The stochastic block model has the overlap graph property for modularity

Shankar Bhamidi
Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, USA.

David Gamarnik
MIT Sloan School of Management.

Remco van der Hofstad
Department of Mathematics and Computer Science, Eindhoven University of Technology, The Netherlands.

Nelly Litvak
Department of Mathematics and Computer Science, Eindhoven University of Technology, The Netherlands.

Paweł Prałat
Department of Mathematics, Toronto Metropolitan University, Canada.

Fiona Skerman[1]
Department of Mathematics, Uppsala University, Sweden.

Yasmin Tousinejad
Department of Mathematics, Uppsala University, Sweden.

Abstract

The overlap gap property (OGP) is a statement about the geometry of near-optimal solutions. Exhibiting OGP implies failure of a class of local algorithms; and has been observed to coincide with conjectured algorithmic limits in problems with statistical computational gap.

We consider the Stochastic Block Model (SBM), where the graph has a planted partition with k equal-size blocks which form the 'communities', and where, for parameters $p > q$, vertices within the same community connect with probability p, while vertices in different communities connect with probability q, independently across pairs of vertices. Modularity–based clustering algorithms have become ubiquitous in applications. This article studies theoretical limits of local algorithms based on the modularity score on the SBM.

We establish that modularity exhibits OGP on the SBM. This rules out a class of local algorithms based on modularity for recovery in the SBM, and shows slow mixing time for a related Markov Chain. Theoretically this is one of the few instances where OGP has been established for a 'planted' model, as most such analyses to date consider the 'null' model.

As part of our analysis, we extend a result by Bickel and Chen 2009, who established that with high probability, the modularity optimal partition of SBM is $o(n)$ local moves away from the planted partition, where n is the graph size. We show that, with high probability, any partition with modularity score sufficiently near the optimal value is close to the planted partition.

2012 ACM Subject Classification Theory of computation → Random network models

Keywords and phrases community detection, average-case complexity, overlap gap property, modularity, Louvain, stochastic block model

Digital Object Identifier 10.4230/LIPIcs.ICALP.2026.32

Related Version https://arxiv.org/abs/2605.10911, [3]

---

[1] Corresponding author.

53rd International Colloquium on Automata, Languages, and Programming (ICALP 2026).
Editors: Sayan Bhattacharya, Danupon Nanongkai, Michael Benedikt, and Gabriele Puppis; Article No. 32; pp. 32:1–32:20
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Funding This material is based upon work supported by the National Science Foundation under Grant No. DMS-1928930, while Bhamidi, Gamarnik, van der Hofstad, Litvak, Prałat and Skerman were in residence at the Simons Laufer Mathematical Sciences Institute (formerly MSRI) in Berkeley, California, during the Spring 2025 semester.

Shankar Bhamidi: Bhamidi was partially supported by NSF DMS-2113662, DMS-2413928, DMS-2434559 and NSF RTG grant DMS-2134107.

David Gamarnik: Gamarnik partially supported by NSF Grant CISE 2233897.

Remco van der Hofstad: Partially supported by the Netherlands Organisation for Scientific Research (NWO) through the Gravitation NETWORKS grant 024.002.003.

Nelly Litvak: Partially supported by the Netherlands Organisation for Scientific Research (NWO) through the Gravitation NETWORKS grant 024.002.003.

Paweł Prałat: Partially supported by NSERC Discovery Grant.

Fiona Skerman: Partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

Yasmin Tousinejad: Partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

Acknowledgements We thank Subhabrata Sen and Ekaterina Toropova for useful discussions.

1 Introduction

Modularity maximisation [27, 28] is one of the main methods used for finding clusters and communities in networks. Modularity maximisation is NP-hard [8]; in practice, modularity has a complex landscape, and there are many high scoring solutions [18]. On the other hand, one major advantage of modularity is that its local updates are easy to compute. Naturally, local updates form the initial phase in the go-to algorithms for modularity maximisation: the Louvain algorithm [6] and its improved version, the Leiden algorithm [33]. However, as already observed in [6], algorithms based on local updates get trapped in a local maximum. To avoid this, both Louvain and Leiden include an additional non-local phase. An alternative approach, proposed by Wang and Kolter [34] via the Locale algorithm, escapes local maxima by replacing the local update with low-cardinality embedding.

Given the abundance of applications of modularity maximisation and Louvain and Leiden algorithms, theoretical understanding of modularity–based algorithms is surprisingly scarce. Cohen-Addad et al. [10] proved that in the Stochastic Block Model (SBM) with sufficient signal and two equal-sized blocks, a local algorithm, very similar to the local phase of Louvain, initiated on a random bisection, recovers the correct partition with high probability. See Section 6 for further discussion including how this relates to our result.

The question remains whether local updates lead to a global optimum in this random model with more than two communities or different initialisation. In this work, we answer this question negatively. Specifically, we establish that in SBM with $k \geq 3$ blocks, modularity exhibits the *overlap gap property* (OGP). The OGP is a statement of the geometry of near-optimal solutions and is considered a signature of algorithmic hardness—see Definition 2.1. Indeed, for many problems known to exhibit a computational-statistical gap, the threshold for OGP coincides with the conjectured algorithmic threshold [15, 16]. The idea has origins in statistical physics, and can be used to prove failure of greedy and Markov Chain Monte Carlo algorithms in average case problems. We show OGP in a planted random model as in [17, 14], though the theory is usually applied to optimisation problems in null models [15].

The practical implication of this result is that in networks with more than two communities, the local updates of Louvain or Leiden can get stuck in a local maximum even for a random graph model with clear communities. The same holds for other algorithms based on local updates. For example, Gösgens et al. [19] offer a new interpretation of modularity as an angular distance in a high-dimensional hypersphere, thus establishing equivalence of modularity maximisation and nearest-neighbour search. Then, OGP implies that greedy algorithms for nearest-neighbour search may return a *local* rather than the *global* maximum. Furthermore, using the OGP we prove that a natural generalisation of the greedy algorithm, namely the Markov Chain Monte Carlo algorithm, mixes exponentially slowly, and, in particular, takes an exponential time to reach partitions close to the planted partition with high probability.

While there are few theoretical results for modularity–based algorithms, there are positive results for recovering ground truth partitions using local-updates based on modularity in the SBM. As mentioned earlier, Cohen-Addad et al. [10] showed that for the SBM with sufficient signal and two equal-sized blocks, a local algorithm recovers the communities with high probability. This was further extended to general k by giving an algorithm with parallel local updates based on the modularity function [11]. For more details and a discussion of how these relate to our OGP results, see Section 6.

More is known about the behaviour of modularity on random models; we briefly review this for context. Considering first random graph models without planted structure, the main results focus on the setting of growing average degree $\bar{d}$ for the random d-regular graph [23, 24, 29] and the Erdős-Rényi random graph [25, 30], which are likely to have modularity of the order $1/\sqrt{\bar{d}}$. For the Preferential Attachment (PA) model [2], very recently, this same behaviour was established up to log factors [31]. For models with planted structure, break-through results were given for the SBM by Bickel and Chen [4] showing that the modularity-optimal partition is close to the planted partition. These were extended to the degree-corrected SBM in [35]. See Theorem 2.4 for details. Finally, the modularity in the Artificial Benchmark for Community Detection (ABCD) model, a model similar to the well-known LFR model [21] used by practitioners, was studied in [20].

2 Results

2.1 Definitions

Let $G = (V, E)$ be a graph with $m = |E| \geq 1$ edges and $n = |V| \geq 1$ nodes. For a partition $\mathcal{A}$ of the vertices of G, the *modularity score* of $\mathcal{A}$ on G is defined as

$$q_{\mathcal{A}}(G) = \frac{1}{2m} \sum_{A \in \mathcal{A}} \sum_{u,v \in A} \left(\mathbf{1}_{\{uv \in E\}} - \frac{d_u d_v}{2m} \right) = \sum_{A \in \mathcal{A}} \frac{e(A)}{m} - \sum_{A \in \mathcal{A}} \left(\frac{\mathrm{vol}(A)}{\mathrm{vol}(V)} \right)^2, \tag{1}$$

where d_u denotes the degree of node u, $e(A)$ denotes the number of edges within $A \subseteq V$ and $\mathrm{vol}(A) = \sum_{v \in A} d_v$ denotes the (degree) volume of the set A. The *modularity* of a graph G is defined as $q^\star(G) = \max_{\mathcal{A}} q_{\mathcal{A}}(G)$, where the maximum is taken over all partitions $\mathcal{A}$ of the nodes of G. It will be useful to express modularity as the difference of the *edge-contribution* or *coverage* $q^E_{\mathcal{A}}$ and the *degree-tax* $q^D_{\mathcal{A}}$ of the partition $\mathcal{A}$, defined as

$$q^E_{\mathcal{A}}(G) = \sum_{A \in \mathcal{A}} e(A)/m \quad \text{and} \quad q^D_{\mathcal{A}}(G) = \sum_{A \in \mathcal{A}} \mathrm{vol}(A)^2/\mathrm{vol}(G)^2. \tag{2}$$

The modularity score, introduced by Newman and Girvan in [28], is a quality function of many popular community detection algorithms such as Louvain [6] or Leiden [33]. Indeed, the modularity score favours partitions of the set of nodes of a graph G in which a large proportion of the edges fall entirely within the partition, but benchmarks it against the

Figure 1 The figure shows the planted partition $\mathcal{P}$ (left) and a 'decoy' partition $\mathcal{D}$ (right) for the SBM which has a planted partition of four equal-size parts. The partitions are denoted by showing the adjacency matrix of G and colouring cell u, v with colour i if nodes u, v are placed together in part / community i in the partition. A 'decoy' partition $\mathcal{D}$ is the one in which two of the planted blocks are clustered together in the same part, and each of the remaining planted blocks form their own part.

expected number of edges one would see in the same partition in a corresponding Chung-Lu random graph model [9] with expected degree sequence taken to be identical to the degree sequence of the observed graph G.

Fix $k \geq 2$ denoting the number of communities. We now formally define the *stochastic block model* (SBM) which will be the focus of this paper. This model is also called the *planted partition model.* The set of nodes $[n] = \{1, \ldots, n\}$ is partitioned into k communities $P_1, \ldots, P_k$ of as equal sizes as possible; that is, $\lfloor n/k \rfloor \leq |P_i| \leq \lceil n/k \rceil$ for all $i \in [k]$. For simplicity, often we will assume that n is divisible by k, but our main result is stated and holds without this restriction. We will refer to this *planted partition* $(P_i)_{i=1}^k$ as $\mathcal{P}$. The probability of observing an edge between two nodes of the same community is equal to p; otherwise, it is equal to q. We will use the notation $G \in \mathcal{G}(n, k, p, q)$ whenever a random graph G is generated with this probability distribution.

Results in the paper are in the large network limit $n \to \infty$. In particular, we will assume that both $p = p(n)$ and $q = q(n)$ are functions of n, and n is large enough for certain statements to be true. On the other hand, the parameter $k \geq 2$ is an arbitrary but *fixed* integer. We emphasise that the notations $o(\cdot)$ and $O(\cdot)$ refer to functions of n, not necessarily positive, whose growth vanishes, respectively, is bounded.

We will establish that modularity exhibits OGP in the SBM. To prove this, we will obtain a more detailed result that characterises the maximum modularity of any partition with given distance to the planted partition. We define the distance as the classification error—see, for example, (4.25) in [1]; in case of two communities, this distance is equal to the 'imbalance' in [10]. Define the distance between a k-part partition $\mathcal{A} = (A_i)_{i=1}^k$ and the planted partition $\mathcal{P} = (P_i)_{i=1}^k$ in SBM as

$$d(\mathcal{A}, \mathcal{P}) = 1 - \frac{1}{n} \max_{\sigma} \sum_{i=1}^{k} |A_{\sigma(i)} \cap P_i|, \tag{3}$$

where the maximum is taken over all permutations $\sigma : [k] \to [k]$ of $[k] = \{1, \ldots, k\}$ that govern how parts of $\mathcal{A}$ are aligned with the ones of $\mathcal{P}$. Note that this $d(\mathcal{A}, \mathcal{P})$ has a natural interpretation: it is the minimum proportion of nodes that need to be re-shuffled to transform the candidate partition $\mathcal{A}$ into the ground truth or planted partition $\mathcal{P}$. Indeed, for a given permutation σ, one can keep nodes in $A_{\sigma(i)} \cap P_i$ where they are and move other nodes to the appropriate parts.

Now that we have a notion of distance between partitions, we may formally define the *overlap gap property* (OGP). The definition of the OGP pertains to a particular instance G_n of a distribution over graphs $\mathcal{G}_n$.

▶ **Definition 2.1 (Overlap gap property – planted model).** *For a graph G_n with planted partition $\mathcal{P}$, the optimisation problem $\max_{\mathcal{A}} q_{\mathcal{A}}(G_n)$ exhibits OGP with values $\mu > 0$ and $0 \leq \nu_1 < \nu_2$ if the following holds: For any partition $\mathcal{B}$ for which $q_{\mathcal{B}}(G_n) \geq q^\star(G_n) - \mu$, it holds that either $d(\mathcal{B}, \mathcal{P}) \leq \nu_1$, or that $d(\mathcal{B}, \mathcal{P}) \geq \nu_2$. Furthermore, $d(\mathcal{B}', \mathcal{P}) \geq \nu_2$ does indeed occur for some partition $\mathcal{B}'$ with $q_{\mathcal{B}'}(G_n) \geq q^\star(G_n) - \mu$.*

2.2 Modularity exhibits OGP

Theorem 2.2 below states that the modularity score exhibits OGP on the SBM, provided that the number of communities, k, is at least 3:

▶ **Theorem 2.2 (Modularity has OGP).** *Fix real numbers $a > b > 0$, and integer $k \geq 3$. For any $\nu \in \left(\frac{1}{2(k-1)}, \frac{1}{k}\right)$ and $\varepsilon > 0$, there exist $\mu = \mu(\nu)$ and $c = c(\varepsilon)$ such that the following holds for n large enough.*

Let $p = p(n) = \omega\, a/n$ and $q = q(n) = \omega\, b/n$ be such that $\omega > c$, and let $G \in \mathcal{G}(n, k, p, q)$ with planted partition $\mathcal{P}$. Then with probability at least $1 - \varepsilon$, for every k-part partition $\mathcal{A}$ with

$$q_{\mathcal{A}}(G) \geq \frac{a-b}{a+(k-1)b}\left(1 - \frac{1}{k} - \frac{2}{k^2}\right) - \mu, \tag{4}$$

either $d(\mathcal{A}, \mathcal{P}) \leq \frac{1}{2(k-1)}$ or $d(\mathcal{A}, \mathcal{P}) \geq \nu$. Moreover, there are partitions $\mathcal{A}$ satisfying the latter.

Intuition. Consider a ‘decoy’ partition $\mathcal{D}$ in which two of the planted blocks are clustered together in the same part, and each of the remaining planted blocks forms its own part. (There are $\binom{k}{2}$ such decoy partitions by symmetry.) We call $\mathcal{D}$ a ‘decoy’ partition since it has a ‘misleadingly’ high modularity score given its distance to the planted partition $\mathcal{P}$; see Figure 1. (We emphasise that the decoy $\mathcal{D}$ is an alternate partition of the nodes rather than an alternate random graph model to generate the graph.)

Given the decoy partition $\mathcal{D}$, the parameters for OGP are set as follows. We set $\mu = q^\star(G) - q_{\mathcal{D}}(G) + \eta$ for some small $\eta > 0$. The idea is that partitions $\mathcal{B}$ with $q_{\mathcal{B}}(G) \geq q^\star(G) - \mu$ will include the decoy partition, those *very* ‘close to’ the decoy, as well as partitions somewhat ‘close to’ the planted (and modularity optimal) partition. Then we set $\nu_2 = d(\mathcal{P}, \mathcal{D}) - \delta = 1/k - \delta$ for some small $\delta > 0$ so that the decoy partition and those very close to it will fall within the second interval. Lastly, we need to set ν_1. Notice that, for small distances from the planted partition, the modularity score of partitions decreases with distance, so we set ν_1 to be some distance by which the likely modularity score has dipped below $q^\star(G) - \mu$. This explains the intuition behind Theorem 2.2.

As an implication of the OGP we show that a natural algorithm based on Markov Chain Monte Carlo (MCMC) updates takes exponential time to reach any proximity of the ground truth. We state our result informally, and defer the formal statement as well as the proof to Appendix B.

▶ **Theorem 2.3** (Informal). *For every $\zeta > 0$, there exists a large enough inverse temperature parameter $\beta > 0$, such that when the chain is initiated at distance at least $1/k$ from the ground truth, the time to reach a partition close to the ground truth is at least $\exp(\Theta(n))$ with high probability.*

2.3 Modularity as a predictor of the planted partition

In a breakthrough result of Bickel and Chen [4], it was shown that in the regime of growing degree, the modularity optimal partition of the stochastic block model is within distance $o(1)$ of the planted partition (see also [5]):

▶ **Theorem 2.4** ([4, 5]). *Fix real numbers* $a > b > 0$*, integer* $k \geq 2$ *and let* $\omega = \omega(n) \to \infty$ *as* $n \to \infty$*. Let* $p = \omega\, a/n$ *and* $q = \omega\, b/n$*, and let* $G \in \mathcal{G}(n, k, p, q)$ *with the planted partition* $\mathcal{P}$*. Then,*

$$d\big(\arg\max q_{\mathcal{A}}(G),\, \mathcal{P}\big) = o(1).$$

Since modularity-based clustering algorithms remain the most popular, despite known problems such as the resolution limit [13], it is reassuring that the modularity optimal partition is quite close to the 'correct' partition in the SBM, a natural simplified model of community structure.

However, we would like to know more. In particular, what about partitions with nearly optimal modularity score? Are these partitions also close to the planted partition? Our Theorem 2.5 (along with some bounds on modularity scores for partitions at higher distances) implies that the answer is 'yes': if the partition has a score within a certain distance of the maximal modularity value, then the partition must be near the planted one; see the full version of the paper for the proof.

todo

▶ **Theorem 2.5 (Partitions with near-optimal modularity are close to the planted).** *Fix* $a > b > 0$ *and integer* $k \geq 2$*. Let* $p = p(n) = \omega\ a/n$ *and* $q = q(n) = \omega\ b/n$*. For any* $\varepsilon > 0$ *and* $\delta < \frac{a-b}{a+(k-1)b} \cdot \frac{2}{k^2}$*, there exists* $c = c(\varepsilon)$ *such that, if* $\omega > c$ *for* n *large enough, then the following holds with probability at least* $1 - \varepsilon$ *provided* n *is large enough.*

Let $G \in \mathcal{G}(n, k, p, q)$ *with the planted partition* $\mathcal{P}$*. Then,* $q_{\mathcal{A}}(G) > q^\star(G) - \delta$ *for some* k*-part partition* $\mathcal{A}$ *implies that* $d(\mathcal{A}, \mathcal{P}) < \delta' + \varepsilon$*, where* $\delta' = \delta'(\delta) \in \big(0, \frac{1}{k(k-1)}\big)$*.*

2.4 Modularity at distance at most $1/k$

The next theorem relates the maximum modularity of a partition to its distance from the planted partition, and is the main ingredient in the proof of Theorem 2.2. For any $k \geq 2$, we approximate the maximum modularity value over all partitions of $k \geq 2$ parts $\mathcal{A}$ at distance t/k (for some $t \in [0, 1]$) from the planted partition $\mathcal{P}$.

▶ **Theorem 2.6 (Maximal modularity at distance at most $1/k$).** *Fix* $a > b > 0$*. Let* $p = p(n) = \omega\ a/n$ *and* $q = q(n) = \omega\ b/n$*. For any* $\varepsilon > 0$*, there exists* $c = c(\varepsilon)$ *such that if* $\omega > c$ *then the following holds with probability at least* $1 - \varepsilon$ *provided* n *is large enough.*

Let $G \in \mathcal{G}(n, k, p, q)$ *with the planted partition* $\mathcal{P}$*. For any* $d \in \big[0, \frac{1}{k}\big]$*, let*

$$H(d) = \max_{\mathcal{A}:|\mathcal{A}| \leq k} \left\{ q_{\mathcal{A}}(G) : d - \frac{1}{\sqrt{n}} \leq d(\mathcal{A}, \mathcal{P}) \leq d + \frac{1}{\sqrt{n}} \right\}.$$

Then,

$$\left| H(d) - \frac{a-b}{a+(k-1)b} \cdot h(d) \right| \leq \varepsilon, \quad \textit{where} \quad h(d) = 1 - \frac{1}{k} - 2d\big(1 - d(k-1)\big).$$

The random variable $H(d)$ is defined as the maximal modularity over all k-part partitions with the distance from $\mathcal{P}$ in an interval $d \pm 1/\sqrt{n}$. We keep the margins $\pm 1/\sqrt{n}$ for the technical reason that there might not be any partition at distance exactly d (say, when

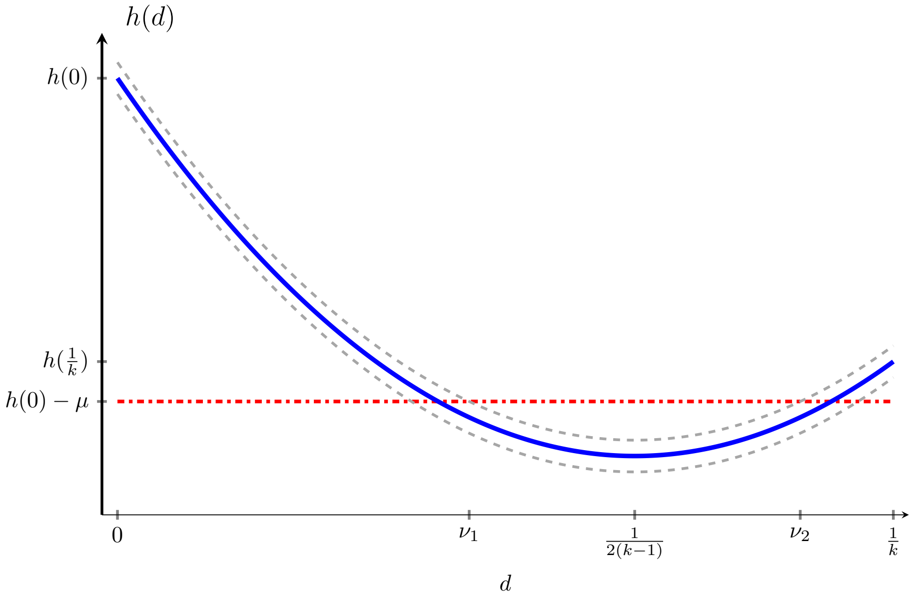

Figure 2 A graph of the function $h(d)$ for $k = 4$ and the relation to the OGP parameters μ, ν_1 and ν_2 – see the text following Theorem 2.6. The dashed grey lines indicate an ε-region above and below $h(d)$. By Theorem 2.6, any partition $\mathcal{A}$ at distance d is likely to have modularity at most $h(d) + \varepsilon$. From this, we will be able to conclude that it will be likely that any partition $\mathcal{A}$ with modularity at least $h(0) - \mu$ (i.e., above the red line) will be at a distance $d \leq \nu_1$ or $d \geq \nu_2$, thus establishing the OGP with these parameters.

$d = \pi/4k$, since the distance is always a rational number). Theorem 2.6 establishes that $H(d)$ is likely to be well approximated by the deterministic expression $\frac{a-b}{a+(k-1)b} \cdot h(d)$.

Note that function $d \mapsto h(d)$ decreases on $\left[0, \frac{1}{2(k-1)}\right]$ and increases on $\left[\frac{1}{2(k-1)}, \frac{1}{k}\right]$. As a result,

- $h(0) = 1 - \frac{1}{k}$,
- $h(d)$ is decreasing on $[0, \frac{1}{2(k-1)}]$ reaching $h\left(\frac{1}{2(k-1)}\right) = 1 - \frac{1}{k} - \frac{1}{2(k-1)}$,
- $h(d)$ is increasing on $[\frac{1}{2(k-1)}, \frac{1}{k}]$ reaching $h(1/k) = 1 - \frac{1}{k} - \frac{2}{k^2}$.

One can easily separate $h(1/k)$ from its local minimum $h(\frac{1}{2(k-1)})$ by introducing a threshold μ in Theorem 2.2, say, $\mu = h(0) - h(\nu)$ with $\nu \in \left(\frac{1}{2(k-1)}, \frac{1}{k}\right)$. In Figure 2, we illustrate $h(d)$ for $d \in [0, \frac{1}{k}]$ and the interplay with the OGP parameters μ, ν_1 and ν_2.

Below, we will show Theorem 2.2, i.e., that modularity has OGP in the SBM – assuming Theorem 2.6, where we recall that Theorem 2.6 characterises the maximum modularity score over partitions $\mathcal{A}$ at given distances $d \in [0, \frac{1}{k}]$ from the planted partition $\mathcal{P}$. We outline the proof of Theorem 2.6 in Sections 3–5, and give some further details Appendix C. We prove Theorem 2.2 using Theorem 2.6 in Appendix D.

The asymptotic maximal modularity value in $\mathcal{G}(n, k, p, q)$ is known and also known to coincide with the modularity of the planted partition; we state it in Theorem 2.7 below for completeness. The result follows from Theorem 2.4 and was also given in [26] for slightly higher p, q. We note that (5) in Theorem 2.7 is a direct corollary of Theorem 2.6 as a special case when $d = 0$. Also (5) follows, for example, by Lemma 4.1.

▶ **Theorem 2.7** **(Maximal modularity).** *Fix* $a > b > 0$ *and integer* $k \geq 2$*. Let* $p = p(n) = \omega\, a/n$ *and* $q = q(n) = \omega\, b/n$*. For any* $\varepsilon > 0$*, there exists* $c = c(\varepsilon)$ *such that, if* $\omega > c$ *for* n *large enough, then the following holds with probability at least* $1 - \varepsilon$ *provided* n *is large enough. Let* $G \in \mathcal{G}(n, k, p, q)$ *and denote the planted partition by* $\mathcal{P}$*. Then,*

$$\left| q_{\mathcal{P}}(G) - \frac{a-b}{a+(k-1)b}\left(1 - \frac{1}{k}\right) \right| \leq \frac{\varepsilon}{2} \tag{5}$$

and

$$\left| q^{\star}(G) - q_{\mathcal{P}}(G) \right| \leq \frac{\varepsilon}{2}. \tag{6}$$

As a result,

$$\left| q^{\star}(G) - \frac{a-b}{a+(k-1)b}\left(1 - \frac{1}{k}\right) \right| \leq \varepsilon.$$

Note that, as expected, $q^{\star}(G) \approx \frac{a-b}{a+(k-1)b}\, h(0)$, where $h(\cdot)$ is defined in Theorem 2.6.

3 Ingredients of the proof : Distances to the planted partition $\mathcal{P}$

For now, for simplicity of exposition, we assume that n is divisible by k, as this will be assumed in the first lemma, Lemma 3.1. However, in the proof of Theorem 2.6, we relax this condition and deal with the general case.

To prove our main result, we need to investigate a family of partitions of $[n]$ with a given distance to the planted partition $\mathcal{P} = (P_i)_{i=1}^k$. It will be convenient to represent partitions as $k \times k$ matrices that capture the way how these partitions overlap with $\mathcal{P}$. Let $\mathcal{A} = (A_i)_{i=1}^k$ be any partition of $[n]$ into k parts. Then, the *signature* of $\mathcal{A}$ (with respect to $\mathcal{P}$) is defined as the matrix $X = X(\mathcal{A}) = (x_{ij})_{1 \leq i \leq j \leq k}$, where

$$x_{ij} = \frac{|A_i \cap P_j|}{n/k}. \tag{7}$$

Note that for the signature to be well defined, one needs to fix the labelling of both $\mathcal{P}$ and $\mathcal{A}$. To compute the distance between $\mathcal{P}$ and $\mathcal{A}$, one needs to align these labels as best as possible (see Lemma 3.1 below), but the modularity score clearly does not depend on the way the labels are aligned (see Lemma 4.1 below). Note also that, trivially, $x_{ij} \geq 0$ for all $i, j \in [k]$. Moreover, since $A_1 \cup \cdots \cup A_k = [n]$, for any $j \in [k]$,

$$\sum_{i \in [k]} x_{ij} = \frac{1}{n/k} \sum_{i \in [k]} |A_i \cap P_j| = \frac{|P_j|}{n/k} = 1. \tag{8}$$

Finally, note that the signature of the planted partition $\mathcal{P}$ is the matrix $Y^{\mathcal{P}} = (y_{ij})_{1 \leq i \leq j \leq k}$ with $y_{ii} = 1$ for all $i \in [k]$ and $y_{ij} = 0$ for all $i \neq j$.

In our first lemma, we rewrite the distance between any partition $\mathcal{A}$ and the planted partition $\mathcal{P}$ in a convenient form, namely, as a function of the signature of $\mathcal{A}$. The proof is short so we give it below.

▶ **Lemma 3.1.** *Suppose that* n *is divisible by* k*. Let* $\mathcal{A}$ *be any partition of* $[n]$ *with* k *parts, and let* $\mathcal{P}$ *be the planted partition. Let* $X = X(\mathcal{A}) = (x_{ij})_{1 \leq i \leq j \leq k}$ *be the signature of* $\mathcal{A}$*. Then,*

$$d(\mathcal{A}, \mathcal{P}) = 1 - \tfrac{1}{k} \max_{\sigma} \sum_{i=1}^{k} x_{\sigma(i)i}, \tag{9}$$

where the maximum is taken over all permutations of $[k]$*.*

Moreover, $d(\mathcal{A},\mathcal{P}) \leq 1 - 1/k$ *and this upper bound is sharp, that is, there exists a partition* $\mathcal{A}$ *such that* $d(\mathcal{A},\mathcal{P}) = 1 - 1/k$*, provided that* n *is divisible by* k^2*.*

Proof of Lemma 3.1. Equality (9) follows immediately from the definition of the signature in (7) and the definition (3) of the distance between the two partitions. Indeed,

$$d(\mathcal{A},\mathcal{P}) = 1 - \frac{1}{n}\max_{\sigma}\sum_{i=1}^{k}|A_{\sigma(i)} \cap P_i| = 1 - \frac{1}{n}\max_{\sigma}\sum_{i=1}^{k}x_{\sigma(i)i}\cdot(n/k) = 1 - \frac{1}{k}\max_{\sigma}\sum_{i=1}^{k}x_{\sigma(i)i}.$$

It remains to show that $d(\mathcal{A},\mathcal{P}) \leq 1 - 1/k$ or, equivalently, that $\max_\sigma \sum_{i=1}^k x_{\sigma(i)i} \geq 1$.

Consider the sum $S = \sum_\sigma \sum_{i=1}^k x_{\sigma(i)i}$, where the outer sum is taken over all $k!$ permutations of $[k]$. Clearly, there are $k! \cdot k$ terms in S and for each $a, b \in [k]$, the term x_{ab} occurs exactly $(k-1)!$ times (i has to be equal to b and there are $(k-1)!$ permutations that map b to a). Hence, for each $j \in [k]$, the sum $\sum_{i=1}^k x_{ij}$ (which is equal to 1 by (8)) occurs $(k-1)!$ times. We conclude that $S = k \cdot (k-1)! = k!$. By an averaging argument, there exists a permutation $\hat{\sigma}$ for which $\sum_{i=1}^k x_{\hat{\sigma}(i)i} \geq 1$. Hence, $\max_\sigma \sum_{i=1}^k x_{\sigma(i)i} \geq \sum_{i=1}^k x_{\hat{\sigma}(i)i} \geq 1$, thus the desired inequality holds.

Assume now that n is not only divisible by k but, in fact, it is divisible by k^2. We construct a partition $\mathcal{A}$ by partitioning each P_j into k equal parts and then picking $1/k$ fraction of each P_j to form A_i of size n/k. This partition has signature X with $x_{ij} = 1/k$ for all $i, j \in [k]$. We get that $\max_\sigma \sum_{i=1}^k x_{\sigma(i)i} = \sum_{i=1}^k 1/k = 1$, which shows that the upper bound is sharp. This finishes the proof of the lemma. ◀

4 Ingredients of the proof: Concentration

The signature of $\mathcal{A}$ not only determines the distance between $\mathcal{A}$ and $\mathcal{P}$ but, more importantly, it predicts (up to an arbitrarily small error in a sufficiently dense graph G) the modularity score $q_{\mathcal{A}}(G)$ of $\mathcal{A}$ for $G \in \mathcal{G}(n,k,p,q)$ generated by the SBM. Indeed we will prove that the modularity score of $\mathcal{A}$ is well-approximated by a scaling of $g(X)$, a function only of the signature X of $\mathcal{A}$ that is defined by

$$g(X) = \sum_{i=1}^{k}\sum_{1\leq j<j'\leq k}(x_{ij} - x_{ij'})^2. \tag{10}$$

▶ **Lemma 4.1 (Concentration of modularity).** *Fix* $a > b > 0$ *and integer* $k \geq 2$*. Let* $p = p(n) = \omega\, a/n$ *and* $q = q(n) = \omega\, b/n$*. For any* $\varepsilon > 0$*, there exists* $c = c(\varepsilon)$ *such that if* $\omega > c$ *for* n *large enough, then the following holds with probability at least* $1 - \varepsilon$ *provided* n *is large enough.*

Let $G \in \mathcal{G}(n,k,p,q)$*. For all* k*-part partitions* $\mathcal{A}$*,*

$$\left| q_{\mathcal{A}}(G) - \frac{a-b}{(a+(k-1)b)}\cdot\frac{g(X)}{k^2}\right| < \varepsilon,$$

where $X = X(\mathcal{A}) = (x_{ij})_{1\leq i,j\leq k}$ *is the signature of* $\mathcal{A}$ *and* $g(X)$ *is defined in* (10)*.*

Recall that the planted partition $\mathcal{P}$ has signature $Y^{\mathcal{P}} = (y_{ij})_{1\leq i\leq j\leq k}$ with $y_{ii} = 1$ for all $i \in [k]$ and $y_{ij} = 0$ for all $i \neq j$. Since $g(Y^{\mathcal{P}}) = k \cdot (k-1)$, we get that $q_{\mathcal{P}}(G)$ is very close to $\left(1 - \frac{1}{k}\right)\cdot\frac{a-b}{a+(k-1)b}$.

As one might want to use this result for other purposes, we state it (and prove it, of course) without assuming that n is divisible by k. The proof of Lemma 4.1 can be found in the full version of the paper, though we describe some intermediate steps below. One of these intermediate steps concerns the modularity score of partitions $\mathcal{A}$ with given signature X on a deterministic *weighted graph*. This can be regarded as a mean-field calculation. Indeed, working with the deterministic weighted graph means we may consider, for example, the edge-weight of edges in part A_i between the planted blocks P_j and $P_{j'}$, which corresponds to the *expected* number of such edges. After that step, it remains only to make a link between the modularity scores of our random graph and those of the deterministic weighted graph.

todo

todo

5 Ingredients of the proof: Optimisation

We have now established that, up to small errors, we may consider $g(X)$ as our objective function, and have re-written our distance to the planted partition in terms of X. Thus, the new aim is to maximise $g(X)$ given the distance of X to the planted partition $Y^{\mathcal{P}}$.

▶ **Lemma 5.1.** *Let* $0 \leq t < 1$*. Define* $\mathcal{X}_k(t)$ *to be the family of matrices* $X = (x_{ij})_{1\leq i\leq j\leq k}$ *such that*

$$\begin{aligned} & x_{ij} \geq 0, && \text{for all } i,j \in [k], \\ & \sum_i x_{ij} = 1 && \text{for all } j \in [k], \qquad (11) \\ & \sum_{i\neq j} x_{ij} = t, \text{ and} \\ & \sum_i x_{ii} \geq \max_\sigma \sum_i x_{\sigma(i)i}. \end{aligned}$$

Let $g(X)$ *be defined as in* (10)*. Then,*

$$\max_{X\in\mathcal{X}_k(t)} g(X) = k(k-1) - 2t\big(k - t(k-1)\big) \text{ and } \arg\max_{X\in\mathcal{X}_k(t)} g(X) = \{X_k^{(ij)}(t) \ : \ i \neq j\},$$

where, for $i \neq j$*,* $X_k^{(ij)}(t)$ *is the* $k \times k$ *matrix with* $x_{ij} = t$*,* $x_{jj} = 1 - t$*, all other diagonal elements* 1 *and non-diagonal elements* 0 *(i.e.,* $x_{aa} = 1$ *for all* $a \neq j$ *and* $x_{ab} = 0$ *for all* $a \neq b$ *such that* $(a, b) \neq (i, j)$*).*

Note that in the optimiser $X_k^{(i,j)}(t)$, the row sums are $(1+t, 1-t, 1, \ldots, 1)$, corresponding to part sizes $(1/k + d, 1/k - d, 1/k, \ldots, 1/k) \cdot n$ in the partition at distance $d < 1/k$ with $d = t/k$ (or, thus, equivalently, $t < 1$).

The following lemma shows that the maximal $g(X)$ for signatures at distance $d > 1/k$ (i.e., for $t > 1$) is at most the maximal $g(X)$ for signatures at distance $d = 1/k$. This result is important to several of our proofs as it will allow us to infer that partitions $\mathcal{A}$ with high modularity scores must be at distance $d \leq 1/k$ from the planted partition.

▶ **Lemma 5.2.** *Consider* $\mathcal{X}_k(t)$ *as defined in* (11)*, and* $g(X)$ *as defined in* (10)*. Then, for any* $t > 1$*,*

$$\max_{X\in\mathcal{X}_k(t)} g(X) < k(k-1) - 2.$$

The optimisation problems in Lemmas 5.1 and 5.2 are quadratic maximisation problems. (Note that in general quadratic maximisation problems are NP-hard [32].) We prove both lemmas in Appendix C. Lemma 5.1 is the last ingredient needed to prove Theorem 2.6, which gives the likely optimal modularity score over all partitions at distance $d \in [0, 1/k]$ from the planted partition. (Recall that we will prove our main OGP result, Theorem 2.2, in Appendix D by appealing to Theorem 2.6.) Lemma 5.2 is used to show slow mixing of a natural MCMC in Theorem B.2.

To understand how these lemmas fit in, recall that the distance to the planted partition $\mathcal{P}$ of a partition $\mathcal{A}$ with signature X is $d(\mathcal{A}, \mathcal{P}) = 1 - \frac{1}{k} \max_\sigma \sum_i x_{\sigma(i)i} = \frac{1}{k} \min_\sigma \sum_{i \neq j} x_{\sigma(i)j}$, by Lemma 3.1 and since $\sum_j x_{ij} = 1$ for each i, so that $x_{\sigma(i)i} = 1 - \sum_{j:\, j \neq i} x_{\sigma(i)j}$. The condition $\sum_{i \neq j} x_{ij} = t$ corresponds to enforcing a distance of t/k from the planted partition, and the arg max tells us the set of partitions that achieve the maximal modularity score at distance t/k. Note that, for $t = 1$, i.e. at distance of $d = 1/k$ from the planted partition, $X_k^{(ij)}(1)$ is the signature for the 'decoy' partition $\mathcal{D}$ where two planted blocks P_i and P_j are placed within the same part, and all other planted blocks are placed within their own part – see also Figure 1.

6 Discussion and the case of optimising only over balanced partitions.

The literature contains positive results, with algorithms using local updates based on the modularity function to recover communities in the SBM [10, 11]. This paper proves OGP, a signature of algorithmic hardness, for such algorithms.

Of course this does not give rise to a contradiction, since the setups in [10, 11] and ours are subtly different and also OGP has been exhibited for problems known to be easy [22]. However, it leads to interesting open questions, probing which of the differences in the two setups are important. We describe this now in more detail.

As mentioned earlier, for $k = 2$ communities and starting with a random partition into equal-size parts, a local algorithm based on the modularity function was shown to recover the planted communities [10]. The local moves for this algorithm take the form of a *swap*, which takes pairs of vertices, one from each part, and swaps them if this increases the modularity score. This naturally maintains equal-size parts. For general k, an algorithm with parallel local updates based on the modularity function was shown to recover the ground truth partition [11]. We note that this parallel algorithm has a random balanced start and also a mechanism to maintain the balanced sizes of parts during the algorithm.

For the overlap gap property that we establish, it was important that the 'decoy' partition $\mathcal{D}$, which is an *unbalanced* partition into $k-1$ parts, has a 'surprisingly high' modularity given its distance of $1/k$ from the ground truth partition $\mathcal{P}$. Furthermore, at distances $0 < d < 1/k$ from $\mathcal{P}$ modularity, optimal partitions are increasingly unbalanced, with part sizes $(1/k + d,\ 1/k - d,\ 1/k, \ldots, 1/k)$; see Lemma 5.1.

Perhaps a crucial difference in the two setups is that the positive results of [10, 11] had *balanced* partitions, both in the initialisation as well as during the algorithm. We finally show that if we introduced this balanced condition into our optimisation problem then it no longer exhibits OGP. In particular, when considering $\max_{\mathcal{A}} q_{\mathcal{A}}$ where the maximisation is over *balanced* k-part partitions instead of all k-part partitions, we see that this no longer exhibits OGP. We should note that this by itself does not imply success of particular greedy algorithms – and hence fast algorithms. More work needs to be done to exhibit one (as done in [10, 11]). In a similar vain, for the problem of sparse regression [17], in the regime where OGP ceases to hold a greedy type algorithm was established to be effective. The construction is not based on OGP. Instead it relies directly on the properties of the model, which is typically the case.

To state this result, we fix the random graph $G \in \mathcal{G}(n, k, p, q)$ under the conditions of Theorem 2.6. It turns out that the maximum modularity of G only over *balanced* partitions at distance about $0 < d < 1 - 1/k$ from the planted partition $\mathcal{P}$ concentrates about some function that now *is* decreasing in d:

▶ **Proposition 6.1.** *Let* $\mathcal{X}'_t$ *be defined as in* (11) *with the additional restriction that* $\sum_j x_{ij} = 1$ *for all* $i \in [k]$*. Then* $g^{\mathrm{bal}}(t) = \max_{X \in \mathcal{X}'_t} g(X)$ *is* strictly *decreasing on* $0 < t < k - 1$.

Proposition 6.1 implies an analogue of Theorem 2.6, optimising now only over *balanced partitions*, for $0 < d < 1 - 1/k$, now with decreasing $\tilde{h}(d)$, contrary to $h(d)$.

References

1 K. Avrachenkov and M. Dreveton. *Statistical Analysis of Networks.* Now Publishers, 2022.

2 A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

3 S. Bhamidi, D. Gamarnik, R. van der Hofstad, N. Litvak, P. Prałat, F. Skerman, and Y. Tousinejad. The stochastic block model has the overlap graph property for modularity. *arXiv preprint arXiv:2605.10911*, 2026.

4 P. J. Bickel and A. Chen. A nonparametric view of network models and Newman–Girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50), 2009.

5 P. J. Bickel, A. Chen, Y. Zhao, E. Levina, and J. Zhu. Correction to the proof of consistency of community detection. *The Annals of Statistics*, 2015.

6 V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.

7 A. Bovier and F. den Hollander. *Metastability*, volume **351** of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer, 2015. `doi:10.1007/978-3-319-24777-9`.

8 U. Brandes, D. Delling, M. Gaertler, R. Gorke, M. Hoefer, Z. Nikoloski, and D. Wagner. On modularity clustering. *IEEE Transactions on Knowledge and Data Engineering*, 20(2):172–188, 2007.

9 F. R. K. Chung and L. Lu. *Complex graphs and networks.* Number 107. American Mathematical Soc., 2006.

10 V. Cohen-Addad, A. Kosowski, F. Mallmann-Trenn, and D. Saulpic. On the power of Louvain in the stochastic block model. *Advances in Neural Information Processing Systems*, 33:4055–4066, 2020.

11 V. Cohen-Addad, F. Mallmann-Trenn, and D. Saulpic. A massively parallel modularity-maximizing algorithm with provable guarantees. In *Proceedings of the 2022 ACM Symposium on Principles of Distributed Computing*, pages 356–365, 2022.

12 A. Coja-Oghlan, A. Galanis, L. A. Goldberg, J. B. Ravelomanana, D. Štefankovič, and E. Vigoda. Metastability of the Potts ferromagnet on random regular graphs. *Comm. Math. Phys.*, **401**(1):185–225, 2023. `doi:10.1007/s00220-023-04644-6`.

13 S. Fortunato and M. Barthelemy. Resolution limit in community detection. *Proceedings of the national academy of sciences*, 104(1):36–41, 2007.

14 D. Gamarnik. The overlap gap property: A topological barrier to optimizing over random structures. *Proceedings of the National Academy of Sciences*, 118(41):e2108492118, 2021.

15 D. Gamarnik. Turing in the shadows of Nobel and Abel: An algorithmic story behind two recent prizes. *Notices of the American Mathematical Society*, 72(5):485–493, 2025.

16 D. Gamarnik, C. Moore, and L. Zdeborová. Disordered systems insights on computational hardness. *Journal of Statistical Mechanics: Theory and Experiment*, 2022(11):114015, 2022.

17 D. Gamarnik and I. Zadik. Sparse high-dimensional linear regression. Estimating squared error and a phase transition. *The Annals of Statistics*, 50(2):880–903, 2022.

18 B. Good, Y.-A. De Montjoye, and A. Clauset. Performance of modularity maximization in practical contexts. *Physical Review E*, 81(4):046106, 2010.

19 M. Gösgens, R. van der Hofstad, and N. Litvak. The hyperspherical geometry of community detection: modularity as a distance. *Journal of Machine Learning Research*, 24(112):1–36, 2023.

20 B. Kamiński, B. Pankratz, P. Prałat, and F. Théberge. Modularity of the ABCD random graph model with community structure. *Journal of Complex Networks*, 10(6):cnac050, 2022.

21 A. Lancichinetti, S. Fortunato, and F. Radicchi. Benchmark graphs for testing community detection algorithms. *Physical Review E*, 78(4):046110, 2008.

22 S. Li and T. Schramm. Some easy optimization problems have the overlap-gap property. In *Proceedings of Thirty Eighth Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pages 3582–3622, 2025.

23 L. Lichev and D. Mitsche. On the modularity of 3-regular random graphs and random graphs with given degree sequences. *Random Structures & Algorithms*, 61(4):754–802, 2022.

24 C. McDiarmid and F. Skerman. Modularity of regular and treelike graphs. *Journal of Complex Networks*, 6(4):596–619, 2018.

25 C. McDiarmid and F. Skerman. Modularity of Erdős-Rényi random graphs. *Random Structures & Algorithms*, 57(1):211–243, 2020.

26 C. McDiarmid and F. Skerman. Modularity and partially observed graphs. *arXiv preprint arXiv:2112.13190*, 2021.

27 M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.

28 M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113, 2004. `doi:10.1103/physreve.69.026113`.

29 L. O. Prokhorenkova, A. Raigorodskii, and P. Pralat. Modularity of complex networks models. *Internet Mathematics*, 2017.

30 K. Rybarczyk and M. Sulkowska. New bounds on the modularity of $G(n, p)$. *arXiv preprint arXiv:2504.16254*, 2025.

31 K. Rybarczyk and M. Sulkowska. Modularity of Preferential Attachment Graphs. In *43rd International Symposium on Theoretical Aspects of Computer Science (STACS 2026)*, pages 76:1–76:19, 2026. `doi:10.4230/LIPIcs.STACS.2026.76`.

32 S. Sahni. Computationally related problems. *SIAM Journal on Computing*, 3(4):262–279, 1974.

33 V. A. Traag, L. Waltman, and N. J. Van Eck. From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*, 9(1):1–12, 2019.

34 P.-W. Wang and J.Z. Kolter. Community detection using fast low-cardinality semidefinite programming. *Advances in Neural Information Processing Systems*, 33:3374–3385, 2020.

35 Y. Zhao, E. Levina, and J. Zhu. Consistency of community detection in networks under degree-corrected stochastic block models. *The Annals of Statistics*, 40(4):2266, 2012.

A A guide to the proofs given in the appendix

A.1 Main results: OGP for SBM and slow-mixing

The main results in the paper are that (a) for three or more communities, the stochastic block model (SBM) exhibits the overlap gap property (OGP); and (b) failure of a natural MCMC algorithm.

For (a), that SBM exhibits OGP, the statement is Theorem 2.2, and is proven in Appendix D. Proving Theorem 2.2 requires understanding of the maximum modularity of partitions within distance $d \leq 1/k$ of the planted partition (Theorem 2.6). Indeed, proving Theorem 2.6 is the main work of the paper, and we outline the steps in Appendix A.2.

For (b), the failure of a natural Markov Chain Monte Carlo (MCMC) algorithm, the statement is Theorem B.2 (an informal version, Theorem 2.3, appeared in the body of the paper). This algorithm is described in Appendix B, and we show in Proposition B.1 that it will likely output a partition within distance $o(1)$ of the planted partition $\mathcal{P}$. However, the main result is a negative one: Theorem B.2 shows that the time taken is exponentially large in n. In Appendix B, we state and prove both Theorem B.2 and Proposition B.1, assuming the OGP result stated in Theorem 2.2 and a bound on the maximum modularity of partitions at distance $d > 1/k$ of the planted partition (Lemma 5.2).

A.2 Putting it all together: Proving Theorem 2.6 (which proves Theorem 2.2)

We outline the proof of Theorem 2.6; in terms of the three ingredients discussed in the main body of the paper: distances to the planted partition (Section 3), concentration (Section 4) and optimisation (Section 5).

An important definition was that of the signature X of a partition $\mathcal{A}$; see (7). The signature of $\mathcal{A}$ determines the distance between $\mathcal{A}$ and $\mathcal{P}$. This was established in Section 3, with all proofs contained within that section. The second and third ingredients, namely, concentration and optimisation, were more involved. The optimisation results are included in Appendix C, for the concentration results see the full version of the paper.

todo

A.3 Ingredient: Concentration

We just saw that the distance of partition $\mathcal{A}$ to the planted partition is determined by the signature of X. Moreover, the signature (up to a small error) defines the modularity score $q_{\mathcal{A}}(G)$ of $\mathcal{A}$ for $G \in \mathcal{G}(n,k,p,q)$ generated by the SBM. This is established by the concentration results in Section 4 which show that for $\mathcal{A}$ with signature X, the modularity score $q_{\mathcal{A}}(G)$ is very close to $\frac{a-b}{a+(k-1)b}\,\frac{1}{k^2}\,g(X)$. See (10) for a definition of $g(X)$. The proofs for the concentration results may be found in the full version of the paper.

todo!

A.4 Ingredient: Optimisation

We are interested in the maximum modularity at a given distance from the planted partition, and the concentration results tell us, loosely, that the modularity score $q_{\mathcal{A}}(G)$ is quite close to $\frac{a-b}{a+(k-1)b}\,\frac{1}{k^2}\,g(X)$, where X is the signature of $\mathcal{A}$. Hence, the remaining challenge is to understand the maximal value of $g(X)$ over signatures X of partitions at a given distance to the planted partition $\mathcal{P}$.

We recall the setup for our optimisation problem. We have defined $\mathcal{X}_k(t)$ to be the family of matrices $X = (x_{ij})_{1\le i\le j\le k}$ such that (11) holds. Recall from (10) that we have defined $g(X)$ to be $g(X) = \sum_{i=1}^{k} \sum_{1\le j<j'\le k}(x_{ij} - x_{ij'})^2$.

The two main optimisation results are to (a) understand $\max_{X\in\mathcal{X}_k(t)} g(X)$ for $t \in [0,1]$ (corresponding to $d \le 1/k$); and (b) upper bound $\max_{X\in\mathcal{X}_k(t)} g(X)$ for $t > 1$ (corresponding to $d > 1/k$). For part (a), the result is stated in Lemma 5.1, and the proof is in Appendix C.1. For part (b), the result is stated in Lemma 5.2, and the proof is in Appendix C.2.

A.5 Auxiliary result : Theorem 2.5

A result of Bickel and Chen [4] states that the modularity-optimal partition is likely to be within small distance of the planted partition. We extend this to say that partitions with modularity score very close to optimal are within a small distance of the planted partition. See Theorem 2.5 for details, and the full version of the paper for proofs. todo

A.6 Outline

We briefly outline the following sections. Appendix B proves the failure of MCMC algorithms, and Appendix C proves the optimisation results. After these sections, we prove our main result, that modularity in SBM has OGP, in Appendix D. See the full version of the paper for the proofs of the auxiliary result Theorem 2.5, and the proof Proposition 6.1 which concerns the setting in which we restrict to optimising only over balanced partitions. todo! todo

B Failure of the Markov Chain Monte Carlo algorithm

Throughout this section we consider k-partitions with k fixed. We denote $q_{\mathcal{A}} = q_{\mathcal{A}}(G)$, $q^\star = q^\star(G)$ for brevity. The OGP immediately implies the failure of a natural Greedy algorithm for finding the planted partition $\mathcal{P}$ when started from the decoy partition $\mathcal{D}$. More specifically, Greedy is an algorithm resulting in a sequence of partitions $\mathcal{A}_t$ built as follows: We initialise the Greedy algorithm in the decoy partition, i.e., $\mathcal{A}_0 = \mathcal{D}$. Given $\mathcal{A}_{t-1}$, the partition $\mathcal{A}_t$ is obtained by changing the membership of at most one node $u \in [n]$ such that the resulting modularity strictly increases. That is, $\mathcal{A}_t$ is any partition satisfying $d(\mathcal{A}_t, \mathcal{A}_{t-1}) = 1/n$ and $q_{\mathcal{A}_t}(G) > q_{\mathcal{A}_{t-1}}(G)$. If no such node u exists, then the algorithm stops and outputs the partition obtained in the final step. By the OGP, the Greedy algorithm, initiated at the decoy partition $\mathcal{D}$, terminates at this partition with distance at least $1/k$ from the ground truth.

A natural generalisation of the Greedy algorithm, which is guaranteed (as we will show) to output a partition approximately matching the planted partition $\mathcal{P}$, is the well-known Markov Chain Monte Carlo (MCMC) algorithm which we now describe. Our main result in this section is the proof of slow mixing of this MCMC. In particular, we will show that the time it takes for the algorithm to approximately produce an planted partition is exponentially large in n.

We begin by describing the MCMC algorithm. A parameter usually called the *inverse temperature* β is fixed. The algorithm proceeds as a Markov chain moving according to the following rules. Given any partition $\mathcal{A}$, consider any partition $\mathcal{A}'$ with $d(\mathcal{A}, \mathcal{A}') = 1/n$. Then the algorithm moves from $\mathcal{A}$ to $\mathcal{A}'$ with probability proportional to $\exp(\beta n q_{\mathcal{A}'})$, namely with probability

$$\mathbb{P}(\mathcal{A} \to \mathcal{A}') \triangleq \frac{\exp(\beta n q_{\mathcal{A}'})}{\sum_{\mathcal{B}} \exp(\beta n q_{\mathcal{B}})},$$

where here the sum is over all partitions $\mathcal{B}$ with $d(\mathcal{A},\mathcal{B}) = 1/n$. (Note that we choose this parametrisation so that the stationary distribution is informative.) It is known that the unique stationary distribution of this chain is the so-called *Gibbs distribution* given by

$$\mathbb{P}_{\text{Gibbs}}(\mathcal{A}) = \frac{\exp(\beta n q_{\mathcal{A}})}{Z},$$

where $Z = \sum_{\mathcal{B}} \exp(\beta n q_{\mathcal{B}})$ is the normalising constant, which is also called the *partition function*, and now the sum is over *all* partitions $\mathcal{B}$.

We first show that this algorithm is sound, in the sense that for large enough β it produces partitions that are close to the ground truth $\mathcal{P}$.

▶ **Proposition B.1.** *For every $\varepsilon > 0, \zeta > 0$, there exists large enough $\beta > 0$ such that*

$$\sum_{\mathcal{A}\colon d(\mathcal{A},\mathcal{P})\leq\zeta} \mathbb{P}_{\text{Gibbs}}(\mathcal{A}) \geq 1-\varepsilon.$$

In words, the probability mass of partitions with distance at most ζ from the planted partition constitutes at least a $1-\varepsilon$ fraction of the total probability mass. Thus, if the algorithm were to run until stationarity, then a partition sampled according to the Gibbs distribution is likely to be at most ζ close to the ground truth, modulo an at most ε likelihood event.

Proof of Proposition B.1. Given any $\zeta > 0$ let $\mathcal{E}(\zeta) = \{\mathcal{A}\colon d(\mathcal{A},\mathcal{P}) \leq \zeta\}$. Fix $\varepsilon, \zeta > 0$. By Theorem 2.5, we can find $\delta > 0$ small enough so that, with probability at least $1-\varepsilon/2$, the event occurs that every $\mathcal{A}$ satisfying $q_{\mathcal{A}} \geq q^\star - \delta$ satisfies $\mathcal{A} \in \mathcal{E}(\zeta)$. On this event,

$$\sum_{\mathcal{A}\notin\mathcal{E}(\zeta)} \exp\left(\beta n q_{\mathcal{A}}\right) \leq k^n \exp(\beta n(q^\star - \delta)).$$

Fix β large enough so that $\log k - \beta\delta \leq -1$ (any strictly negative constant will do). Then, this sum is at most $\exp(-n)\exp(\beta n q^\star)$, so that

$$\begin{aligned}\sum_{\mathcal{A}\notin\mathcal{E}(\zeta)} \mathbb{P}_{\text{Gibbs}}(\mathcal{A}) &\leq \frac{\exp(-n)\exp(\beta n q^\star)}{Z}\\ &\leq \frac{\exp(-n)\exp(\beta n q^\star)}{\exp(\beta n q^\star)}\\ &= \exp(-n) \leq \varepsilon/2,\end{aligned}$$

for large enough n. Combining this event with the complement event, which occurs with probability at most $\varepsilon/2$, we complete the proof. ◀

Our main result, described next, shows that unfortunately the time to stationarity is exponentially large in n when the chain is initiated at a large enough distance from the partition, appropriately defined. Worse than that, the time to reach even *one* partition close to the ground truth is exponentially large in n, again when the chain is initiated at a large distance from the ground truth. This means that the set of such starting points acts as a *metastable set.* See [7] for a discussion on metastability, and [12] for an example of metastability in random regular graphs.

In preparation for the proof of this claim, we recap some of the properties implied by the OGP and an auxiliary lemma: there exist $0 < \nu_1 < \nu_2 < 1$ and $c_1, c_2 > 0$ such that the following holds. Let

$$\begin{aligned}\mathcal{E}_{\text{close}} &= \{\mathcal{A}\colon d(\mathcal{P},\mathcal{A}) \leq \nu_1\},\\ \mathcal{E}_{\text{far}} &= \{\mathcal{A}\colon d(\mathcal{P},\mathcal{A}) \geq \nu_2\},\\ \mathcal{E}_{\text{btw}} &= \{\mathcal{A}\colon d(\mathcal{P},\mathcal{A}) \in (\nu_1,\nu_2)\}.\end{aligned}$$

Then, by Theorem 2.2 and Lemma 5.2 with probability at least $1-\varepsilon$,

$$\arg\max q_{\mathcal{A}} \subseteq \mathcal{E}_{\text{close}}, \tag{12}$$

$$\max_{\mathcal{A}:\mathcal{A}\in\mathcal{E}_{\text{far}}} q_{\mathcal{A}} \geq q^{\star} - c_1, \tag{13}$$

$$\max_{\mathcal{A}:\mathcal{A}\in\mathcal{E}_{\text{btw}}} q_{\mathcal{A}} \leq q^{\star} - c_1 - c_2. \tag{14}$$

We now state our lower bound result.

▶ **Theorem B.2.** *Consider the MCMC initiated at the Gibbs distribution conditioned on being in $\mathcal{E}_{\text{far}}$. Namely, suppose*

$$\mathbb{P}\left(\mathcal{A}_0 = \mathcal{A}\right) = \frac{\mathbb{P}_{\text{Gibbs}}(\mathcal{A})}{\sum_{\mathcal{B}\in\mathcal{E}_{\text{far}}} \mathbb{P}_{\text{Gibbs}}(\mathcal{B})},$$

for all $\mathcal{A} \in \mathcal{E}_{\text{far}}$, and $\mathbb{P}\left(\mathcal{A}_0 = \mathcal{A}\right) = 0$ otherwise. Let $\tau = \min\{t\colon \mathcal{A}_t \in \mathcal{E}_{\text{close}}\}$. There exist $c_3, c_4 > 0$ such that with probability at least $1-\varepsilon$ (with respect to the randomness of the graph)

$$\mathbb{P}\left(\tau \geq \exp(c_3 n)\right) \geq 1 - \exp(-c_4 n),$$

where the probability is with respect to the random choices of the MCMC.

Proof. We assume that the events (12)-(14) hold, which is the case with probability at least $1-\varepsilon$. We note that the events $\mathcal{A}_0 \in \mathcal{E}_{\text{far}}, \mathcal{A}_\tau \in \mathcal{E}_{\text{close}}$ imply the existence of $s < \tau$ such that $\mathcal{A}_s \in \mathcal{E}_{\text{btw}}$. For every positive integer t,

$$\begin{aligned}\mathbb{P}\left(\mathcal{A}_t \in \mathcal{E}_{\text{btw}} | \mathcal{A}_0 \in \mathcal{E}_{\text{far}}\right) &= \frac{\mathbb{P}\left(\mathcal{A}_t \in \mathcal{E}_{\text{btw}}, \mathcal{A}_0 \in \mathcal{E}_{\text{far}}\right)}{\mathbb{P}\left(\mathcal{A}_0 \in \mathcal{E}_{\text{far}}\right)} \\ &\leq \frac{\mathbb{P}\left(\mathcal{A}_t \in \mathcal{E}_{\text{btw}}\right)}{\mathbb{P}\left(\mathcal{A}_0 \in \mathcal{E}_{\text{far}}\right)}.\end{aligned}$$

We have

$$\begin{aligned}&\mathbb{P}\left(\mathcal{A}_t \in \mathcal{E}_{\text{btw}}\right) \leq k^n \exp(\beta n(q^{\star} - c_1 - c_2)), \\ &\mathbb{P}\left(\mathcal{A}_t \in \mathcal{E}_{\text{far}}\right) \geq \exp(\beta n(q^{\star} - c_1)).\end{aligned}$$

Thus the ratio is at most $k^n \exp(-\beta n c_2)$. Assuming β is large enough so that $\log k - \beta c_2 \leq -1$ (again any negative constant suffices), this ratio is at most $\exp(-n)$. By the union bound, we obtain that $\mathbb{P}(\tau \leq \exp(n/2)) \leq \mathrm{e}^{n/2}\mathrm{e}^{-n} = \mathrm{e}^{-n/2}$, completing the proof. ◀

C Optimisation lemmas for modularity at a given distance

In this section we prove that the function $g(X)$ has the required behaviour. Recall that by Theorem 2.6 with probability close to one, the modularity of any partition $\mathcal{A}$ with signature X can be well approximated by $\frac{a-b}{a+(k-1)b}\,\frac{1}{k^2}g(X)$.

In particular, we are interested in the maximum modularity score for partitions at a given distance d; and this we can control by understanding the maximum of $g(X)$ for signatures X at distance d; or, equivalently, the maximum of $\frac{1}{k^2}\,g(X)$, which we call $\tilde{h}(d)$. (See also the discussion after Theorem 2.6.) This will correspond to the maximisation problem $\tilde{h}(d) = \frac{1}{k^2}\max_{X\in\mathcal{X}_k(dk)} g(X)$ for the k-block SBM, and re-scaled distance $t = dk$.

These results naturally split into two. For $d < 1/k$, we solve this optimisation in Lemma 5.1, while, for $d > 1/k$, we show in Lemma 5.2 that $h(d) < h(1/k)$, i.e., that the value of g is bounded above by its maximum at $d = 1/k$.

C.1 Proof of the general case for distance $d < 1/k$

Proof of Lemma 5.1. Let $t = dk$ and consider first $0 < t < 1$. We will prove the result by successively eliminating the variables.

We begin by giving an equivalent optimisation problem with $k(k-1)$ variables. Since the columns of the matrix X sum to 1, we may reduce the number of variables by writing $x_{ii} = 1 - \sum_{i':i'\neq i} x_{i'i}$, and maximising the resulting $\tilde{g}(x_{12}, \dots, x_{k-1,\,k})$ obtained from $g(X)$ by making these substitutions for x_{ii}.

In particular, let

$$\widetilde{\mathcal{X}} = \big\{(x_{ij})_{i\neq j} \ : \ x_{ij} \geq 0 \ \forall i \neq j, \ \sum_{i:\ i\neq j} x_{ij} \leq 1 \ \forall j \quad \text{and} \quad \sum_{i\neq j} x_{ij} = t\big\}. \tag{15}$$

Thus it will suffice to show $\max_{\underline{x}\in\widetilde{\mathcal{X}}} \tilde{g}(\underline{x})$ has the bounds claimed. (We write $\underline{x}$, since we have now a subset of variables of the matrix X.)

Notice that if we have equality for $x_{ij} + x_{i'j} = 1$ for some i, i', j all distinct, then $\sum_{i\neq j} x_{ij} \geq 1$, which yields a contradiction since $t < 1$. Thus, it is equivalent to consider the maximisation problem over $\widetilde{\mathcal{X}}_1$, where we require strict inequality for the terms, i.e., $x_{ij} + x_{i'j} < 1$. Let

$$\widetilde{\mathcal{X}}_1 = \big\{(x_{ij})_{i\neq j} \ : \ x_{ij} \geq 0 \ \forall i \neq j, \ \sum_{i:\ i\neq j} x_{ij} < 1 \ \forall j \quad \text{and} \quad \sum_{i\neq j} x_{ij} = t\big\}. \tag{16}$$

It now suffices to show that $\max_{\underline{x}\in\widetilde{\mathcal{X}}_1} \tilde{g}(\underline{x})$ satisfies the claimed bounds.

We may remove one more variable. Note from the constraints that we have

$$x_{21} = t - \sum_{\substack{i\neq j \\ (i,j)\neq(2,1)}} x_{ij}.$$

Let $\tilde{g}_2$ be $\tilde{g}$ making this substitution (i.e., $\tilde{g}_2$ is a function on $k^2 - k - 1$ variables), and let

$$\begin{aligned} \widetilde{\mathcal{X}}_2 = \big\{(x_{ij})_{i\neq j,\ (i,j)\neq(2,1)} \quad : \quad & x_{ij} \geq 0 \ \forall (i,j) \text{ with } i \neq j, (i,j) \neq (1,2), \ \sum_{i:\ i\geq 3} x_{i1} < 1 \\ & \sum_{i:\ i\neq j} x_{ij} < 1 \ \forall j \geq 2 \quad \text{and} \quad \sum_{\substack{i\neq j \\ (i,j)\neq(2,1)}} x_{ij} \leq t\big\}. \end{aligned} \tag{17}$$

Thus, finally, it suffices to show that $\max_{\underline{x}\in\widetilde{\mathcal{X}}_2} \tilde{g}_2(\underline{x})$ satisfies the claimed bounds.

Recall that the maximum of a convex function is obtained on the boundary. Note also that we may apply this recursively, since if we set any inequality in (17) to equality, then we may make a substitution to reduce the number of variables by 1, and we obtain a maximisation problem for a convex function (on one fewer variables). Thus, the max of $\tilde{g}_2(\underline{x})$ over $\underline{x} \in \widetilde{\mathcal{X}}_2$ is obtained when at least $k^2 - k - 1$ of the inequalities in (17) are equality.

Observe that there are $k^2 - k - 1$ inequalities in (17) (which are not strict inequalities), so the maximum is obtained when we set all but (at most) one of these to equality.

We consider two cases. The first case is that we have $x_{ij} = 0$ for all (i,j) with $i \neq j$, $(i,j) \neq (2,1)$. Notice that this fixes the value of *all* variables, and that the remaining inequalities in (17) are all satisfied. The value of $\tilde{g}_2(x)$ attained is that of $g(X)$ with $x_{21} = t$ (and $x_{ij} = 0$ for all (i,j) with $i \neq j$, $(i,j) \neq (2,1)$, and $x_{ii} = \sum_{i':\ i'\neq i} x_{i'i}$ for all i). Note that this yields the matrix $X^{(2,1)}(t)$, and, moreover, that $g(X^{(2,1)}(t)) = k(k-1) - 2t\big(k - t(k-1)\big)$.

The second case is that there exist a, b (where $(a,b) \neq (2,1)$) such that

$$\sum_{\substack{i\neq j \\ (i,j)\neq(2,1)}} x_{ij} = t \quad \text{and} \quad x_{ij} = 0 \text{ for all } (i,j) \text{ with } i \neq j, \ (i,j) \neq (2,1), (a,b).$$

Notice that this fixes the value of all $k^2 - k - 1$ variables: $x_{ab} = t$ and all remaining variables take the value 0, and that this assignment satisfies the remaining inequalities in (17). The value of $\tilde{g}_2(x)$, attained is that of $g(X)$ with $x_{ab} = t$ (and $x_{ij} = 0$ for all (i, j) with $i \neq j$, $(i, j) \neq (a, b)$, and $x_{ii} = \sum_{i':i' \neq i} x_{i'i}$ for all i). Note that this yields the matrix $X^{(a,b)}(t)$, and, moreover, that again $g(X^{(a,b)}(t)) = k(k-1) - 2t\big(k - t(k-1)\big)$ for any such (a, b). This completes the proof. ◀

C.2 Proof for the general case $d > 1/k$

Here we prove Lemma 5.2, which will allow us to infer that partitions $\mathcal{A}$ with high modularity scores must be at distance $d \leq 1/k$ from the planted partition.

Proof. Consider any partition $\mathcal{A}$ with signature $X = (x_{ij})_{1 \leq i \leq j \leq k}$ that is at distance more than $1/k$ from the planted partition $\mathcal{P}$. Our goal is to show that $g(X) \leq k(k-1) - 2$.

Suppose that there exists a column $j \in [k]$ (say, $j = 1$) in X with at least two non-zero entries (say, $x_{11} > 0$ and $x_{21} > 0$). Consider a family of signatures $X(s)$ that is parametrised by variable s, as follows: for a given $s \in [-x_{11}, x_{21}]$, matrix $X(s)$ is exactly the same as X but x_{11} is replaced with $x_{11} + s$ and x_{21} is replaced with $x_{21} - s$. Of course, $X = X(0)$.

Note that $g(X(s))$ is a quadratic function of s with a positive coefficient in front of s^2. As a result, at least one of the following two properties holds: (i) $g(X(s))$ increases as s increases from $s = 0$ to $s = x_{21}$ (which is equivalent to transferring a weight from x_{21} to x_{11}); or (ii) $g(X(s))$ increases as s decreases from $s = 0$ to $s = -x_{11}$ (which is equivalent to transferring a weight from x_{11} to x_{21}). Note that it might be the case that $g(X(s))$ attains its local minimum at $s = 0$ and both properties hold at the same time. Regardless, we may start from $s = 0$ and either increase or decrease s to gradually increase $g(X(s))$. While we do this, the distance to the planted partition $\mathcal{P}$ does not need to behave monotonically. However, it is easy to see (but it is crucial for the argument) that the distance to $\mathcal{P}$ is a continuous function of s. Once we are done with transferring the weight, we get an additional zero entry in our signature matrix, and we can move on to the next pair of non-zero entries, possibly in a different column j.

We need to consider two cases now. Suppose first that during the above process the distance to $\mathcal{P}$ is equal to $1/k$. We prematurely stop the process at the very first time this happens, and let Y be the signature we have at that moment. Since we kept increasing the function g along the way, it follows from Lemma 5.1 (applied with $t = 1$) that

$$g(X) < g(Y) \leq \max_{X \in \mathcal{X}_k(1)} g(X) = k(k-1) - 2,$$

which finishes the proof of the lemma in this case.

Suppose now that the distance from $\mathcal{A}$ to $\mathcal{P}$ was always more than $1/k$, but we had to stop the above process of transferring weights at some point because each column had exactly one non-zero entry (of course these non-zero entries must be equal to one). This means that each part of $\mathcal{A}$ is a union of some planted parts, and the distance between $\mathcal{A}$ and $\mathcal{P}$ is equal to i/k for some integer $i \geq 2$.

Now, take any part which is a union of $r \geq 2$ planted parts and split it into two parts, consisting of $r - 1$ and a single planted part, respectively. After such partition refinement, the function g increases. Indeed, before splitting the part with r planted parts, the contribution of that part to the function g is $r(k-r)$, but after splitting, the two resulting parts contribute $(r-1)(k-r+1) + (k-1)$, which is equal to $r(k-r) + 2(r-1) > r(k-r)$. We continue such refinements of partitions until the distance to $\mathcal{P}$ is $1/k$. At that point, the partition $\mathcal{A}$

with signature Y has one part consisting of two planted parts, while the remaining parts consist of just one planted part. The conclusion is as before:

$$g(X) < g(Y) = \max_{X \in \mathcal{X}_k(1)} g(X) = k(k-1) - 2.$$

This finishes the proof of Lemma 5.2. ◀

D Proof of Theorem 2.2 : SBM has OGP

Proof of Theorem 2.2. We will prove Theorem 2.2 using Theorem 2.6, our concentration result in Lemma 4.1, and the optimisation result in Lemma 5.2.

First, we will set up the range for the distances. Recall that $h(d)$ is a quadratic function of d, with minimum at $\frac{1}{2(k-1)}$. See also Figure 2. Take any $\nu \in \left(\frac{1}{2(k-1)}, \frac{1}{k}\right)$, choose $\nu' = \frac{2}{3} \cdot \nu + \frac{1}{3} \cdot \frac{1}{k}$, and let $\nu'' \in \left[0, \frac{1}{2(k-1)}\right)$ be the unique number such that $h(\nu'') = h(\nu')$. Let n_0 be the smallest integer such that $1/\sqrt{n_0} < \nu' - \nu < \frac{1}{2(k-1)} - \nu''$.

Next, we define $\mu = \frac{a-b}{a+(k-1)b}\big(h(0) - h(\nu')\big)$. Choose a positive $\varepsilon < \frac{a-b}{a+(k-1)b}\left(h(\nu') - h(\nu)\right)$. Then, by Theorem 2.6, we obtain that parts (i) and (ii) below hold with probability at least $1 - \varepsilon$, and for $n > n_0$ large enough:

(i) For all partitions $\mathcal{A}$ at distance $d \in [0, \nu' - \nu)$ and at distance $d \in (1/k - (\nu' - \nu), 1/k]$, it holds that

$$q_{\mathcal{A}}(G) > \frac{a-b}{a+(k-1)b}\, h(d) - \varepsilon > \frac{a-b}{a+(k-1)b}\, h(\nu') = \frac{a-b}{a+(k-1)b}\, h(0) - \mu.$$

(ii) For any partition $\mathcal{A}$ at distance $d \in \left(\frac{k}{2(k-1)}, \nu\right) \subset \left[0, \frac{1}{k}\right]$,

$$q_{\mathcal{A}}(G) < \frac{a-b}{a+(k-1)b}\, h(d) + \varepsilon < \frac{a-b}{a+(k-1)b}\, h(\nu') = \frac{a-b}{a+(k-1)b}\, h(0) - \mu.$$

Together, (i) and (ii) establish Theorem 2.2. ◀