \frac{2/3+(1-\epsilon^3)\beta'}{2/3-(1+\epsilon^3)\beta} = 1+\frac{(1-\epsilon^3)(\beta+\beta')}{2/3-(1-\epsilon^3)\beta} \geqslant 1+2\epsilon^2.$$

Consequently,

$$Q_2 \geqslant (1+2\epsilon^2)^{d-1}. \tag{3.73}$$

Combining (3.72) and (3.73) and recalling that $d \gg \epsilon^{-2}$, we obtain the assertion. □

Lemma 3.25. *Suppose that* $\|\Delta(L_2)\|_\infty > (\epsilon d)^{-1}$*. Then* $\eta^a_{v\to w}(L_2+1) \geqslant 0.99$ *if* $v \in V_a$*, and* $\eta^a_{v\to w}(L_2+2) \leqslant 0.01$ *if* $v \notin V_a$*.*

Proof. Since $\|\Delta(L_2)\|_\infty > (\epsilon d)^{-1}$, (3.64) and (3.65) yield

$$\mu \geqslant (2\epsilon d)^{-1}. \tag{3.74}$$

Without loss of generality we may consider a vertex $v \in V_1$ and a neighbour $w \in N(v)$. We will prove that $\eta^1_{v\to w}(L_2+1)/\eta^2_{v\to w}(L_2+1) > 1000$. Since $\sum_{j=1}^3 \eta^j_{v\to w}(L_2+1) = 1$, this implies the assertion. To bound the quotient from below, we decompose

$$\frac{\eta^1_{v\to w}(L_2+1)}{\eta^2_{v\to w}(L_2+1)} = Q_2 \cdot Q_3, \quad \text{where} \tag{3.75}$$
$$Q_j = \prod_{u \in V_j \cap N(v)\setminus w} \frac{1-\eta^1_{u\to v}(L_2)}{1-\eta^2_{u\to v}(L_2)}, \quad \text{for } j = 2, 3,$$

With respect to Q_3, (3.64) and (3.65) imply that, for $u \in V_3$,

$$\frac{1-\eta^1_{u\to v}(L_2)}{1-\eta^2_{u\to v}(L_2)} \geqslant \frac{2/3+(1/2-\epsilon^3)\mu)}{2/3+(1/2+\epsilon^3)\mu} = 1-\frac{2\epsilon^3\mu}{2/3+(1/2+\epsilon^3)\mu} \geqslant 1-3\epsilon^3\mu.$$

Hence,

$$Q_3 \geqslant (1-3\epsilon^3\mu)^d. \tag{3.76}$$

Further, (3.64) and (3.65) yield that, for $u \in V_2$,

$$\frac{1-\eta^1_{u\to v}(L_2)}{1-\eta^2_{u\to v}(L_2)} \geqslant \frac{2/3+(1/2-\epsilon^3)\mu)}{2/3-(1-\epsilon^3)\mu} = 1 + \frac{(3/2-2\epsilon^3)\mu}{2/3+(1/2-(1-\epsilon^3))\mu} \geqslant 1+2\mu.$$

Therefore,

$$Q_2 \geqslant (1+2\mu)^{d-1}. \tag{3.77}$$

Thus, combining (3.74)–(3.77), we obtain

$$\frac{\eta^1_{v\to w}(L_2+1)}{\eta^2_{v\to w}(L_2+1)} = Q_2 \cdot Q_3 \geqslant (1-3\epsilon^3\mu)^d(1+2\mu)^{d-1} \geqslant (1+\mu)^{d-1} \geqslant 1000,$$

which implies the assertion. □

Finally, Proposition 3.7 is a direct consequence of Corollary 3.24 and Lemma 3.25.

4. Proof of Corollary 1.2

Throughout this section, we assume that $d \geqslant d_0$ for a sufficiently large constant $d_0 > 0$, and that $n > n_0 = n_0(d)$ for a large enough n_0. Set $p = d/n$.

Let $G = G_{n,d,3}$ be a random graph with vertex set $V = \{1, \dots, 3n\}$ and 'planted' 3-colouring V_1, V_2, V_3. In order to analyse the adjacency $A(G)$, we shall employ the following lemma, which follows immediately from the 'converse expander mixing lemma' from [3].

Lemma 4.1. *Let $B = (V' \cup V'', E_B)$ be a bipartite d-regular graph such that $|V'| = |V''|$. Assume that*

$$\forall S \subset V',\ T \subset V'' : |e_B(S,T) - |S||T|p| \leqslant d^{0.51}\sqrt{|S||T|}, \tag{4.78}$$

where $e_B(S,T)$ is the number of S–T-edges in B. Then the adjacency matrix $A(B)$ enjoys the property:

For any two vectors $\xi, \eta \in \mathbb{R}^{V' \cup V''}$ such that both ξ, η are perpendicular to $\vec{1}_{V'}$ and $\vec{1}_{V''}$, the inequality $\langle A(B)\xi, \eta\rangle \leqslant d^{0.52}\|\xi\|\|\eta\|$ holds.

Moreover, the following lemma can be derived using standard techniques from the theory of random regular graphs; see Chapter 9 of [10].

Lemma 4.2. *With high probability, G has the following property. Let $1 \leqslant i < j \leqslant 3$. Then*

$$\forall S \subset V_i,\ T \subset V_j : |e_G(S,T) - |S||T|p| \leqslant d^{0.51}\sqrt{|S||T|}.$$

Corollary 4.3. *With high probability, G is $(d, 0.01)$-regular.*

Proof. Let $A(G) = (a_{v,w})_{v,w\in V}$ denote the adjacency matrix of G. Moreover, let

$$a^{ij}_{vw} = \begin{cases} a_{vw} & \text{if } v,w \in V_i \cup V_j \\ 0 & \text{otherwise} \end{cases} \qquad (1 \leqslant i < j \leqslant 3).$$

Then $A^{ij} = (a^{ij}_{vw})_{v,w \in V}$ is the adjacency matrix of the bipartite subgraph of G induced on $V_i \cup V_j$. Let $\mathcal{E}$ be the subspace of $\mathbb{R}^V$ spanned by the three vectors $\vec{1}_{V_k}$ $(k = 1,2,3)$. Combining Lemma 4.1 with Lemma 4.2, we conclude that w.h.p. $\langle A^{ij}\xi, \eta\rangle \leqslant d^{0.52}\|\xi\|\|\eta\|$ for all $\xi, \eta \perp \mathcal{E}$ and any $1 \leqslant i < j \leqslant 3$. Since $A(G) = \sum_{1 \leqslant i < j \leqslant 3} A^{ij}$, this implies that

$$\forall \xi, \eta \perp \mathcal{E} : \langle A(G)\xi, \eta\rangle \leqslant 0.01d\|\xi\|\|\eta\| \tag{4.79}$$

(provided that d is sufficiently large). Furthermore, as the construction of G ensures that each vertex $v \in V_i$ has exactly d neighbours in each class $V_j \neq V_i$, we can compute the vector $\zeta^i = A(G)\vec{1}_{V_i}$ as follows. For any $v \in V$,

$$\zeta^i_v = \sum_{w \in V_i} a_{vw} = \begin{cases} 0 & \text{if } v \in V_i, \\ d & \text{if } v \notin V_i. \end{cases}$$

Hence, $\zeta^i = A(G)\vec{1}_{V_i} = d\sum_{j \neq i} \vec{1}_{V_j}$. Therefore, for any $1 \leqslant i < j \leqslant 3$ we have

$$A(G)(\vec{1}_{V_i} - \vec{1}_{V_j}) = -d(\vec{1}_{V_i} - \vec{1}_{V_j}). \tag{4.80}$$

Combining (4.79) and (4.80), we see that G is $(d, 0.01)$-regular w.h.p. □

Finally, Corollary 1.2 follows from Theorem 1.1 and Corollary 4.3.

5. Conclusion

We have shown that `BPCol` 3-colours $(d, 0.01)$-regular graphs in polynomial time. Three potentially interesting extensions suggest themselves, which may be the subject of future work.

(1) In $(d, 0.01)$-regular graphs every vertex has precisely d neighbours in each colour class except for its own. By comparison, in the planted random graph model studied in [2], the number of neighbours that a vertex has in another colour class is Poisson with mean d. It would be interesting to see if/how the present analysis can be modified to deal with such a more irregular degree distribution.

(2) Survey propagation ('SP') is a more involved version of belief propagation (although SP can be rephrased as BP on a different model [15]) and performs very well empirically on random graphs $G(n, p)$. It would be interesting to extend our analysis to SP.

(3) In a $(d, 0.01)$-regular graph there is exactly one 3-colouring (up to permutations of the colour classes). Nonetheless, we think that the techniques of our analysis can be extended to more complicated 'solution spaces'. For instance, it should be straightforward to deal with graphs that have a bounded number of distinct 3-colourings.

References

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

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

[3] Bilu, Y. and Linial, N. (2006) Lifts, discrepancy and nearly optimal spectral gap. *Combinatorica* **26** 495–519.
[4] Braunstein, A., Mézard, M., Weigt, M. and Zecchina, R. (2005) Constraint satisfaction by survey propagation. In *Computational Complexity and Statistical Physics* (A. Percus, G. Istrate and C. Moore, eds), Oxford University Press.
[5] Braunstein, A., Mézard, M. and Zecchina, R. (2005) Survey propagation: An algorithm for satisfiability. *Random Struct. Alg.* **27** 201–226.
[6] Braunstein, A., Mulet, R., Pagnani, A., Weigt, M. and Zecchina, R. (2003) Polynomial iterative algorithms for coloring and analyzing random graphs. *Phys. Rev. E* **68** 036702.
[7] Brightwell, G. R. and Winkler, P. (2002) Random colorings of a Cayley tree. In *Contemporary Combinatorics*, Vol. 10 of *Bolyai Society Mathematical Studies*, János Bolyai Math. Soc., pp. 247–276.
[8] Chung, F. and Graham, R. (2002) Sparse quasi-random graphs. *Combinatorica* **22** 217–244.
[9] Feige, U., Mossel, E. and Vilenchik, D. (2006) Complete convergence of message passing algorithms for some satisfiability problems. In *Random 2006*, Vol. 4110 of *Lecture Notes in Computer Science*, pp. 339–350.
[10] Janson, S., Łuczak, T. and Ruciński, A. (2000) *Random Graphs*, Wiley.
[11] Jonasson, J. (2002) Uniqueness of uniform random colorings of regular trees. *Statist. Probab. Lett.* **57** 243–248.
[12] Krzakala, F., Montanari, A., Ricci-Tersenghi, F., Semerjianc, G. and Zdeborova, L. (2007) Gibbs states and the set of solutions of random constraint satisfaction problems. *Proc. Nat. Acad. Sci.* **104** 10318–10323.
[13] Luby, M., Mitzenmacher, M., Shokrollahi, M. A. and Spielman, D. (1998) Analysis of low density parity check codes and improved designs using irregular graphs. In *Proc. 30th ACM Symposium on the Theory of Computing*, pp. 249–258.
[14] Luby, M., Mitzenmacher, M., Shokrollahi, M. A. and Spielman, D. (2001) Efficient erasure correcting codes. *IEEE Trans. Inform. Theory* **47** 569–584.
[15] Maneva, E., Mossel, E. and Wainwright, M. (2005) A new look at survey propagation and its generalizations. In *Proc. 16th ACM–SIAM Symposium on Discrete Algorithms*, pp. 1089–1098.
[16] Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Francisco, CA.
[17] Richardson, T., Shokrollahi, A. and Urbanke, R. (2001) Design of capacity-approaching irregular low-density parity check codes. *IEEE Trans. Inform. Theory* **47** 619–637.
[18] Tatikonda, S. and Jordan, M. I. (2002) Loopy belief propagation and Gibbs measures. In *Uncertainty in Artificial Intelligence (UAI): Proc. 18th Conference*.
[19] Weitz, D. (2006) Counting independent sets up to the tree threshold. In *Proc. 38th Annual ACM Symposium on the Theory of Computing*, pp. 140–149.
[20] Yamamoto, M. and Watanabe, O. (2007) Belief propagation and spectral methods. Report C–248, Department of Mathematical and Computing Sciences, Tokyo Institute of Technology.