The Minimum Bisection in the Planted Bisection Model*

Amin Coja-Oghlan[1], Oliver Cooley[2], Mihyun Kang[2], and Kathrin Skubch[1]

1 **Goethe University, Mathematics Institute**
10 Robert Mayer St, Frankfurt 60325, Germany
{acoghlan,skubch}@math.uni-frankfurt.de

2 **Graz University of Technology, Institute of Optimization and Discrete Mathematics (Math B)**
Steyrergasse 30, 8010 Graz, Austria
{cooley,kang}@math.tugraz.at

Abstract

In the *planted bisection* model a random graph $\boldsymbol{G}(n, p_+, p_-)$ with n vertices is created by partitioning the vertices randomly into two classes of equal size (up to ± 1). Any two vertices that belong to the same class are linked by an edge with probability p_+ and any two that belong to different classes with probability $p_- < p_+$ independently. The planted bisection model has been used extensively to benchmark graph partitioning algorithms. If $p_\pm = 2d_\pm/n$ for numbers $0 \le d_- < d_+$ that remain fixed as $n \to \infty$, then w.h.p. the "planted" bisection (the one used to construct the graph) will not be a minimum bisection. In this paper we derive an asymptotic formula for the minimum bisection width under the assumption that $d_+ - d_- > c\sqrt{d_+ \ln d_+}$ for a certain constant $c > 0$.

1998 ACM Subject Classification G.2.2 Graph Theory

Keywords and phrases Random graphs, minimum bisection, planted bisection, belief propagation

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2015.710

1 Introduction

1.1 Background and motivation

Since the early days of computational complexity graph partitioning problems have played a central role in computer science [18, 24]. Over the years they have inspired some of the most important algorithmic techniques that we have at our disposal today, such as network flows or semidefinite programming [3, 17, 19, 25, 38].

In the context of the probabilistic analysis of algorithms, it is hard to think of a more intensely studied problem than the *planted bisection model*. In this model a random graph $\boldsymbol{G} = \boldsymbol{G}(n, p_{+1}, p_{-1})$ on $[n] = \{1, \ldots, n\}$ is created by choosing a map $\boldsymbol{\sigma} : V \to \{-1, 1\}$ uniformly at random subject to $||\boldsymbol{\sigma}^{-1}(1)| - |\boldsymbol{\sigma}^{-1}(-1)|| \le 1$ and connecting any two vertices

* The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 278857–PTCC. Supported by Austrian Science Fund (FWF): P26826 and W1230, Doctoral Program "Discrete Mathematics".

18th Int'l Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX'15) / 19th Int'l Workshop on Randomization and Computation (RANDOM'15).
Editors: Naveen Garg, Klaus Jansen, Anup Rao, and José D. P. Rolim; pp. 710–725
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

$v \neq w$ with probability $p_{\boldsymbol{\sigma}(v)\boldsymbol{\sigma}(w)}$ independently, where $0 \le p_{-1} < p_{+1} \le 1$. To ease notation, we often write p_+ for p_{+1} and p_- for p_{-1}, and handle subscripts similarly for other parameters.

Given the random graph $\boldsymbol{G}$ (but not the planted bisection $\boldsymbol{\sigma}$), the task is to find a *minimum bisection* of $\boldsymbol{G}$, i.e., to partition the vertices into two disjoint sets $S, \bar{S} = [n] \setminus S$ whose sizes satisfy $||S| - |\bar{S}|| \le 1$ such that the number of S-$\bar{S}$-edges is minimum. The planted bisection model has been employed to gauge algorithms based on spectral, semidefinite programming, flow and local search techniques, to name but a few [5, 6, 7, 8, 9, 11, 14, 15, 16, 22, 23, 27, 31, 29].

Remarkably, for a long time the algorithm with the widest range of $n, p_\pm$ for which a minimum bisection can be found efficiently was one of the earliest ones, namely Boppana's spectral algorithm [6]. It succeeds if

$$n(p_+ - p_-) \ge c\sqrt{np_+ \ln n}$$

for a certain constant $c > 0$. Under this assumption the planted bisection is minimum w.h.p. In fact, recently the critical value $c^* > 0$ for which this statement is true was identified explicitly [36]. In particular, for $n(p_+ - p_-) > c^*\sqrt{np_+ \ln n}$ the minimum bisection width simply equals $(\frac{1}{4} + o(1))n^2 p_-$ w.h.p.

But if $n(p_+ - p_-) < c^*\sqrt{np_+ \ln n}$, then the minimum bisection width will be strictly smaller than the width of the planted bisection w.h.p. Yet there is another spectral algorithm [9] that finds a minimum bisection w.h.p. under the weaker assumption that

$$n(p_+ - p_-) \ge c\sqrt{np_+ \ln(np_+)}, \tag{1.1}$$

for a certain constant $c > 0$, and even certifies the optimality of its solution. However, [9] does not answer what is arguably the most immediate question: what is the *typical* value of the minimum bisection width?

In this paper we derive the value to which the (suitably scaled) minimum bisection width converges in probability. We confine ourselves to the case that $\frac{n}{2}p_\pm = d_\pm$ remain fixed as $n \to \infty$. Hence, the random graph $\boldsymbol{G}$ has bounded average degree. This is arguably the most interesting case because the discrepancy between the planted and the minimum bisection gets larger as the graphs get sparser. In fact, it is easy to see that in the case of fixed $\frac{n}{2}p_\pm = d_\pm$ the difference between the planted and the minimum bisection width is $\Theta(n)$ as the planted bisection is not even locally optimal w.h.p.

Although we build upon some of the insights from [9], it seems difficult to prove our main result by tracing the fairly complicated algorithm from that paper. Instead, our main tool is an elegant message passing algorithm called *Warning Propagation* that plays an important role in the study of random constraint satisfaction problems via ideas from statistical physics [32]. Running Warning Propagation on $\boldsymbol{G}$ naturally corresponds to a fixed point problem on the 2-simplex, and the minimum bisection width can be cast as a function of the fixed point.

1.2 The main result

To state the fixed point problem, we consider the functions

$$\psi : \mathbb{R} \to \mathbb{R}, \quad x \mapsto \begin{cases} -1 & \text{if } x < -1 \\ x & \text{if } -1 \le x \le 1 \\ 1 & \text{if } x > 1, \end{cases} \qquad \tilde{\psi} : \mathbb{R} \to \mathbb{R}, \quad x \mapsto \begin{cases} -1 & \text{if } x \le -1 \\ 1 & \text{if } x > -1. \end{cases}$$

Let $\mathcal{P}(\{-1,0,1\})$ be the set of probability measures on $\{-1,0,1\}$. Clearly, we can identify $\mathcal{P}(\{-1,0,1\})$ with the set of all maps $p:\{-1,0,1\}\to[0,1]$ such that $p(-1)+p(0)+p(1)=1$, i.e., the 2-simplex. Further, let us define a map

$$\mathcal{T}_{d_+,d_-}:\mathcal{P}(\{-1,0,1\})\to\mathcal{P}(\{-1,0,1\}) \tag{1.2}$$

as follows. Given $p\in\mathcal{P}(\{-1,0,1\})$, let $(\eta_{p,i})_{i\geq 1}$ be a family of i.i.d. $\{-1,0,1\}$-valued random variables with distribution p. Moreover, let $\gamma_\pm=\mathrm{Po}(d_\pm)$ be Poisson variables that are independent of each other and of the $\eta_{p,i}$. Let

$$Z_{p,d_+,d_-}:=\sum_{i=1}^{\gamma_+}\eta_{p,i}-\sum_{i=\gamma_++1}^{\gamma_++\gamma_-}\eta_{p,i}. \tag{1.3}$$

Then we let $\mathcal{T}_{d_+,d_-}(p)\in\mathcal{P}(\{-1,0,1\})$ be the distribution of $\psi(Z_{p,d_+,d_-})$. Further, with $(\eta_{p,i})_{i\geq 1}$ and $\gamma_\pm$ as before, let

$$\begin{aligned}\varphi_{d_+,d_-}:&\mathcal{P}(\{-1,0,1\})\to\mathbb{R},\\ &p\mapsto\frac{1}{2}\mathbb{E}\left[\sum_{i=1}^{\gamma_+}\mathbf{1}\{\eta_{p,i}=-\tilde{\psi}(Z_{p,d_+,d_-})\}+\sum_{i=\gamma_++1}^{\gamma_++\gamma_-}\mathbf{1}\{\eta_{p,i}=\tilde{\psi}(Z_{p,d_+,d_-})\}\right].\end{aligned}$$

Moreover, let us call $p\in\mathcal{P}(\{-1,0,1\})$ *skewed* if $p(1)\geq 1-d_+^{-10}$. Finally, we denote the minimum bisection width of a graph G by $\mathrm{bis}(G)$.

▶ **Theorem 1.1.** *There exists a constant $c>0$ such that for any $d_\pm>0$ satisfying $d_+\geq 2$ and $d_+-d_-\geq c\sqrt{d_+\ln d_+}$ the map $\mathcal{T}_{d_+,d_-}$ has a unique skewed fixed point p^* and $n^{-1}\mathrm{bis}(\boldsymbol{G})$ converges in probability to $\varphi_{d_+,d_-}(p^*)$.*

Note that $\mathcal{T}_{d_+,d_-}$ may have further fixed points besides p^*, but p^* is the only fixed point which is skewed. We also note that the condition $d_+\geq 2$ is not optimised – any constant larger than 1 would do as a lower bound, but then in any case the condition $d_+\geq 2$ follows from the lower bound on d_+-d_- for sufficiently large c.

In the following sections we will use that the assumptions of Theorem 1.1 allow us to assume that also d_+ is sufficiently large.

1.3 Further related work

Determining the minimum bisection width of a graph is NP-hard [18] and there is evidence that the problem does not even admit a PTAS [26]. On the positive side, it is possible to approximate the minimum bisection width within a factor of $O(\ln n)$ for graphs on n vertices in polynomial time [38].

The planted bisection model has been studied in statistics under the name "stochastic block model" [20]. However, in the context of statistical inference the aim is to recover the planted partition $\boldsymbol{\sigma}$ as best as possible given $\boldsymbol{G}$ rather than to determine the minimum bisection width. Recently there has been a lot of progress, much of it inspired by non-rigorous work [12], on the statistical inference problem. The current status of the problem is that matching upper and lower bounds are known for the values of $d_\pm$ for which it is possible to obtain a partition that is non-trivially correlated with $\boldsymbol{\sigma}$ [30, 33, 35]. Furthermore, there are algorithms that recover a best possible approximation to $\boldsymbol{\sigma}$ under certain conditions on $d_\pm$ [1, 34, 36]. But since our objective is different, the methods employed in the present paper are somewhat different and, indeed, rather simpler.

Finally, there has been recent progress on determining the minimum bisection width on the Erdős-Rényi random graph. Although its precise asymptotics remain unknown in the case of bounded average degrees d, it was proved in [13] that the main correction term corresponds to the "Parisi formula" in the Sherrington-Kirkpartrick model [39]. Additionally, regarding the case of very sparse random graphs (i.e. with constant average degree), there is a sharp threshold (at $np = \ln 4$) for the minimum bisection width to be linear in n [28].

Generally speaking, the approach that we pursue is somewhat related to the notion of "local weak convergence" of graph sequences as it was used in [2]. More specifically, we are going to argue that the minimum bisection width of $\boldsymbol{G}$ is governed by the "limiting local structure" of the graph, which is a two-type Galton-Watson tree. The fixed point problem in Theorem 1.1 mirrors the execution of a message passing algorithm on the Galton-Watson tree. The study of this fixed point problem, for which we use the *contraction method* [37], is the key technical ingredient of our proof. We believe that this strategy provides an elegant framework for tackling many other problems in the theory of random graphs as well. In fact, in a recent paper [10] we combined Warning Propagation with a fixed point analysis on Galton-Watson trees to the k-core problem, and in [4] Warning Propagation was applied to the random graph coloring problem.

2 Outline

From here on we keep the notation and the assumptions of Theorem 1.1. In particular, we assume that $d_+ - d_- \geq c\sqrt{d_+ \ln d_+}$ for a large enough constant $c > 0$ and that $d_\pm$ remain fixed as $n \to \infty$. Furthermore we assume that d_+ is bounded from below by a large enough constant. Throughout the paper all graphs will be locally finite and of countable size.

Three main insights enable the proof of Theorem 1.1. The first one, which we borrow from [9], is that w.h.p. $\boldsymbol{G}$ features a fairly large set $\mathcal{C}$ of vertices such that for any two optimal bisections τ_1, τ_2 of $\boldsymbol{G}$ (i.e. maps $\tau_1, \tau_2 : V(\boldsymbol{G}) \to \{\pm 1\}$), we either have $\tau_1(v) = \tau_2(v)$ for all $v \in \mathcal{C}$ or $\tau_1(v) = -\tau_2(v)$ for all $v \in \mathcal{C}$. In the language of random constraint satisfaction problems, the vertices in $\mathcal{C}$ are "frozen". While there remain $\Omega(n)$ unfrozen vertices, the subgraph that they induce is subcritical, i.e., all components are of size $O(\ln n)$ and indeed most are of bounded size.

The second main ingredient is an efficient message passing algorithm called *Warning Propagation*, (cf. [32, Chapter 19]). We will show that a bounded number of Warning Propagation iterations suffice to arrange almost all of the unfrozen vertices optimally (i.e. to assign almost all of the vertices to two classes such that there is a minimum bisection respecting this assignment) and thus to obtain a very good approximation to the minimum bisection w.h.p. (Proposition 2.2). This insight reduces our task to tracing Warning Propagation for a bounded number of rounds.

This last problem can be solved by studying Warning Propagation on a suitable Galton-Watson tree, because $\boldsymbol{G}$ only contains a negligible number of short cycles w.h.p. (Lemma 2.3). Thus, the analysis of Warning Propagation on the random tree is the third main ingredient of the proof. This task will turn out to be equivalent to studying the fixed point problem from Section 1.2 (Proposition 2.5). We proceed to outline the three main components of the proof.

2.1 The core

Given a vertex u of a graph G let $\partial_G u$ denote the neighbourhood of u in G. We sometimes omit the subscript G when the graph is clear from the context. More particularly, in the random graph $\boldsymbol{G}$, let $\partial_\pm u$ denote the set of all neighbours w of u in $\boldsymbol{G}$ with $\boldsymbol{\sigma}(w)\boldsymbol{\sigma}(v) = \pm 1$. Following [9], we define $\mathcal{C}$ as the largest subset $U \subset [n]$ such that

$$||\partial_\pm u| - d_\pm| \leq \frac{c}{4}\sqrt{d_+ \ln d_+} \quad \text{and} \quad |\partial u \setminus U| \leq 100 \quad \text{for all } u \in U. \tag{2.1}$$

Clearly, the set $\mathcal{C}$, which we call the *core*, is uniquely defined because any union of sets U that satisfy (2.1) also has the property. Let $\boldsymbol{\sigma}_{\mathcal{C}} : \mathcal{C} \to \{\pm 1\}$, $v \mapsto \boldsymbol{\sigma}(v)$ be the restriction of the "planted assignment" to $\mathcal{C}$.

Furthermore, for a graph G, a set $U \subset V(G)$ and a map $\sigma : U \to \{-1, 1\}$ we let

$$\mathrm{cut}(G, \sigma) := \min\left\{ \sum_{\{v,w\} \in E(G)} \frac{1 - \tau(v)\tau(w)}{2} \;\middle|\; \tau : V(G) \to \{\pm 1\} \text{ satisfies } \tau(v) = \sigma(v) \text{ for all } v \in U \right\}.$$

In words, $\mathrm{cut}(G, \sigma)$ is the smallest number of edges in a cut of G that separates the vertices in $U \cap \sigma^{-1}(-1)$ from those in $U \cap \sigma^{-1}(1)$. In particular, $\mathrm{cut}(G, \sigma_{\mathcal{C}})$ is the smallest cut of G that separates the vertices in the core $\mathcal{C}$ that are frozen to -1 from those that are frozen to 1.

Finally, for any vertex v we define a set $\mathcal{C}_v = \mathcal{C}_v(G, \sigma)$ of vertices via the following process.

C1 Let $\mathcal{C}_v^{(0)} = \{v\} \cup \partial_G v$.

C2 Inductively, let $\mathcal{C}_v^{(t+1)} = \mathcal{C}_v^{(t)} \cup \bigcup_{u \in \mathcal{C}_v^{(t)} \setminus \mathcal{C}} \partial_G u$ and let $\mathcal{C}_v = \bigcup_{t \geq 0} \mathcal{C}_v^{(t)}$.

▶ **Lemma 2.1** ([9], Proposition 19 and Section 3.6)**.** *We have* $\mathrm{bis}(\boldsymbol{G}) = \mathrm{cut}(\boldsymbol{G}, \boldsymbol{\sigma}_{\mathcal{C}})$ *and* $|\mathcal{C}| \geq n(1 - d_+^{-100})$ *w.h.p. Furthermore, for any* $\varepsilon > 0$ *there exists* $\omega > 0$ *such that w.h.p.* $\sum_{v \in [n]} |\mathcal{C}_v| \cdot \mathbf{1}\{|\mathcal{C}_v| \geq \omega\} \leq \varepsilon n$.

2.2 Warning Propagation

To calculate $\mathrm{cut}(\boldsymbol{G}, \boldsymbol{\sigma}_{\mathcal{C}})$ we adopt the *Warning Propagation* ("WP") message passing algorithm[1]. Let us first introduce WP for a generic graph $G = (V(G), E(G))$ and a map $\sigma : U \subset V(G) \to \{-1, 1\}$. At each time $t \geq 0$, WP sends a "message" $\mu_{v \to w}(t|G, \sigma) \in \{-1, 0, 1\}$ from v to w for any edge $\{v, w\} \in E(G)$. The messages are directed objects, i.e., $\mu_{v \to w}(t|G, \sigma)$ and $\mu_{w \to v}(t|G, \sigma)$ may differ. They are defined inductively by

$$\mu_{v \to w}(0|G, \sigma) := \begin{cases} \sigma(v) & \text{if } v \in U, \\ 0 & \text{otherwise,} \end{cases} \qquad \mu_{v \to w}(t+1|G, \sigma) := \psi\left(\sum_{u \in \partial v \setminus w} \mu_{u \to v}(t|G, \sigma)\right). \tag{2.2}$$

Note that U does not appear explicitly in the notation $\mu_{v \to w}(t|G, \sigma)$ despite being integral to the definition – it is however implicit in the notation since U is the domain of σ.

Thus, the WP messages are initialised according to σ. Subsequently, v sends message ± 1 to w if it receives more ± 1 than ∓ 1 messages from its neighbours $u \neq w$. If there is a tie, v

[1] A discussion of Warning Propagation in the context of the "cavity method" from statistical physics can be found in [32].

sends out 0. Finally, for $t \geq 0$ define

$$\mu_v(t|G,\sigma) := \sum_{w \in \partial v} \mu_{w \to v}(t|G,\sigma).$$

The intuition is that the message $\mu_{v \to w}$ which v sends to w indicates which class v is most likely to be in based on the current local information it receives from its other neighbours. To minimise the cut, we would like to place v into the class in which most of its neighbours lie. The initialisation is given by the set U, which we will choose to be the core.

▶ **Proposition 2.2.** *For any $\varepsilon > 0$ there exists $t_0 = t_0(\varepsilon, d_+, d_-)$ such that for all $t \geq t_0$ w.h.p.*

$$\left| \mathrm{cut}(\boldsymbol{G}, \boldsymbol{\sigma}_{\mathcal{C}}) - \frac{1}{2} \sum_{v \in [n]} \sum_{w \in \partial v} \mathbf{1}\{\mu_{w \to v}(t|\boldsymbol{G},\boldsymbol{\sigma}) = -\tilde{\psi}\left(\mu_v(t|\boldsymbol{G},\boldsymbol{\sigma})\right)\} \right| \leq \varepsilon n.$$

We defer the proof of Proposition 2.2 to Section 3.

2.3 The local structure

Proposition 2.2 shows that w.h.p. in order to approximate $\mathrm{cut}(\boldsymbol{G}, \boldsymbol{\sigma}_{\mathcal{C}})$ up to a small error of εn we merely need to run WP for a number t_0 of rounds that is bounded in terms of ε. The upshot is that the WP messages $\mu_{w \to v}(t|\boldsymbol{G},\boldsymbol{\sigma})$ that are required to figure out the minimum bisection width are determined by the *local* structure of $\boldsymbol{G}$. We show that the local structure of $\boldsymbol{G}$ "converges to" a suitable Galton-Watson tree. For this purpose, for simplicity we always say that the number of potential neighbours of any vertex in each class is $n/2$. This ignores the fact that if n is odd the classes do not have quite this size and the fact that a vertex cannot be adjacent to itself. However, ignoring these difficulties will not affect our calculations in any significant way.

Our task boils down to studying WP on that Galton-Watson tree. Specifically, let $\boldsymbol{T} = \boldsymbol{T}_{d_+,d_-}$ be the Galton-Watson tree with two types $+1, -1$ and offspring matrix

$$\begin{pmatrix} \mathrm{Po}(d_+) & \mathrm{Po}(d_-) \\ \mathrm{Po}(d_-) & \mathrm{Po}(d_+) \end{pmatrix}. \tag{2.3}$$

Hence, a vertex of type ± 1 spawns $\mathrm{Po}(d_+)$ vertices of type ± 1 and independently $\mathrm{Po}(d_-)$ vertices of type ∓ 1. Moreover, the type of the root vertex $r_{\boldsymbol{T}}$ is chosen uniformly at random. Let $\boldsymbol{\tau} = \boldsymbol{\tau}_{d_+,d_-} : V(\boldsymbol{T}) \to \{\pm 1\}$ assign each vertex of $\boldsymbol{T}$ its type.

The random graph $(\boldsymbol{G}, \boldsymbol{\sigma})$ "converges to" $(\boldsymbol{T}, \boldsymbol{\tau})$ in the following sense. For two triples (G, r, σ), (G', r', σ') of graphs G, G', root vertices $r \in V(G)$, $r' \in V(G')$ and maps $\sigma : V(G) \to \{\pm 1\}$, $\sigma' : V(G') \to \{\pm 1\}$ we write $(G, \sigma) \cong (G', \sigma')$ if there is a graph isomorphism $\varphi : G \to G'$ such that $\varphi(r) = r'$ and $\sigma = \sigma' \circ \varphi$. Further, we denote by $\partial^t(G, r, \sigma)$ the rooted graph obtained from (G, r) by deleting all vertices at distance greater than t from r together with the restriction of σ to this subgraph. The following lemma characterises the local structure of $(\boldsymbol{G}, \boldsymbol{\sigma})$.

▶ **Lemma 2.3.** *Let $t > 0$ be an integer and let T be any tree with root r and map $\tau : V(T) \to \{\pm 1\}$. Then*

$$\frac{1}{n} \sum_{v \in [n]} \mathbf{1}\left\{\partial^t(\boldsymbol{G}, v, \boldsymbol{\sigma}) \cong \partial^t(T, r, \tau)\right\} \quad \stackrel{n \to \infty}{\to} \quad \mathbb{P}\left[\partial^t(\boldsymbol{T}, r_{\boldsymbol{T}}, \boldsymbol{\tau}) \cong \partial^t(T, r, \tau)\right] \quad \text{in probability.}$$

Furthermore, w.h.p. $\boldsymbol{G}$ does not contain more than $\ln n$ vertices v such that $\partial^t(\boldsymbol{G}, v, \boldsymbol{\sigma})$ contains a cycle.

Proof. Given a tree T with root r and map $\tau : V(T) \to \{\pm 1\}$, let

$$X_t = X_t(T, r, \tau) = \frac{1}{n} \sum_{v \in [n]} \mathbf{1}\left\{\partial^t(\boldsymbol{G}, v, \boldsymbol{\sigma}) \cong \partial^t(T, r, \tau)\right\}$$

and

$$p_t = p_t(T, r, \tau) = \mathbb{P}\left[\partial^t(\boldsymbol{T}, r_{\boldsymbol{T}}, \boldsymbol{\tau}) \cong \partial^t(T, r, \tau)\right].$$

The proof proceeds by induction on t. If $t = 0$, pick a vertex $\boldsymbol{v} \in [n]$ uniformly at random, then $X_0 = \mathbb{P}_{\boldsymbol{v}}(\boldsymbol{\sigma}(\boldsymbol{v}) = \tau(r)) = \frac{1}{2}$ and $p_0 = \mathbb{P}_{\boldsymbol{T}}(\boldsymbol{\tau}(r_{\boldsymbol{T}}) = \tau(r)) = \frac{1}{2}$ for any $\tau(r) \in \{\pm 1\}$. To proceed from t to $t+1$, let d denote the number of children $v_1, \ldots, v_d$ of r in T. For each $i = 1, \ldots, d$, let T_i denote the tree rooted at v_i in the forest obtained from T by removing r and let $\tau_i : V(T_i) \to \{\pm 1\}$ denote the restriction of τ to the vertex set of T_i. Finally, let $C_1, \ldots, C_{\tilde{d}}$ for some $\tilde{d} \le d$ denote the distinct isomorphism classes among $\{\partial^t(T_i, v_i, \tau_i) : i = 1, \ldots, d\}$, and let $c_j = |\{i : \partial^t(T_i, v_i, \tau_i) \in C_j\}|$. Let $v \in [n]$ be an arbitrary vertex in $\boldsymbol{G}$. Our aim is to determine the probability of the event $\{\partial^{t+1}(\boldsymbol{G}, v, \boldsymbol{\sigma}) \cong \partial^{t+1}(T, r, \tau)\}$. Therefore, we think of $\boldsymbol{G}$ as being created in three rounds. First, partition $[n]$ in two classes. Second, randomly insert edges between vertices in $[n] \setminus \{v\}$ according to their planted sign. Finally, reveal the neighbours of v. For the above event to happen, v must have d neighbours in $\boldsymbol{G}$. Since $|\partial_{\pm} v|$ are independent binomially distributed random variables with parameters $\frac{n}{2}$ and $p_{\pm}$ and because $\frac{n}{2} p_{\pm} = d_{\pm}$, we may approximate $|\partial_{\pm} v|$ with a poisson distribution, and v has degree d with probability

$$\frac{(d_+ + d_-)^d}{d! \exp(d_+ + d_-)} + \mathrm{o}(1).$$

Conditioned on v having degree d, by induction v is adjacent to precisely c_j vertices with neighbourhood isomorphic to $\partial^t(T_i, v_i, \tau_i) \in C_j$ with probability

$$\binom{d}{c_1 \ldots c_{\tilde{d}}} \prod_{j=1}^{\tilde{d}} p_t(C_j) + \mathrm{o}(1).$$

The number of cycles of length $\ell \le 2t + 3$ in $\boldsymbol{G}$ is stochastically bounded by the number of such cycles in $\boldsymbol{G}(n, d_+/n)$ (the standard 1-type binomial random graph). For each ℓ, this number tends in distribution to a poisson variable with bounded mean (see e.g. Theorem 3.19 in [21]) and so the total number of such cycles is bounded w.h.p. Thus all the pairwise distances (in $\boldsymbol{G} - v$) between neighbours of v are at least $2t + 1$ w.h.p. (and in particular this proves the second part of the lemma). Therefore

$$\mathbb{E}_{\boldsymbol{G}}[X_{t+1}] = \frac{(d_+ + d_-)^d}{d! \exp(d_+ + d_-)} \binom{d}{c_1 \ldots c_{\tilde{d}}} \prod_{j=1}^{\tilde{d}} p_t(C_j) + \mathrm{o}(1).$$

By definition of $\boldsymbol{T}$, we obtain $\mathbb{E}[X_{t+1}] = p_{t+1} + \mathrm{o}(1)$. To apply Chebyshev's inequality, it remains to determine $\mathbb{E}[X_{t+1}^2]$. Let $\boldsymbol{v}, \boldsymbol{w} \in [n]$ be two randomly choosen vertices. Then w.h.p. $\boldsymbol{v}$ and $\boldsymbol{w}$ have distance at least $2t + 3$ in $\boldsymbol{G}$, conditioned on which $\partial^{t+1}(\boldsymbol{G}, \boldsymbol{v}, \boldsymbol{\sigma})$ and $\partial^{t+1}(\boldsymbol{G}, \boldsymbol{w}, \boldsymbol{\sigma})$ are independent. Therefore we obtain

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{v}, \boldsymbol{w}}\left(\partial^{t+1}(\boldsymbol{G}, \boldsymbol{v}, \boldsymbol{\sigma}) \cong \partial^{t+1}(T, r, \tau) \wedge \partial^{t+1}(\boldsymbol{G}, \boldsymbol{w}, \boldsymbol{\sigma}) \cong \partial^{t+1}(T, r, \tau)\right) \\
&\quad = \mathbb{P}_{\boldsymbol{v}}\left(\partial^{t+1}(\boldsymbol{G}, \boldsymbol{v}, \boldsymbol{\sigma}) \cong \partial^{t+1}(T, r, \tau)\right) \mathbb{P}_{\boldsymbol{w}}\left(\partial^{t+1}(\boldsymbol{G}, \boldsymbol{w}, \boldsymbol{\sigma}) \cong \partial^{t+1}(T, r, \tau)\right) + \mathrm{o}(1)
\end{aligned}$$

And finally

$$\begin{aligned}\mathbb{E}_{\boldsymbol{G}}[X_{t+1}^2] =& \mathbb{E}_{\boldsymbol{G}}\left[\mathbb{P}_{\boldsymbol{v}}\left(\partial^{t+1}(\boldsymbol{G},\boldsymbol{v},\boldsymbol{\sigma}) \cong \partial^{t+1}(T,r,\tau)\right)\mathbb{P}_{\boldsymbol{w}}\left(\partial^{t+1}(\boldsymbol{G},\boldsymbol{w},\boldsymbol{\sigma}) \cong \partial^{t+1}(T,r,\tau)\right)\right] \\ &+ \frac{1}{n}\mathbb{E}_{\boldsymbol{G}}[X_{t+1}] + \mathrm{o}(1) \\ =& \mathbb{E}_{\boldsymbol{G}}[X_{t+1}]^2 + \mathrm{o}(1).\end{aligned}$$

The first assertion follows from Chebyshev's inequality. ◀

2.4 The fixed point

Let (T, r, τ) be a rooted tree together with a map $\tau : V(T) \to \{\pm 1\}$. Then for any pair v, w of adjacent vertices we have the WP messages $\mu_{v \to w}(t|T, \tau)$, $t \geq 0$, as defined in (2.2). Since we are going to be particularly interested in the messages directed towards the root, we introduce the following notation. Given the root r, any vertex $v \neq r$ of T has a unique parent vertex w (the neighbour of v on the unique path from v to r). Initially, let

$$\mu_{v\uparrow}(0|T, r, \tau) = \tau(v) \tag{2.4}$$

and define

$$\mu_{v\uparrow}(t|T, r, \tau) = \mu_{v\to w}(t|T, \tau) \tag{2.5}$$

for $t > 0$. In addition, set $\mu_{r\uparrow}(0|T, r, \tau) = \tau(r)$ and let

$$\mu_{r\uparrow}(t+1|T, r, \tau) = \psi\left(\sum_{v \in \partial_{\boldsymbol{T}} r} \mu_{v\uparrow}(t|T, r, \tau)\right) \qquad (t \geq 0) \tag{2.6}$$

be the message that r would send to its parent if there was one.

For $p = (p(-1), p(0), p(1)) \in \mathcal{P}(\{-1, 0, 1\})$ we let $\bar{p} = (p(1), p(0), p(-1))$. Remembering the map

$$\mathcal{T} = \mathcal{T}_{d_+, d_-} : \mathcal{P}(\{-1, 0, 1\}) \to \mathcal{P}(\{-1, 0, 1\})$$

from Section 1.2 and writing $\mathcal{T}^t$ for its t-fold iteration, we observe the following.

▶ **Lemma 2.4.** *Let $p_t = \mathcal{T}^t(0, 0, 1)$.*

1. *Given that $\boldsymbol{\tau}(r_{\boldsymbol{T}}) = +1$, the message $\mu_{r_{\boldsymbol{T}}\uparrow}(t|\boldsymbol{T}, r_{\boldsymbol{T}}, \boldsymbol{\tau})$ has distribution p_t.*
2. *Given that $\boldsymbol{\tau}(r_{\boldsymbol{T}}) = -1$, the message $\mu_{r_{\boldsymbol{T}}\uparrow}(t|\boldsymbol{T}, r_{\boldsymbol{T}}, \boldsymbol{\tau})$ has distribution $\bar{p}_t$.*

Proof. The proof is by induction on t. In the case $t = 0$ the assertion holds because $\mu_{r_{\boldsymbol{T}}\uparrow}(0|\boldsymbol{T}, r_{\boldsymbol{T}}, \boldsymbol{\tau}) = \boldsymbol{\tau}(r_{\boldsymbol{T}})$. Now, assume that the assertion holds for t. To prove it for $t+1$, let $C_\pm$ be the set of all children v of $r_{\boldsymbol{T}}$ with $\boldsymbol{\tau}(r_{\boldsymbol{T}})\boldsymbol{\tau}(v) = \pm 1$. By construction, $|C_\pm|$ has distribution $\mathrm{Po}(d_\pm)$. Furthermore, let $(\boldsymbol{T}_v, v, \tau_v)$ signify the subtree pending on a child v of $r_{\boldsymbol{T}}$. Because $\boldsymbol{T}$ is a Galton-Watson tree, the random subtrees $\boldsymbol{T}_v$ are mutually independent. Moreover, each $\boldsymbol{T}_v$ is distributed as a Galton-Watson tree with offspring matrix (2.3) and a root vertex of type $\pm\boldsymbol{\tau}(r_{\boldsymbol{T}})$ for each $v \in C_\pm$. Therefore, by induction the message $\mu_{v\uparrow}(t|\boldsymbol{T}_v, v, \boldsymbol{\tau}_v)$ has distribution p_t if $\boldsymbol{\tau}(v) = 1$ resp. $\bar{p}_t$ if $\boldsymbol{\tau}(v) = -1$. As a consequence,

$$\mu_{r_{\boldsymbol{T}}\uparrow}(t+1|\boldsymbol{T}, r_{\boldsymbol{T}}, \boldsymbol{\tau}) = \psi\left(\sum_{v \in C_+} \mu_{v\uparrow}(t|\boldsymbol{T}_v, v, \boldsymbol{\tau}_v) + \sum_{v \in C_-} \mu_{v\uparrow}(t|\boldsymbol{T}_v, v, \boldsymbol{\tau}_v)\right)$$

has distribution p_{t+1} if $\boldsymbol{\tau}(r_{\boldsymbol{T}}) = 1$ and $\bar{p}_{t+1}$ otherwise. ◀

Lemma 2.4 shows that the operator $\mathcal{T}$ mimics WP on the Galton-Watson tree $(\boldsymbol{T}, r_{\boldsymbol{T}}, \boldsymbol{\tau})$. Hence, to understand the behaviour of WP after a large enough number of iterations we need to investigate the fixed point to which $\mathcal{T}^t(0,0,1)$ converges as $t \to \infty$. In Section 4 we will establish the following.

▶ **Proposition 2.5.** *The operator $\mathcal{T}$ has a unique skewed fixed point p^* and $\lim_{t\to\infty} \mathcal{T}^t(0,0,1) = p^*$.*

Proof of Theorem 1.1. Consider the random variables

$$X_n := \frac{1}{n}\mathrm{bis}(\boldsymbol{G}), \qquad Y_n^{(t)} := \frac{1}{2}\frac{1}{n}\sum_{v\in[n]}\sum_{w\in\partial_{\boldsymbol{G}} v} \mathbf{1}\left\{\mu_{w\to v}(t|\boldsymbol{G},\boldsymbol{\sigma}) = -\tilde{\psi}\left(\mu_v(t|\boldsymbol{G},\boldsymbol{\sigma})\right)\right\}.$$

Then Lemma 2.1 and Proposition 2.2 imply that for any $\varepsilon > 0$,

$$\lim_{t\to\infty}\lim_{n\to\infty} \mathbb{P}\left[|X_n - Y_n^{(t)}| > \varepsilon\right] = 0. \tag{2.7}$$

By Definition (2.2), $\mu_{w\to v}(t|\boldsymbol{G},\boldsymbol{\sigma})$ and $\mu_v(t|\boldsymbol{G},\boldsymbol{\sigma})$ are determined by $\partial^t_{\boldsymbol{G}} v$ and the initialisation $\mu_{u\to w}(0|\boldsymbol{G},\boldsymbol{\sigma})$ for all $u, w \in \partial^t_{\boldsymbol{G}} v$, $\{u,w\} \in E(\boldsymbol{G})$. Since (2.5) and (2.6) match the recursive definition (2.2) of $\mu_{w\to v}(t|\boldsymbol{G},\boldsymbol{\sigma})$ and $\mu_v(t|\boldsymbol{G},\boldsymbol{\sigma})$, Lemma 2.3 implies that for any fixed $t > 0$ (as n tends to infinity),

$$Y_n^{(t)} \overset{n\to\infty}{\to} x^{(t)} := \frac{1}{2}\mathbb{E}\left[\sum_{w\in\partial_{\boldsymbol{T}} r_{\boldsymbol{T}}} \mathbf{1}\{\mu_{w\uparrow}(t|\boldsymbol{T}, r_{\boldsymbol{T}}, \boldsymbol{\tau}) = -\psi(\mu_{r_{\boldsymbol{T}}}(t|\boldsymbol{T}, r_{\boldsymbol{T}}, \boldsymbol{\tau}))\}\right] \quad \text{in probability.} \tag{2.8}$$

Now let p^* denote the unique skewed fixed point of $\mathcal{T}$ guaranteed by Proposition 2.5. Since each child of $r_{\boldsymbol{T}}$ can be considered a root of an independent instance of $\boldsymbol{T}$ to which we can apply Lemma 2.4, we obtain that given $(\boldsymbol{\tau}(w))_{w\in\partial r_{\boldsymbol{T}}}$ the sequence $(\mu_{w\uparrow}(t|\boldsymbol{T}, r_{\boldsymbol{T}}, \boldsymbol{\tau}))_{w\in\partial r_{\boldsymbol{T}}}$ converges to a sequence of independent random variables $(\eta_w)_{w\in\partial r_{\boldsymbol{T}}}$ with distribution p^* (if $\boldsymbol{\tau}(w) = 1$) and $\bar{p}^*$ (if $\boldsymbol{\tau}(w) = -1$). By definition $\mu_{r_{\boldsymbol{T}}}(t|\boldsymbol{T}, r_{\boldsymbol{T}}, \boldsymbol{\tau})$ converges to $\sum_{w\in\partial r_{\boldsymbol{T}}, \boldsymbol{\tau}(w)=1} \eta_w + \sum_{w\in\partial r_{\boldsymbol{T}}, \boldsymbol{\tau}(w)=-1} \eta_w$. Considering the offspring distributions of $r_{\boldsymbol{T}}$ in both cases, i.e. $\boldsymbol{\tau}(r_{\boldsymbol{T}}) = \pm 1$, we obtain from $\varphi_{d_+,d_-}(p) = \varphi_{d_+,d_-}(\bar{p})$ for all $p \in \mathcal{P}(\{-1,0,1\})$ that

$$\lim_{t\to\infty} x^{(t)} = \varphi_{d_+,d_-}(p^*). \tag{2.9}$$

Finally, combining (2.7)–(2.9) completes the proof. ◀

3 Proof of Proposition 2.2

▶ **Lemma 3.1.** *If $v \in \mathcal{C}$ and $w \in \partial_{\boldsymbol{G}} v$, then $\mu_{v\to w}(t|\boldsymbol{G},\boldsymbol{\sigma}) = \boldsymbol{\sigma}(v) = \mu_{v\to w}(t|\boldsymbol{G},\boldsymbol{\sigma}_{\mathcal{C}})$ for all $t \geq 0$.*

Proof. We proceed by induction on t. For $t = 0$ the assertion is immediate from the initialisation of the messages. To go from t to $t+1$, consider $v \in \mathcal{C}$ and $w \in \partial_{\boldsymbol{G}} v$. We may assume without loss of generality that $\boldsymbol{\sigma}(v) = 1$. By the definition of the WP message,

$$\mu_{v\to w}(t+1|\boldsymbol{G},\boldsymbol{\sigma}) = \psi\left(\sum_{u\in\partial_{\boldsymbol{G}} v\setminus\{w\}} \mu_{u\to v}(t|\boldsymbol{G},\boldsymbol{\sigma})\right) = \psi\left(S_+ + S_- + S_0\right) \tag{3.1}$$

where

$$S_+ := \sum_{u \in \mathcal{C} \cap \boldsymbol{\sigma}^{-1}(+1) \cap \partial_{\boldsymbol{G}} v \setminus \{w\}} \mu_{u \to v}(t|\boldsymbol{G}, \boldsymbol{\sigma}),$$

$$S_- := \sum_{u \in \mathcal{C} \cap \boldsymbol{\sigma}^{-1}(-1) \cap \partial_{\boldsymbol{G}} v \setminus \{w\}} \mu_{u \to v}(t|\boldsymbol{G}, \boldsymbol{\sigma}),$$

$$S_0 := \sum_{u \in \partial_{\boldsymbol{G}} v \setminus (\mathcal{C} \cup \{w\})} \mu_{u \to v}(t|\boldsymbol{G}, \boldsymbol{\sigma}).$$

Now, (2.1) ensures that

$$S_+ \geq d_+ - \frac{c}{4}\sqrt{d_+ \ln d_+}, \quad S_- \geq -d_- - \frac{c}{4}\sqrt{d_+ \ln d_+}, \quad S_0 \leq 100 \leq \frac{c}{4}\sqrt{d_+ \ln d_+}, \tag{3.2}$$

provided that the constant $c > 0$ is chosen large enough. Combining (3.1) and (3.2), we see that $S_+ + S_- + S_0 \geq 1$ and thus $\mu_{v \to w}(t+1|\boldsymbol{G}, \boldsymbol{\sigma}) = 1$. The exact same argument works for $\mu_{v \to w}(t+1|\boldsymbol{G}, \boldsymbol{\sigma}_{\mathcal{C}}) = 1$. ◀

Let $\boldsymbol{G}_v$ denote the subgraph of $\boldsymbol{G}$ induced on $\mathcal{C}_v$. To prove Proposition 2.2, fix $s > 0$ large enough. Let $\mathcal{S} = \mathcal{S}(s)$ be the set of all vertices such that either $|\mathcal{C}_v| > \sqrt{s}$ or $\boldsymbol{G}_v$ is cyclic. Then Lemma 2.1 (with slightly smaller ε) and Lemma 2.3 imply that $|\mathcal{S}| \leq \varepsilon n$ w.h.p. For the rest of this section, let $v \notin \mathcal{S}$ be fixed.

For $w \in \mathcal{C}_v \setminus \{v\}$ we let $w_{\uparrow v}$ be the neighbour of w on the path from w to v. We define $\boldsymbol{G}_{w \to v}$ as the component of w in the graph obtained from $\boldsymbol{G}_v$ by removing the edge $\{w, w_{\uparrow v}\}$. The vertex set of $\boldsymbol{G}_{w \to v}$ will be denoted by $\mathcal{C}_{w \to v}$. Further, $h_{w \to v}$ is the maximum distance between w and any other vertex in $\boldsymbol{G}_{w \to v}$. Additionally, h_v is the maximum distance between v and any other vertex in $\boldsymbol{G}_v$. Finally, let $\boldsymbol{\sigma}_v : \mathcal{C}_v \to \{\pm 1\}$, $w \mapsto \boldsymbol{\sigma}(w)$ and let $\boldsymbol{\sigma}_{\mathcal{C},v} : \mathcal{C}_v \cap \mathcal{C} \to \{\pm 1\}$, $w \mapsto \boldsymbol{\sigma}_{\mathcal{C}}(w)$.

▶ **Lemma 3.2.**
1. *For any $w \in \mathcal{C}_v \setminus \{v\}$ and any $t > h_{w \to v}$ we have* $\mu_{w \to w_{\uparrow v}}(t|\boldsymbol{G}, \boldsymbol{\sigma}) = \mu_{w \to w_{\uparrow v}}(h_{w \to v} + 1|\boldsymbol{G}, \boldsymbol{\sigma}) = \mu_{w \to w_{\uparrow v}}(t|\boldsymbol{G}, \boldsymbol{\sigma}_{\mathcal{C}})$.
2. *For any $t \geq h_v$ we have* $\mu_v(t|\boldsymbol{G}, \boldsymbol{\sigma}) = \mu_v(h_v + 1|\boldsymbol{G}, \boldsymbol{\sigma}) = \mu_v(t|\boldsymbol{G}, \boldsymbol{\sigma}_{\mathcal{C}})$.

Proof. The proof of (1) proceeds by induction on $h_{w \to v}$. The construction **C1–C2** of $\mathcal{C}_v$ ensures that any $w \in \mathcal{C}_v$ with $h_{w \to v} = 0$ either belongs to $\mathcal{C}$ or has no neighbour besides $w_{\uparrow v}$. Hence for the first case the assumption follows from Lemma 3.1. If $\partial_{\boldsymbol{G}} w \setminus \{w_{\uparrow v}\} = \emptyset$ we obtain that $\mu_{w \to w_{\uparrow v}}(t|\boldsymbol{G}, \boldsymbol{\sigma}) = \mu_{w \to w_{\uparrow v}}(t|\boldsymbol{G}, \boldsymbol{\sigma}_{\mathcal{C}}) = 0$ for all $t \geq 1$ by the definition of the WP messages. Now, assume that $h_{w \to v} > 0$ and let $t > h_{w \to v}$. Then all neighbours $u \neq w_{\uparrow v}$ of w in $\boldsymbol{G}_{w \to v}$ satisfy $h_{u \to v} < h_{w \to v}$. Thus, by induction

$$\begin{aligned}
\mu_{w \to w_{\uparrow v}}(t|\boldsymbol{G}, \boldsymbol{\sigma}) &= \psi\left(\sum_{u \in \partial_{\boldsymbol{G}} w \setminus \{w_{\uparrow v}\}} \mu_{u \to w}(t-1|\boldsymbol{G}, \boldsymbol{\sigma})\right) \\
&= \psi\left(\sum_{u \in \partial_{\boldsymbol{G}} w \setminus \{w_{\uparrow v}\}} \mu_{u \to w}(h_{u \to v}+1|\boldsymbol{G}, \boldsymbol{\sigma})\right) = \mu_{w \to w_{\uparrow v}}(h_{w \to v}+1|\boldsymbol{G}, \boldsymbol{\sigma}).
\end{aligned}$$

An analogous argument applies to $\mu_{w \to w_{\uparrow v}}(t|\boldsymbol{G}, \boldsymbol{\sigma}_{\mathcal{C}})$. The proof of (2) is similar. ◀

For each vertex $w \in \mathcal{C}_v$, $w \neq v$, let $\mu^*_{w \to v} = \mu_{w \to w_{\uparrow v}}(s|\boldsymbol{G}, \boldsymbol{\sigma})$. Further, let $\mu^*_w = \mu_w(s|\boldsymbol{G}, \boldsymbol{\sigma})$. In addition, for $z \in \{\pm 1\}$ let

$$\boldsymbol{\sigma}^z_{w \to v} : \mathcal{C}_{w \to v} \cap (\{w\} \cup \mathcal{C}) \to \{\pm 1\}, \qquad u \mapsto \begin{cases} z & \text{if } u = w, \\ \boldsymbol{\sigma}(u) & \text{otherwise.} \end{cases}$$

In words, $\boldsymbol{\sigma}^z_{w\to v}$ freezes w to z and all other $u \in \mathcal{C}_{w\to v}$ that belong to the core to $\boldsymbol{\sigma}(u)$. Analogously, let

$$\boldsymbol{\sigma}^z_v : \mathcal{C}_v \cap (\{v\} \cup \mathcal{C}) \to \{\pm 1\}, \qquad u \mapsto \begin{cases} z & \text{if } u = v, \\ \boldsymbol{\sigma}(u) & \text{otherwise.} \end{cases}$$

▶ **Lemma 3.3.** *Suppose that $u \in \mathcal{C}_v \setminus \{v\}$, such that $h_{u\to v} \geq 1$.*

1. *If $z = \mu^*_{u\to v} \in \{-1, 1\}$, then*

$$\mathrm{cut}(\boldsymbol{G}_{u\to v}, \boldsymbol{\sigma}^z_{u\to v}) < \mathrm{cut}(\boldsymbol{G}_{u\to v}, \boldsymbol{\sigma}^{-z}_{u\to v}). \tag{3.3}$$

Similarly, if $z = \psi(\mu^*_v) \in \{-1, 1\}$, then

$$\mathrm{cut}(\boldsymbol{G}_v, \boldsymbol{\sigma}^z_v) < \mathrm{cut}(\boldsymbol{G}_v, \boldsymbol{\sigma}^{-z}_v). \tag{3.4}$$

2. *If $\mu^*_{u\to v} = 0$, then*

$$\mathrm{cut}(\boldsymbol{G}_{u\to v}, \boldsymbol{\sigma}^{+1}_{u\to v}) = \mathrm{cut}(\boldsymbol{G}_{u\to v}, \boldsymbol{\sigma}^{-1}_{u\to v}). \tag{3.5}$$

Similarly, if $\mu^*_v = 0$, then

$$\mathrm{cut}(\boldsymbol{G}_v, \boldsymbol{\sigma}^{+1}_v) = \mathrm{cut}(\boldsymbol{G}_v, \boldsymbol{\sigma}^{-1}_v). \tag{3.6}$$

Proof. We prove (3.3) and (3.5) by induction on $h_{u\to v}$. If $h_{u\to v} = 1$ then we have that all neighbours $w \in \partial_{\mathcal{C}_{u\to v}} u$ of u with $\mu^*_{u\to v} \neq 0$ are in $\mathcal{C}$, i.e. fixed under $\boldsymbol{\sigma}^z_{u\to v}$. Since $\mathcal{C}_{u\to v} = \partial_{\boldsymbol{G}} u \setminus \{u_{\uparrow v}\} \cup \{u\}$, we obtain

$$\mathrm{cut}(\mathcal{C}_{u\to v}, \boldsymbol{\sigma}^{-z}_{u\to v}) - \mathrm{cut}(\mathcal{C}_{u\to v}, \boldsymbol{\sigma}^{z}_{u\to v}) = \left| \sum_{w \in \partial_{\boldsymbol{G}} u \setminus \{u_{\uparrow v}\}} \mu^*_{w\to v} \right| \tag{3.7}$$

by definition of z. By the induction hypothesis and because $\boldsymbol{G}_{u\to v}$ is a tree (as $v \notin \mathcal{S}$) we have that (3.7) holds for $h_{u\to v} > 1$ as well. A similar argument yields (3.4) and (3.6). ◀

Now, let $\mathcal{U}_v$ be the set of all $w \in \mathcal{C}_v$ such that $\mu^*_{w\to v} \neq 0$. Furthermore, let

$$\boldsymbol{\sigma}_{\uparrow v} : \mathcal{U}_v \cup \{v\} \to \{-1, +1\}, \qquad w \mapsto \begin{cases} \tilde{\psi}(\mu^*_v) & \text{if } w = v, \\ \mu^*_{w\to v} & \text{otherwise.} \end{cases}$$

Thus, $\boldsymbol{\sigma}_{\uparrow v}$ sets all $w \in \mathcal{C}_v \cap \mathcal{C} \setminus \{v\}$ to their planted sign and all $w \in \mathcal{U}_v \setminus \mathcal{C}$ to $\mu^*_{w\to v}$. Moreover, $\boldsymbol{\sigma}_{\uparrow v}$ sets v to $\psi(\mu^*_v)$ if $\psi(\mu^*_v) \neq 0$ and to 1 if there is a tie.

▶ **Corollary 3.4.** *We have* $\mathrm{cut}(\boldsymbol{G}_v, \boldsymbol{\sigma}_{\mathcal{C}}) = \mathrm{cut}(\boldsymbol{G}_v, \boldsymbol{\sigma}_{\uparrow v})$.

Proof. This is immediate from Lemma 3.3. ◀

Hence, in order to determine an optimal cut of $\boldsymbol{G}_v$ we merely need to figure out the assignment of the vertices in $\mathcal{C}_v \setminus (\{v\} \cup \mathcal{U}_v)$. Suppose that $\boldsymbol{\sigma}^*_{\uparrow v} : \mathcal{C}_v \to \{\pm 1\}$ is an optimal extension of $\boldsymbol{\sigma}_{\uparrow v}$ to a cut of $\boldsymbol{G}_v$, i.e.,

$$\mathrm{cut}(\boldsymbol{G}_v, \boldsymbol{\sigma}_{\uparrow v}) = \sum_{\{u,w\} \in E(\boldsymbol{G}_v)} \frac{1}{2}(1 - \boldsymbol{\sigma}^*_{v\uparrow}(u)\boldsymbol{\sigma}^*_{v\uparrow}(w)).$$

▶ **Corollary 3.5.** *It holds that* $\sum_{w \in \partial_{\boldsymbol{G}} v} \frac{1}{2}(1 - \boldsymbol{\sigma}^*_{v\uparrow}(v)\boldsymbol{\sigma}^*_{v\uparrow}(w)) = \sum_{w \in \partial_{\boldsymbol{G}} v} \mathbf{1}\left\{\mu^*_{w\to v} = -\tilde{\psi}\left(\mu_v\right)\right\}$.

Proof. Part (2) of Lemma 3.3 implies that $\boldsymbol{\sigma}^*_{v\uparrow}(v)\boldsymbol{\sigma}^*_{v\uparrow}(w) = 1$ for all $w \in \partial_{\boldsymbol{G}} v$ such that $\mu^*_{w\to v} = 0$. ◀

Proof of Proposition 2.2. Given $\varepsilon > 0$ choose $\delta = \delta(\varepsilon, d_+, d_-)$ sufficiently small and $s = s(\varepsilon, \delta, d_+, d_-) > 0$ sufficiently large. In particular, pick s large enough so that

$$\mathbb{P}\left(|\mathcal{S}| \geq \delta n\right) < \varepsilon. \tag{3.8}$$

Provided that δ is suitable small, the Chernoff bound implies that for large n

$$\mathbb{P}\left(\frac{1}{2}\sum_{v\in\mathcal{S}} |\partial_{\boldsymbol{G}} v| \geq \varepsilon n \,\middle|\, |\mathcal{S}| < \delta n\right) < \varepsilon. \tag{3.9}$$

Now, suppose that $\boldsymbol{\sigma}^*_{\mathcal{C}}$ is an optimal extension of $\boldsymbol{\sigma}_{\mathcal{C}}$ to a cut of $\boldsymbol{G}$ and let $v \notin \mathcal{S}$. Then using the definition of $\mathcal{C}_v$, Corollary 3.4 implies that

$$\sum_{w\in\partial_{\boldsymbol{G}} v} (1 - \boldsymbol{\sigma}^*_{\mathcal{C}}(v)\boldsymbol{\sigma}^*_{\mathcal{C}}(w)) = \sum_{w\in\partial_{\boldsymbol{G}} v} (1 - \boldsymbol{\sigma}^*_{v\uparrow}(v)\boldsymbol{\sigma}^*_{v\uparrow}(w)).$$

Therefore, we obtain

$$\mathbb{P}\left(\left|\mathrm{cut}(\boldsymbol{G}, \boldsymbol{\sigma}_{\mathcal{C}}) - \frac{1}{2}\sum_{v\notin\mathcal{S}}\sum_{w\in\partial_{\boldsymbol{G}} v} (1 - \boldsymbol{\sigma}^*_{v\uparrow}(v)\boldsymbol{\sigma}^*_{v\uparrow}(w))\right| \geq \varepsilon n\right) \leq \mathbb{P}\left(\frac{1}{2}\sum_{v\in\mathcal{S}} |\partial_{\boldsymbol{G}} v| \geq \varepsilon n\right) \leq 2\varepsilon.$$

The assertion follows from Lemma 3.2 for $t \geq s$. ◀

4 Proof of Proposition 2.5

We continue to denote the set of probability measures on $\mathcal{X} \subset \mathbb{R}^k$ by $\mathcal{P}(\mathcal{X})$. For a $\mathcal{X}$-valued random variable X we denote by $\mathcal{L}(X) \in \mathcal{P}(\mathcal{X})$ the distribution of X. Furthermore, if $p, q \in \mathcal{P}(\mathcal{X})$, then $\mathcal{P}_{p,q}(\mathcal{X})$ denotes the set of all probability measures μ on $\mathcal{X} \times \mathcal{X}$ such that the marginal distribution of the first (resp. second) component coincides with p (resp. q). The space $\mathcal{P}(\{-1, 0, 1\})$ is complete with respect to (any and in particular) the L_1-Wasserstein metric, defined by

$$\ell_1(p, q) = \inf\left\{\mathbb{E}|X - Y| : X, Y \text{ are random variables with } \mathcal{L}(X, Y) \in \mathcal{P}_{p,q}(\{-1, 0, 1\})\right\}.$$

In words, the infimum of $\mathbb{E}|X - Y|$ is over all couplings (X, Y) of the distributions p, q. Such a coupling (X, Y) is *optimal* if $\ell_1(p, q) = \mathbb{E}|X - Y|$. Finally, let $\mathcal{P}^*(\{-1, 0, 1\})$ be the set of all skewed probability measures on $\{-1, 0, 1\}$. Being a closed subset of $\mathcal{P}(\{-1, 0, 1\})$, $\mathcal{P}^*(\{-1, 0, 1\})$ is complete with respect to $\ell_1(\,\cdot\,, \cdot\,)$.

As in the definition (1.2)-(1.3) of the operator $\mathcal{T} = \mathcal{T}_{d_+,d_-}$ for $p \in \mathcal{P}(\{-1, 0, 1\})$ we let $(\eta_{p,i})_{i\geq 1}$ be a family of independent random variables with distribution p. Further, let $\gamma_\pm = \mathrm{Po}(d_\pm)$ be independent of each other and of the $(\eta_{p,i})_{i\geq 1}$. We introduce the shorthands

$$Z_p = Z_{p,d_+,d_-}, \qquad Z_{p,+} = \sum_{i=1}^{\gamma_+} \eta_{p,i}, \qquad Z_{p,-} = \sum_{i=\gamma_++1}^{\gamma_++\gamma_-} \eta_{p,i} \quad \text{so that} \quad Z_p = Z_{p,+} - Z_{p,-}.$$

Also set $\lambda = c\sqrt{d_+ \ln d_+}$ and recall that $c > 0$ is a constant that we assume to be sufficiently large.

▶ **Lemma 4.1.** *The operator $\mathcal{T}$ maps $\mathcal{P}^*(\{-1, 0, 1\})$ into itself.*

Proof. Suppose that $p \in \mathcal{P}(\{-1,0,1\})$ is skewed. Then

$$\mathbb{P}\left(Z_p < 1\right) \leq \mathbb{P}\left(Z_{p,+} \leq d_+ - \frac{\lambda-1}{2}\right) + \mathbb{P}\left(Z_{p,-} \geq d_- + \frac{\lambda-1}{2}\right). \tag{4.1}$$

Since $|\eta_{p,i}| \leq 1$ for all i, we can bound the second summand from above by invoking the Chernoff bound on a binomial approximation of the Poisson distribution to obtain

$$\mathbb{P}\left(\gamma_- \geq d_- + \frac{c}{2}\sqrt{d_+ \ln d_+} - \frac{1}{2}\right) < \frac{1}{3} d_+^{-10}, \tag{4.2}$$

provided c is large enough. To bound the other summand from above we use that $(\eta_{p,i})_{i \geq 1}$ is a sequence of independent *skewed* random variables, whence by the Chernoff bound

$$\begin{aligned}
&\mathbb{P}\left(Z_{p,+} \leq d_+ - \frac{\lambda-1}{2}\right) \\
&\leq \mathbb{P}\left(|\gamma_+ - d_+| > \lambda/8\right) + \mathbb{P}\left(Z_{p,-} \leq d_+ - \frac{\lambda-1}{2} \middle| \gamma_+ \geq d_+ - \lambda/8\right) \\
&\leq \frac{1}{3} d_+^{-10} + \mathbb{P}\left[\mathrm{Bin}(d_+ - \lambda/8, 1 - d_+^{-10}) \leq d_+ - \lambda/7\right] < \frac{2}{3} d_+^{-10},
\end{aligned} \tag{4.3}$$

provided that c is sufficiently big. Combining (4.1)–(4.3) completes the proof. ◀

▶ **Lemma 4.2.** *The operator $\mathcal{T}$ is ℓ_1-contracting on $\mathcal{P}^*(\{-1,0,1\})$.*

Proof. Let $p, q \in \mathcal{P}^*(\{-1,0,1\})$. We aim to show that $\ell_1(\mathcal{T}(p), \mathcal{T}(q)) \leq \frac{1}{2}\ell_1(p,q)$. To this end, we let $(\eta_{p,i}, \eta_{q,i})_{i \geq 1}$ be a family of random variables with distribution p resp. q such that $(\eta_{p,i})_{i \geq 1}$ are independent and $(\eta_{q,i})_{i \geq 1}$ are independent but such that the pair $(\eta_{p,i}, \eta_{q,i})$ is an optimal coupling for every i. Then by the definition of $\ell_1(\,\cdot\,,\,\cdot\,)$,

$$\ell_1(\mathcal{T}(p), \mathcal{T}(q)) \leq \mathbb{E}\left|\psi(Z_p) - \psi(Z_q)\right|. \tag{4.4}$$

To estimate the r.h.s., let $\tilde{\eta}_{p,i} = \mathbf{1}\{\eta_{p,i} = 1\}$, $\tilde{\eta}_{q,i} = \mathbf{1}\{\eta_{q,i} = 1\}$. Further, let $\mathfrak{F}_i$ be the σ-algebra generated by $\tilde{\eta}_{p,i}, \tilde{\eta}_{q,i}$ and let $\mathfrak{F}$ be the σ-algebra generated by γ_+, γ_- and the random variables $(\tilde{\eta}_{p,i}, \tilde{\eta}_{q,i})_{i \geq 1}$. Additionally, let $\gamma = \gamma_+ + \gamma_-$ and consider the three events

$$\mathfrak{A}_1 = \left\{\sum_{i=1}^{\gamma} \tilde{\eta}_{p,i}\tilde{\eta}_{q,i} \geq \gamma - 10\right\}, \qquad \mathfrak{A}_2 = \left\{\gamma \geq 2d_+\right\}, \qquad \mathfrak{A}_3 = \left\{\gamma_+ - \gamma_- \leq 20\right\}.$$

We are going to bound $|\psi(Z_p) - \psi(Z_q)|$ on $\mathfrak{A}_1 \setminus (\mathfrak{A}_2 \cup \mathfrak{A}_3)$, $\overline{\mathfrak{A}_1 \cup \mathfrak{A}_2 \cup \mathfrak{A}_3}$, $\mathfrak{A}_2$ and $\mathfrak{A}_3 \setminus \mathfrak{A}_2$ separately. The bound on the first event is immediate: if $\mathfrak{A}_1 \setminus (\mathfrak{A}_2 \cup \mathfrak{A}_3)$ occurs, then $\psi(Z_p) = \psi(Z_q) = 1$ with certainty. Hence,

$$\mathbb{E}\left[|\psi(Z_p) - \psi(Z_q)| \cdot \mathbf{1}_{\mathfrak{A}_1 \setminus (\mathfrak{A}_2 \cup \mathfrak{A}_3)}\right] = 0. \tag{4.5}$$

Let us turn to the second event $\overline{\mathfrak{A}_1 \cup \mathfrak{A}_2 \cup \mathfrak{A}_3}$. Because the pairs $(\eta_{p,i}, \eta_{q,i})_{i \geq 1}$ are mutually independent, we find

$$\mathbb{E}\left[|\eta_{p,i} - \eta_{q,i}| \,|\, \mathfrak{F}\right] = \mathbb{E}\left[|\eta_{p,i} - \eta_{q,i}| \,|\, \mathfrak{F}_i\right] \qquad \text{for all } i \geq 1. \tag{4.6}$$

Clearly, if $\tilde{\eta}_{p,i}\tilde{\eta}_{q,i} = 1$, then $\eta_{p,i} - \eta_{q,i} = 0$. Consequently,

$$\mathbb{E}\left[|\eta_{p,i} - \eta_{q,i}| \,|\, \mathfrak{F}_i\right] \leq \frac{\mathbb{E}|\eta_{p,i} - \eta_{q,i}|}{\mathbb{P}[\tilde{\eta}_{p,i}\tilde{\eta}_{q,i} = 0]} = \frac{\mathbb{E}|\eta_{p,1} - \eta_{q,1}|}{\mathbb{P}[\tilde{\eta}_{p,1}\tilde{\eta}_{q,1} = 0]}. \tag{4.7}$$

Since the events $\mathfrak{A}_1, \mathfrak{A}_2, \mathfrak{A}_3$ are $\mathfrak{F}$-measurable and because $\bar{\mathfrak{A}}_2$ ensures that $\gamma < 2d_+$, (4.6) and (4.7) yield

$$\mathbb{E}[|\psi(Z_p) - \psi(Z_q)|\,|\mathfrak{F}]\mathbf{1}_{\overline{\mathfrak{A}_1 \cup \mathfrak{A}_2 \cup \mathfrak{A}_3}} \leq \frac{2d_+ \mathbb{E}|\eta_{p,1} - \eta_{q,1}|}{\mathbb{P}[\tilde{\eta}_{p,1}\tilde{\eta}_{q,1} = 0]} \cdot \mathbf{1}_{\overline{\mathfrak{A}_1 \cup \mathfrak{A}_2 \cup \mathfrak{A}_3}}. \tag{4.8}$$

Further, because the pairs $(\eta_{p,i}, \eta_{q,i})_{i \geq 1}$ are independent and because p, q are skewed,

$$\mathbb{P}\left(\overline{\mathfrak{A}_1 \cup \mathfrak{A}_2 \cup \mathfrak{A}_3}\right) \leq \mathbb{P}\left(\gamma \leq 2d_+, \sum_{i=1}^{\gamma} \tilde{\eta}_{p,i}\tilde{\eta}_{q,i} \leq \gamma - 10\right) \leq \left(2d_+ \mathbb{P}\left(\tilde{\eta}_{p,1}\tilde{\eta}_{q,1} = 0\right)\right)^{10}. \tag{4.9}$$

Combining (4.8) and (4.9), we obtain

$$\mathbb{E}\left[\mathbb{E}\left[|\psi(Z_p) - \psi(Z_q)||\mathfrak{F}\right] \mathbf{1}_{\overline{\mathfrak{A}_1 \cup \mathfrak{A}_2 \cup \mathfrak{A}_3}}\right] \leq (2d_+)^{11}\, \mathbb{P}\left(\tilde{\eta}_{p,1}\tilde{\eta}_{q,1} = 0\right)^9 \mathbb{E}|\eta_{p,1} - \eta_{q,1}|. \tag{4.10}$$

Since p, q are skewed, we furthermore obtain $\mathbb{P}\left(\tilde{\eta}_{p,1}\tilde{\eta}_{q,1} = 0\right) \leq 2d_+^{-10}$. Therefore

$$\begin{aligned}\mathbb{E}\left[|\psi(Z_p) - \psi(Z_q)|\mathbf{1}_{\overline{\mathfrak{A}_1 \cup \mathfrak{A}_2 \cup \mathfrak{A}_3}}\right] &= \mathbb{E}\left[\mathbb{E}\left[|\psi(Z_p) - \psi(Z_q)||\mathfrak{F}\right] \mathbf{1}_{\overline{\mathfrak{A}_1 \cup \mathfrak{A}_2 \cup \mathfrak{A}_3}}\right] \\ &\leq 2^{20} d_+^{-79} \mathbb{E}|\eta_{p,1} - \eta_{q,1}|.\end{aligned}$$

With respect to $\mathfrak{A}_2$, the triangle inequality yields

$$\mathbb{E}[|\psi(Z_p) - \psi(Z_q)|\mathbf{1}_{\mathfrak{A}_2}] \leq 2\mathbb{E}|\eta_{p,1} - \eta_{q,1}| \cdot \mathbb{E}[\gamma \mathbf{1}_{\mathfrak{A}_2}]. \tag{4.11}$$

Further, since $\gamma = \mathrm{Po}(d_+ + d_-)$, the Chernoff bound entails that $\mathbb{E}[\gamma \mathbf{1}_{\mathfrak{A}_2}] \leq d_+^{-1}$ if the constant c is chosen large enough. Combining this estimate with (4.11), we get

$$\mathbb{E}[|\psi(Z_p) - \psi(Z_q)|\mathbf{1}_{\mathfrak{A}_2}] \leq 2d_+^{-1} \mathbb{E}|\eta_{p,1} - \eta_{q,1}|. \tag{4.12}$$

Finally, on $\mathfrak{A}_3 \setminus \mathfrak{A}_2$ we have

$$\mathbb{E}[|\psi(Z_p) - \psi(Z_q)|\mathbf{1}_{\mathfrak{A}_3 \setminus \mathfrak{A}_2}] \leq 4d_+ \mathbb{E}|\eta_{p,1} - \eta_{q,1}|\, \mathbb{P}\left[\gamma_+ - \gamma_- \leq 20\right]. \tag{4.13}$$

Since $\gamma_\pm = \mathrm{Po}(d_\pm)$ and $d_+ - d_- \geq \lambda$, the Chernoff bound yields $\mathbb{P}\left[\gamma_+ - \gamma_- \leq 20\right] \leq d_+^{-2}$, if c is large enough. Hence, (4.13) implies

$$\mathbb{E}[|\psi(Z_p) - \psi(Z_q)|\mathbf{1}_{\mathfrak{A}_3 \setminus \mathfrak{A}_2}] \leq 4d_+^{-1} \mathbb{E}|\eta_{p,1} - \eta_{q,1}|. \tag{4.14}$$

Finally, the assertion follows from (4.4), (4.5), (4.10), (4.12) and (4.14). ◀

Proof of Proposition 2.5. The assertion follows from Lemmas 4.1 and 4.2 and the Banach fixed point theorem. ◀

Acknowledgements. We would like to thank an anonymous referee for detailed comments on an earlier draft of the paper.

References

1 E. Abbe, A. Bandeira, and G. Hall. Exact recovery in the stochastic block model. *arXiv preprint arXiv:1405.3267*, 2014.

2 D. Aldous and J. Steele. The objective method: probabilistic combinatorial optimization and local weak convergence. In *Probability on discrete structures*, pages 1–72. Springer, 2004.

3 S. Arora, S. Rao, and U. Vazirani. Expander flows, geometric embeddings and graph partitioning. *Journal of the ACM (JACM)*, 56:5, 2009.

4 V. Bapst, A. Coja-Oghlan, S. Hetterich, F. Rassmann, and D. Vilenchik. The condensation phase transition in random graph coloring. *arXiv preprint arXiv:1404.5513*, 2014.

5 B. Bollobás and A. Scott. Max cut for random graphs with a planted partition. *Combinatorics, Probability and Computing*, 13:451–474, 2004.

6 R. Boppana. Eigenvalues and graph bisection: An average-case analysis. In *Proc. 28th Foundations of Computer Science*, pages 280–285. IEEE, 1987.

7 T. Bui, S. Chaudhuri, T. Leighton, and M. Sipser. Graph bisection algorithms with good average case behavior. *Combinatorica*, 7:171–191, 1987.

8 T. Carson and R. Impagliazzo. Hill-climbing finds random planted bisections. In *Proceedings of the twelfth annual ACM-SIAM symposium on Discrete algorithms*, pages 903–909. Society for Industrial and Applied Mathematics, 2001.

9 A. Coja-Oghlan. A spectral heuristic for bisecting random graphs. In *Random Structures and Algorithms*, pages 351–398, 2006.

10 A. Coja-Oghlan, O. Cooley, M. Kang, and K. Skubch. How does the core sit inside the mantle? *arXiv preprint arXiv:1503.09030*, 2015.

11 A. Condon and R. Karp. Algorithms for graph partitioning on the planted partition model. *Random Structures and Algorithms*, 18:116–140, 2001.

12 A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84:066106, 2011.

13 A. Dembo, A. Montanari, and S. Sen. Extremal cuts of sparse random graphs. *arXiv preprint arXiv:1503.03923*, 2015.

14 T. Dimitriou and R. Impagliazzo. Go with the winners for graph bisection. In *Proc. 9th SODA*, pages 510–520. ACM/SIAM, 1998.

15 M. Dyer and A. Frieze. The solution of some random np-hard problems in polynomial expected time. *Journal of Algorithms*, 10:451–489, 1989.

16 U. Feige and J. Kilian. Heuristics for semirandom graph problems. *Journal of Computer and System Sciences*, 63:639–671, 2001.

17 U. Feige and R. Krauthgamer. A polylogarithmic approximation of the minimum bisection. *SIAM Journal on Computing*, 31:1090–1118, 2002.

18 M. Garey, D. Johnson, and L. Stockmeyer. Some simplified np-complete graph problems. *Theoretical computer science*, 1:237–267, 1976.

19 M. Goemans and D. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42:1115–1145, 1995.

20 P. Holland, K. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5:109–137, 1983.

21 S. Janson, T. Łuczak, and A. Ruciński. *Random graphs.* John Wiley & Sons, 2000.

22 M. Jerrum and G. Sorkin. The metropolis algorithm for graph bisection. *Discrete Applied Mathematics*, 82:155–175, 1998.

23 A. Juels. *Topics in black-box combinatorial function optimization.* PhD thesis, UC Berkeley, 1996.

24 R. Karp. *Reducibility among combinatorial problems.* Springer, 1972.

25 M. Karpinski. Approximability of the minimum bisection problem: An algorithmic challenge. In *Proc. 27th Mathematical Foundations of Computer Science*, pages 59–67. Springer, 2002.

26 S. Khot. Ruling out ptas for graph min-bisection, dense k-subgraph, and bipartite clique. *SIAM Journal on Computing*, 36:1025–1071, 2006.

27 L. Kučera. Expected complexity of graph partitioning problems. *Discrete Applied Mathematics*, 57:193–212, 1995.

28 M. Luczak and C. McDiarmid. Bisecting sparse random graphs. *Random Structures and Algorithms*, 18:31–38, 2001.

29 K. Makarychev, Y. Makarychev, and A. Vijayaraghavan. Approximation algorithms for semi-random partitioning problems. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 367–384. ACM, 2012.

30 L. Massoulié. Community detection thresholds and the weak ramanujan property. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 694–703. ACM, 2014.

31 F. McSherry. Spectral partitioning of random graphs. In *Proc. 42nd Foundations of Computer Science*, pages 529–537. IEEE, 2001.

32 M. Mézard and A. Montanari. *Information, physics, and computation.* Oxford University Press, 2009.

33 E. Mossel, J. Neeman, and A. Sly. Stochastic block models and reconstruction. *arXiv preprint arXiv:1202.1499*, 2012.

34 E. Mossel, J. Neeman, and A. Sly. Belief propagation, robust reconstruction, and optimal recovery of block models. *arXiv preprint arXiv:1309.1380*, 2013.

35 E. Mossel, J. Neeman, and A. Sly. A proof of the block model threshold conjecture. *arXiv preprint arXiv:1311.4115*, 2013.

36 E. Mossel, J. Neeman, and A. Sly. Consistency thresholds for the planted bisection model. *arXiv preprint arXiv:1407.1591 v2*, 2014.

37 R. Neininger and L. Rüschendorf. A general limit theorem for recursive algorithms and combinatorial structures. *The Annals of Applied Probability*, 14:378–418, 2004.

38 H. Räcke. Optimal hierarchical decompositions for congestion minimization in networks. In *Proc. 40th ACM symposium on Theory of computing*, pages 255–264. ACM, 2008.

39 M. Talagrand. The parisi formula. *Annals of Mathematics*, 163:221–263, 2006.