Electronic Journal of Probability

Electron. J. Probab. **21** (2016), no. 21, 1–24.
ISSN: 1083-6489 DOI: 10.1214/16-EJP4185

Consistency thresholds for the planted bisection model

Elchanan Mossel* Joe Neeman† Allan Sly‡

Abstract

The planted bisection model is a random graph model in which the nodes are divided into two equal-sized communities and then edges are added randomly in a way that depends on the community membership. We establish necessary and sufficient conditions for the asymptotic recoverability of the planted bisection in this model. When the bisection is asymptotically recoverable, we give an efficient algorithm that successfully recovers it. We also show that the planted bisection is recoverable asymptotically if and only if with high probability every node belongs to the same community as the majority of its neighbors.

Our algorithm for finding the planted bisection runs in time almost linear in the number of edges. It has three stages: spectral clustering to compute an initial guess, a "replica" stage to get almost every vertex correct, and then some simple local moves to finish the job. An independent work by Abbe, Bandeira, and Hall establishes similar (slightly weaker) results but only in the case of logarithmic average degree.

Keywords: stochastic block model; planted partition model; consistency; threshold; phase transition; community detection; random network.
AMS MSC 2010: 05C80.
Submitted to EJP on March 12, 2015, final version accepted on January 27, 2016.
Supersedes arXiv:1407.1591.

1 Introduction

The "planted bisection model" is a random graph model with $2n$ vertices that are divided into two classes with n vertices each. Edges within the classes are added to the graph independently with probability p_n each, while edges between the classes are added with probability q_n. Following Bui et al, [5] who studied a related model, Dyer and Frieze [9] introduced the planted bisection model in order to study the average-case complexity of the Min-Bisection problem, which asks for a bisection of a graph that cuts

*U.C. Berkeley and the University of Pennsylvania. Supported by NSF grants DMS-1106999 and CCF 1320105 and DOD ONR grant N000141110140. E-mail: mossel@stat.berkeley.edu

†U.T. Austin and the University of Bonn. Supported by NSF grant DMS-1106999 and DOD ONR grant N000141110140. E-mail: joeneeman@gmail.com

‡U.C. Berkeley and the Australian National University. Supported by an Alfred Sloan Fellowship and NSF grant DMS-1208338. E-mail: sly@stat.berkeley.edu

the smallest possible number of edges. This problem is known to be NP-complete in the worst case [14], but on a random graph model with a "planted" small bisection one might hope that it is usually easy. Indeed, Dyer and Frieze showed that if $p_n = p > q = q_n$ are fixed as $n \to \infty$ then with high probability the bisection that separates the two classes is the minimum bisection, and it can be found in expected $O(n^3)$ time.

These models were introduced slightly earlier in the statistics literature [12] (under the name "stochastic block model") in order to study the problem of community detection in random graphs. Here, the two parts of the bisection are interpreted as latent "communities" in a network, and the goal is to identify them from the observed graph structure. If $p_n > q_n$, the maximum a posteriori estimate of the true communities is exactly the same as the minimum bisection (see the discussion leading to Lemma 4.1), and so the community detection problem on a stochastic block model is exactly the same as the Min-Bisection problem on a planted bisection model; hence, we will use the statistical and computer science terminologies interchangeably. We note, however, the statistics literature is slightly more general, in the sense that it often allows $q_n > p_n$, and sometimes relaxes the problem by allowing the detected communities to contain some errors.

Our main contribution is a necessary and sufficient condition on p_n and q_n for recoverability of the planted bisection. When the bisection can be recovered, we provide an efficient algorithm for doing so.

2 Definitions and results

Definition 2.1 (Planted bisection model)**.** *Given $n \in \mathbb{N}$ and $p, q \in [0,1]$, we define the random $2n$-node labelled graph $(G,\sigma) \sim \mathcal{G}(2n,p,q)$ as follows: first, choose a balanced labelling σ uniformly at random from $\{\tau \in \{1,-1\}^{V(G)} : \sum_u \tau_u = 0\}$. Then, for every distinct pair $u, v \in V(G)$ independently, add an edge between u and v with probability p if $\sigma_u = \sigma_v$, and with probability q if $\sigma_u \neq \sigma_v$.*

The oldest and most fundamental question about planted partition models is the label reconstruction problem: if we were given the graph G but not the labelling σ, could we reconstruct σ (up to its sign) from G? This problem is usually framed in the asymptotic regime, where the number of nodes $n \to \infty$, and p and q are allowed to depend on n.

Definition 2.2 (Strong consistency)**.** *Given sequences p_n and q_n in $[0,1]$, and given a map $\mathcal{A}$ from graphs to vertex labellings, we say that $\mathcal{A}$ is* strongly consistent *(or sometimes just* consistent*) if*

$$\Pr_n(\mathcal{A}(G) = \sigma \text{ or } \mathcal{A}(G) = -\sigma) \to 1,$$

where the probability $\Pr_n$ is taken with respect to $(G,\sigma) \sim \mathcal{G}(2n,p_n,q_n)$.

Depending on the application, it may also make sense to ask for a labelling which is almost completely accurate, in the sense that it correctly labels all but a vanishingly small fraction of nodes. Amini et al. [2] suggested the term "weak consistency" for this notion.

Definition 2.3 (Weak consistency)**.** *Given $\sigma, \tau \in \{1,-1\}^{2n}$, define*

$$\Delta(\sigma,\tau) = 1 - \frac{1}{2n}\left|\sum_{i=1}^{2n} \sigma_i \tau_i\right|.$$

Given sequences p_n and q_n in $[0,1]$, and given a map $\mathcal{A}$ from graphs to vertex labellings, we say that $\mathcal{A}$ is weakly consistent *if*

$$\Delta(\sigma, \mathcal{A}(G)) \xrightarrow{P} 0,$$

where "$\xrightarrow{P}$" means convergence in probability, and the probability is taken with respect to $(G, \sigma) \sim \mathcal{G}(2n, p_n, q_n)$.

Our main result is a characterization of the sequences p_n and q_n for which consistent or weakly consistent estimators exist. Note that the characterization of weak consistency was obtained previously by Yun and Proutiere [27], but we include it here for completeness.

Definition 2.4. *Given m, n, p, and q, let*

$$X \sim \mathrm{Binom}(m, \max\{p, q\})$$
$$Y \sim \mathrm{Binom}(n, \min\{p, q\}).$$

We define

$$P(m, n, p, q) = \Pr(Y \geq X).$$

When $m = n$, we will abbreviate by $P(n, p, q) = P(n, n, p, q)$.

Theorem 2.5 (Characterization of consistency)**.** *Consider sequences p_n and q_n in $[0, 1]$. There exists a strongly consistent estimator for $\mathcal{G}(2n, p_n, q_n)$ if and only if $P(n, p_n, q_n) = o(n^{-1})$. There exists a weakly consistent estimator for $\mathcal{G}(2n, p_n, q_n)$ if and only if $P(n, p_n, q_n) \to 0$.*

In order to provide some intuition for Definition 2.4 and its appearance in our characterization, we note the following graph-theoretic interpretation of $P(n, p, q)$:

Definition 2.6. *Given a labelled graph $(G, \sigma) \sim \mathcal{G}(2n, p, q)$ and a node $v \in V(G)$, we say that v has a majority of size k if either*

$$p > q \text{ and } \#\{u \sim v : \sigma_u = \sigma_v\} \geq \#\{u \sim v : \sigma_u \neq \sigma_v\} + k$$

or

$$p < q \text{ and } \#\{u \sim v : \sigma_u \neq \sigma_v\} \geq \#\{u \sim v : \sigma_u = \sigma_v\} + k.$$

We say that v has a majority if it has a majority of size one. If v does not have a majority, we say that it has a minority.

Proposition 2.7. *Fix sequences p_n and q_n in $[0, 1]$ and let $(G, \sigma) \sim \mathcal{G}(n, p_n, q_n)$. Then*

- *$P(n, p_n, q_n) = o(n^{-1})$ if and only if a.a.s. every $v \in V(G)$ has a majority; and*
- *$P(n, p_n, q_n) \to 0$ if and only if a.a.s. at most $o(n)$ nodes in $V(G)$ fail to have a majority.*

Proposition 2.7 suggests some intuition for Theorem 2.5: namely, that a node can be labelled correctly if and only if it has a majority. In fact, having a majority is necessary for correct labelling (and we will use this to prove one direction of Theorem 2.5); however, it is not sufficient. For example, there are regimes in which 51% of nodes have majorities, but only 50% of them can be correctly labelled (see [22]).

We note that Theorem 2.5 has certain parallels with local-to-global threshold phenomena in random graphs. For example, Erdős and Rényi showed [10] that for $\mathcal{G}(n, p_n)$, if p_n is large enough so that with high probability every node has a neighbor then the graph is connected with high probability. On the other hand, every node having a neighbor is clearly necessary for the graph to be connected. An analogous story holds for the existence of Hamiltonian cycles: Komlós and Szemerédi [15] showed that $\mathcal{G}(n, p_n)$ has a Hamiltonian cycle with high probability if and only if with high probability every node has degree at least two.

These results on connectedness and Hamiltonicity have a feature in common: in both cases, an obviously necessary local condition turns out to also be sufficient (on random graphs) for a global condition. One can interpret Theorem 2.5 similarly: the minimum bisection in $\mathcal{G}(n, p_n, q_n)$ equals the planted bisection with high probability if and only if with high probability every node has more neighbors of its own label than those of the other label.

2.1 The algorithm

In order to prove the positive direction of Theorem 2.5, we provide an algorithm that recovers the planted bisection with high probability whenever $P(n, p_n, q_n) = o(n^{-1})$. Moreover, this algorithm runs in time $\tilde{O}(n^2(p_n + q_n))$, where $\tilde{O}$ hides polylogarithmic factors. That is, it runs in time that is almost linear in the number of edges. In addition, we remark that the algorithm does not need to know p_n and q_n. For simplicity, we assume that we know whether $p_n > q_n$ or vice versa, but this can be checked easily from the data (for example, by checking the sign of the second-largest-in-absolute-value eigenvalue of the adjacency matrix; see Section 4.1).

Our algorithm comes in three steps, each of which is based on an idea that has already appeared in the literature. Our first step is a spectral algorithm, along the lines of those developed by Boppana [4], McSherry [20], and Coja-Oghlan [7]. Yun and Proutiere [27] recently made some improvements to (a special case of) Coja-Oghlan's work, showing that a spectral algorithm can find a bisection with $o(n)$ errors if $n\frac{(p_n-q_n)^2}{p_n+q_n} \to \infty$; this is substantially weaker than McSherry's condition for strong consistency, which would require converging to infinity with a rate of at least $\log n$.

The second stage of our algorithm is to apply a "replica trick." We hold out a small subset U of vertices and run a spectral algorithm on the subgraph induced by $V \setminus U$. Then we label vertices in U by examining the edges between U and $V \setminus U$. By repeating the process for many subsets U, we dramatically reduce the number of errors made by the spectral algorithm. More importantly, we get extra information about the structure of the errors; for example, we can show that the set of incorrectly-labelled vertices is very poorly connected. Similar ideas are used by Condon and Karp [8], who used successive augmentation to build an initial guess on a subset of vertices, and then used that guess to correctly classify the remaining vertices. The authors [21] also used a similar idea in the $p_n, q_n = \Theta(n^{-1})$ regime, with a more complicated replica trick based on belief propagation.

The third step of our algorithm is a hill-climbing algorithm, or a sequence of local improvements. We simply relabel vertices so that they agree with the majority of their neighbors. An iterative version of this procedure was considered in [6], and a randomized version (based on simulated annealing) was studied by Jerrum and Sorkin [13]. Our version has better performance guarantees because we begin our hill-climbing just below the summit: as we will show, we need to relabel only a tiny fraction of the vertices and each of those will be relabelled only once.

As noted above, none of the ingredients in our algorithm are novel on their own. However, the way that we combine them is new (and also crucial to the correctness of the resulting algorithm). For example, McSherry [20] used a spectral algorithm with a "clean-up" stage, but his clean-up stage was different from our second and third stages.

2.2 Formulas in terms of p_n and q_n

Although Theorem 2.5 is not particularly explicit in terms of p_n and q_n, one can obtain various explicit characterizations in particular regimes (for example, in order to better compare our results with the existing literature). We will focus our attention on the case where p_n and q_n are bounded away from one; for concreteness, suppose $p_n, q_n \le 2/3$. Because of the symmetry of the problem, this case suffices: indeed, replacing $G \sim \mathcal{G}(n, p_n, q_n)$ by its complement (the graph in which two vertices are connected if they are not connected in G) corresponds to replacing p_n by $1 - p_n$ and q_n by $1 - q_n$. Hence, if we handle the case $p_n, q_n \leq 2/3$ then we also handle the case $p_n, q_n \geq 1/3$. There remains the case in which $\min\{p_n, q_n\} \leq 1/3$ and $2/3 \leq \max\{p_n, q_n\}$, but this case is trivial: $P(n, p_n, q_n)$ decreases exponentially fast in n, and even very simple algorithms are known to be strongly consistent.

One can easily see that to obtain strong consistency, at least one of p_n or q_n must be at least $n^{-1} \log n$ asymptotically. Indeed, suppose $q_n \leq p_n = n^{-1} \log n$ and let $X \sim \mathrm{Binom}(n, p_n)$, $Y \sim \mathrm{Binom}(n, q_n)$. Then $\Pr(X = 0) = \Theta(n^{-1})$, and so certainly $P(n, p_n, q_n) = \Pr(Y \geq X) = \Omega(n^{-1})$, which means that strong consistency is impossible for these parameters. However, strong consistency is possible for some other parameters in the range $\Theta(n^{-1} \log n)$. Using a Poisson approximation, we can characterize explicitly which of these sequences allow for strong consistency:

Proposition 2.8. *Let $p_n = a_n n^{-1} \log n$ and $q_n = b_n n^{-1} \log n$. If there is a constant C such that $C^{-1} \leq a_n, b_n \leq C$ for all but finitely many n then $P(n, p_n, q_n) = o(n^{-1})$ if and only if*

$$(a_n + b_n - 2\sqrt{a_n b_n} - 1) \log n + \frac{1}{2} \log \log n \to \infty.$$

In a denser regime, it is tempting to approximate $\mathrm{Binom}(n, p_n)$ and $\mathrm{Binom}(n, q_n)$ by the normal random variables $\mathcal{N}(np_n, n\sigma_p^2)$ and $\mathcal{N}(nq_n, n\sigma_q^2)$, where $\sigma_p = \sqrt{p(1-p)}$ and $\sigma_q = \sqrt{q(1-q)}$. That is,

$$\begin{aligned}
\Pr(Y \geq X) &\approx \Pr(\mathcal{N}(np_n, n\sigma_p^2) \geq \mathcal{N}(nq_n, n\sigma_q^2)) \\
&= \Pr(\sigma_p \mathcal{N}(0,1) \geq \sqrt{n}(q_n - p_n) + \sigma_q \mathcal{N}(0,1)) \\
&= \Pr(\mathcal{N}(0,1) \geq \sigma^{-1} \sqrt{n}(q_n - p_n)),
\end{aligned}$$

where $\sigma = \sqrt{\sigma_p^2 + \sigma_q^2}$. The central limit theorem implies that the normal approximation is correct in the bulk of the distribution if $np_n \to \infty$ and $nq_n \to \infty$. However, we are interested in applying this approximation for the tail, which requires a faster increase of np_n and a more delicate argument.

Proposition 2.9. *Suppose $p_n, q_n = \omega\left(n^{-1} \log^3 n\right)$ and $p_n, q_n \leq 2/3$. Then the following conditions are equivalent*

- $P(n, p_n, q_n) = o(1/n)$
- $n \Pr\left(\mathcal{N}(0,1) \geq \sigma_n^{-1} \sqrt{n}(p_n - q_n)\right) \to 0$
- $\frac{\sqrt{n}\sigma_n}{p_n - q_n} \exp(-\frac{n(p_n - q_n)^2}{2\sigma_n^2}) \to 0$,

where $\sigma_n = \sqrt{p_n(1-p_n) + q_n(1-q_n)}$.

In particular, the third condition in Proposition 2.9 gives an explicit formula for checking whether a strongly consistent estimator exists.

The formula for weak consistency is rather simpler:

Proposition 2.10. *$P(n, p_n, q_n) \to 0$ if and only if $\frac{n(p_n - q_n)^2}{p_n + q_n} \to \infty$.*

One direction of Proposition 2.10 follows from Chebyshev's inequality, while the other follows from the central limit theorem.

2.3 Relation to prior work

Over the years, various authors have improved on the seminal work of Dyer and Frieze [9] by proving weaker sufficient conditions on the sequences p_n and q_n for which the planted bisection can be recovered. (Various results also generalized the problem by allowing more than two labels, but we will ignore this generalization here.) For example, Jerrum and Sorkin [13] required $p_n - q_n = \Omega(n^{-1/6+\epsilon})$, while Condon and Karp improved this to $p_n - q_n = \Omega(n^{-1/2+\epsilon})$. McSherry [20] made a big step by showing that if

$$\frac{p_n - q_n}{p_n} \geq C\sqrt{\frac{\log n}{p_n n}}$$

for a large enough constant C then spectral methods can exactly recover the labels. This was significant because it allowed p_n and q_n to be as small as $\Theta(n^{-1}\log n)$, which is order-wise the smallest possible. A similar result for a slightly different random graph model had been claimed earlier by Boppana [4], but the proof was incomplete. Carson and Impagliazzo [6] showed that with slightly worse poly-logarithmic factors, a simple hill-climbing algorithm also works. Analogous results were later obtained by by Bickel and Chen [3] using modularity maximization (for which no efficient algorithm is known).

Until now, none of the sufficient conditions in the literature were also necessary; in fact, necessary conditions on p_n and q_n have only rarely been discussed. It is instructive to keep the example $p_n = 1/2$, $q_n = 1/2 - r_n$ in mind. In this case McSherry's condition is the same as requiring that $r_n \geq C\sqrt{n^{-1}\log n}$. On the other hand, Carson and Impagliazzo [6] pointed out that if $r_n \leq c\sqrt{n^{-1}\log n}$ for some small constant c then the minimum bisection no longer coincides with the planted bisection (as far as we are aware, this was the only necessary condition in the literature). From a statistical point of view, this means that the true communities can no longer be reconstructed perfectly. Our contribution closes the gap between McSherry's sufficient condition and Carson-Impagliazzo's necessary condition. In the above case, for example, Proposition 2.9 shows that the critical constant is $C = c = 1$.

2.4 Parallel independent work

Abbe et al. [1] independently studied the same problem in the logarithmic sparsity regime. They consider $p_n = (a \log n)/n$ and $q_n = (b \log n)/n$ for constants a and b; they show that $(a+b) - 2\sqrt{ab} > 1$ is sufficient for strong consistency and that $(a+b) - 2\sqrt{ab} \geq 1$ is necessary. Note that these are implied by Proposition 2.8, which is more precise. Abbe et al. also consider a semidefinite programming algorithm for recovering the labels; they show that it performs well under slightly stronger assumptions.

2.5 Other related work, and an open problem

Consistency is not the only interesting notion that one can study on the planted partition model. Earlier work by the authors [22, 23] and by Massoulié [19] considered a much weaker notion of recovery: they only asked whether one could find a labelling that was positively correlated with the true labels.

There are also model-free notions of consistency. Kumar and Kannan [16] considered a deterministic spatial clustering problem and showed that if every point is substantially closer to the center of its own cluster than it is to the center of the other cluster then one can exactly reconstruct the clusters. This is in much the same spirit as Theorem 2.5.

Makarychev, Makarychev, and Vijayaraghavan [17, 18] proposed semi-random models for planted bisections. These models allow for adversarial noise, and also allow edge distributions that are not independent, but only invariant under permutations. They then give approximation algorithms for Min-Bisection, which they prove to work under expansion conditions that hold with high probability for their semi-random model.

We ask whether the techniques developed here could sharpen the results obtained by Makarychev et al. For example, exact recovery under adversarial noise is clearly impossible, but if the adversary is restricted to adding $o(n)$ edges, then maybe one can guarantee almost exact recovery.

3 Binomial probabilities and graph structure

In this section, we will prove Proposition 2.7, which relates the binomial probabilities $P(n, p_n, q_n)$ to the structure of random graphs $G \sim \mathcal{G}(2n, p_n, q_n)$.

From now on, the letters c and C refer to positive constants, whose value may change from line to line. We adopt the convention that C refers to a "sufficiently large" constant, so that any statement involving C will remain true if C is replaced by a larger constant. Similarly, c refers to a "sufficiently small" constant.

3.1 Binomial perturbation estimates

We begin by stating some estimates on how binomial probabilities respond to perturbations, which we will prove in Section 6. For example, we will use the following proposition for two main applications: when $n = m$ and $\ell = (np)^{1/2} \log^{-1/2} n$, it can be used to get large majorities "for free," by implying that if every node has a majority a.a.s., then in fact every node has a majority of size $(np)^{1/2} \log^{-1/2} n$ a.a.s. On the other hand, we will also apply Proposition 3.1 with $m = n - 1$ and $\ell = 1$, which will be useful (later in this section) for showing that whether u has a majority is almost independent of whether v has a majority.

Proposition 3.1. *Let $X \sim \mathrm{Binom}(m, p)$ and $Y \sim \mathrm{Binom}(n, q)$, where $mp \ge 64 \log m$ and $p \le 2/3$. For any $1 \le \ell \le \sqrt{mp \log m}$,*

$$\Pr(Y \ge X + \ell) \ge \Pr(Y \ge X) e^{\left(-C\ell\sqrt{\frac{\log m}{mp}}\right)} - 2m^{-2} \tag{3.1}$$

$$\Pr(Y \ge X - \ell) \le \Pr(Y \ge X) e^{\left(C\ell\sqrt{\frac{\log m}{mp}}\right)} + 2m^{-2}, \tag{3.2}$$

where $C > 0$ is a universal constant.

Note that the condition $mp \ge 64 \log m$ is not only a technical one (although the constant 64 is certainly not optimal). For example, if $p = m^{-1} \log m$ and $q = 0$ then (3.2) fails to hold, because $\Pr(Y \ge X) = \Pr(X = 0) \sim m^{-1}$ but $\Pr(Y \ge X - 1) = \Pr(X \le 1) \sim m^{-1} \log m$.

Nevertheless, it is still possible to consider similar estimates in the sparse case. Here is an analogue of (3.2) that holds with $p = O(m^{-1} \log m)$.

Proposition 3.2. *If $\frac{1}{2} \log m \le mp \le 128 \log m$ and $1 \le \ell \le \log m$ then*

$$\Pr(Y \ge X - \ell) \le \left(\frac{C \log m}{\ell}\right)^{C\ell} \Pr(Y \ge X),$$

where $C > 0$ is a universal constant.

3.2 Majorities are uncorrelated

The preceding propositions may be combined to show that the event that u has a minority is essentially independent of the event that v has a minority. First, we observe that removing one trial from a binomial random variable doesn't change very much.

Lemma 3.3. *There is a universal constant $C > 0$ such that for all m, n and all $p, q \le 2/3$,*

$$(1 - Cm^{-1/3}) P(m-1, n, p, q) - 2m^{-2} \le P(m, n, p, q) \le (1 + Cn^{-1/3}) P(m, n-1, p, q) + 2n^{-2}.$$

Proof. Assume without loss of generality that $p \ge q$. Let $X' \sim \mathrm{Binom}(m-1, p)$, $Y' \sim \mathrm{Binom}(n-1, q)$, $\xi_X \sim \mathrm{Bernoulli}(p)$ and $\xi_Y \sim \mathrm{Bernoulli}(q)$ be independent, and then take

$X = X' + \xi_X$ and $Y = Y' + \xi_Y$. In terms of these variables, the left-hand inequality above may be written as

$$(1 - Cm^{-1/3}) \Pr(Y \geq X') - 2m^{-2} \leq \Pr(Y \geq X)$$

We will focus on this inequality (since the other inequality is essentially identical). Now,

$$\begin{aligned} \Pr(Y \geq X) &= \Pr(\xi_X = 0, Y \geq X') + \Pr(\xi_X = 1, Y \geq X' + 1) \\ &= (1-p) \Pr(Y \geq X') + p \Pr(Y \geq X' + 1). \end{aligned} \tag{3.3}$$

If we assume that $(m-1)p \geq 64 \log(m-1)$ then (3.1) implies that

$$\begin{aligned} p \Pr(Y \geq X' + 1) &\geq p \left(1 - C\sqrt{\frac{\log m}{mp}}\right) \Pr(Y \geq X') - 2m^{-2} \\ &\geq \left(p - C\sqrt{\frac{\log m}{m}}\right) \Pr(Y \geq X') - 2m^{-2}. \end{aligned}$$

Plugging this into (3.3) yields

$$(1 - Cm^{-1/3}) \Pr(Y \geq X') - 2m^{-2} \leq \Pr(Y \geq X),$$

which implies the claim. On the other hand, if $(m-1)p \leq 64 \log(m-1)$ then directly from (3.3) we have

$$\Pr(Y \geq X) \geq (1-p) \Pr(Y \geq X') \geq (1 - Cm^{-1/3}) \Pr(Y \geq X'). \qquad \square$$

Next, we show that $\{u$ has a minority$\}$ and $\{v$ has a minority$\}$ are essentially uncorrelated. We recall that if A and B are events then $\mathrm{Cov}(A, B) = \Pr(A \cap B) - \Pr(A) \Pr(B)$.

Lemma 3.4. *Fix nodes u and v. Let A and B be the events that u and v respectively have minorities. If $p, q \leq 2/3$ then*

$$|\mathrm{Cov}(A, B)| \leq Cn^{-1/3} \Pr(A) \Pr(B) + Cn^{-4}.$$

Proof. Assume that $p > q$ and that $\sigma_u = +$ and $\sigma_v = -$ (the other cases are very similar). Let ξ be the indicator that $u \sim v$, and let A and B be the events that u and v respectively have minorities. Note that A and B are conditionally independent given ξ, which means that

$$\begin{aligned} \mathrm{Cov}(A, B) &= \mathrm{Cov}(\Pr(A \mid \xi), \Pr(B \mid \xi)) \\ &\leq \sqrt{\mathrm{Var}(\Pr(A \mid \xi)) \mathrm{Var}(\Pr(B \mid \xi))} \\ &= \mathrm{Var}(\Pr(A \mid \xi)), \end{aligned}$$

where the last equality holds because A and B have the same distribution given ξ.

Define $\alpha = P(n-1, n, p, q) = \Pr(u$ has a minority$) = \Pr(v$ has a minority$)$. By our assumption that $\sigma_u \neq \sigma_v$ and $p > q$, we have $\Pr(A \mid \xi = 0) \leq \Pr(A \mid \xi = 1)$. On the other hand,

$$\Pr(A \mid \xi = 0) = P(n-1, n-1, p, q) \geq (1 - Cn^{-1/3})\alpha - 2n^{-2}.$$

by Lemma 3.3.

Next, we consider $\Pr(A \mid \xi = 1)$. Note that

$$\begin{aligned} \Pr(A \mid \xi = 1) &= \Pr(1 + \mathrm{Binom}(n-1, q) \geq \mathrm{Binom}(n-1, p)) \\ &\leq \Pr(1 + \mathrm{Binom}(n, q) \geq \mathrm{Binom}(n-1, p)). \end{aligned}$$

By applying either (3.2) or Proposition 3.2 to the right hand side above, we have

$$\Pr(A \mid \xi = 1) \le \begin{cases} (1 + Cn^{-1/6})\alpha + 2n^{-2} & p \ge n^{-1/2} \\ \alpha \log^C n + 2n^{-2} & \text{otherwise.} \end{cases}$$

(To get the second case, we are either applying (3.2) for $64 \log n \le np \le n^{1/2}$ or we are applying Proposition 3.2.) In the first case, the random variable $\Pr(A \mid \xi)$ is supported on an interval of width at most $Cn^{-1/6}\alpha + Cn^{-2}$ and so its variance is at most $Cn^{-1/3}\alpha^2 + Cn^{-4}$. In the second case, $\Pr(\xi = 1) = q \le p \le n^{-1/2}$, and so

$$\begin{aligned} \mathrm{Var}(\Pr(A \mid \xi)) &\le \mathbb{E}(\Pr(A \mid \xi) - \alpha)^2 \\ &\le \Pr(\xi = 0) C\alpha^2 n^{-2/3} + \Pr(\xi = 1) C\alpha^2 \log^{2C} n + Cn^{-4}, \end{aligned}$$

which is bounded by $C\alpha^2 n^{-1/3} + Cn^{-4}$. □

3.3 Graph structure

Finally, we will use our preceding estimates to prove Proposition 2.7. Most of the proof essentially follows by straightforward first moment arguments. The most complicated part is showing that $P(n, p_n, q_n) = \Omega(n^{-1})$ implies that with constant probability there exists a node with a minority. This uses a fairly standard second moment argument, the main technical part of which is contained in Lemma 3.4.

Proof of Proposition 2.7. Fix a node $v \in V(G)$ and suppose without loss of generality that $\sigma_v = +$. For notational convenience, we will also suppose that $p > q$; an essentially identical proof works for $p < q$. Let X and Y denote the number of $+$- and $-$-labelled neighbors of v. Then

$$\begin{aligned} X &\sim \mathrm{Binom}(n-1, p_n) \\ Y &\sim \mathrm{Binom}(n, q_n). \end{aligned}$$

Suppose first that $P(n, p_n, q_n) = o(1)$. Then

$$\Pr(v \text{ has a minority}) = \Pr(Y \ge X) = P(n-1, n, p_n, q_n) = o(1)$$

by Lemma 3.3. Summing over $v \in V(G)$, we have

$$\mathbb{E}(\#\text{ of nodes with a minority}) = o(n),$$

and so Markov's inequality implies that a.a.s. all but $o(n)$ nodes have a majority.

The case where $P(n, p_n, q_n) = o(n^{-1})$ is very similar, except that we conclude with $\mathbb{E}(\#\text{ of nodes with a minority}) = o(1)$, which implies that a.a.s. every node has a majority.

For the rest of the proof, we will assume that $p_n, q_n \le 2/3$. As we explained in Section 2.2, this case suffices: if $p_n, q_n \ge 1/3$ then we may apply the result with p_n and q_n replaced by $1 - p_n$ and $1 - q_n$; if $q_n \le 1/3$ and $p_n \ge 2/3$ then $P(n, p_n, q_n) = o(n^{-1})$ and we have already given that part of the proof.

Suppose that the number of nodes without a majority is not $o(n)$ a.a.s. Then there is some $\epsilon > 0$ such that for infinitely many n, the probability of having ϵn nodes with a minority is at least ϵ. Thus, the expected number of nodes with a minority is at least $\epsilon^2 n$ for infinitely many n, which in turn implies that $P(n-1, n, p_n, q_n) = \Pr(Y \ge X) \ge \epsilon^2$ for infinitely many n. By Lemma 3.3, $P(n, p_n, q_n) \not\to 0$.

It remains to prove that all nodes have a majority a.a.s. only if $P(n, p_n, q_n) = o(n^{-1})$. This requires a second moment argument: let ξ_u be the indicator that u has a minority and let $N = \sum_u \xi_u$ be the number of nodes with a minority. If $\alpha = \Pr(u \text{ has a minority})$ (which is the same for all u) then

$$\begin{aligned}\mathrm{Var}(N) &= \sum_u \mathrm{Var}(\xi_u) + \sum_{u \neq v} \mathrm{Cov}(\xi_u, \xi_v) \\ &\leq n\alpha + Cn^2\alpha^2 n^{-1/3} + Cn^{-2},\end{aligned}$$

where the last line follows from Lemma 3.4. In particular, we may bound $\mathrm{Var}(N) \leq C \max\{\mathbb{E}N, (\mathbb{E}N)^2, n^{-2}\}$. Now, if $P(n, p_n, q_n)$ is not $o(n^{-1})$ then there is some $\epsilon > 0$ and infinitely many N for which $\mathbb{E}N \geq \epsilon$. By the Paley-Zygmund inequality and our bound on $\mathrm{Var}(N)$, there is some $\delta > 0$ such that for infinitely many n, $\Pr(N \geq \delta) \geq \delta$. Since $\{N > 0\} = \{\exists u \text{ with a minority}\}$, this implies that the event of having only majorities is not asymptotically almost sure. □

4 Sufficient condition for strong consistency

The rough idea behind our strongly consistent labelling algorithm is to first run a weakly consistent algorithm and then try to improve it. The natural way to improve an almost-accurate labelling τ is to search for nodes u that have a minority with respect to τ and flip their signs. In fact, if the errors in τ were independent of the neighbors of u then this would work quite well: assuming that u has a decently large majority (which it will, for most u, by Proposition 3.1), then having a labelling τ with few errors is like observing each neighbor of u with a tiny amount of noise. This tiny amount of noise is very unlikely to flip u's neighborhood from a majority to a minority. Therefore, choosing u's sign to give it a majority is a reasonable approach.

There are two important problems with the argument outlined in the previous paragraph: it requires the errors in τ to be independent, and it is only guaranteed to work for those u that have a sizeable majority (i.e., almost, but not quite, all the nodes in G). Nevertheless, this procedure is a good starting point and it motivates the first clean-up stage of our algorithm (Algorithm 1). By removing u from the graph before looking for the almost-accurate labelling τ, we ensure the required independence properties (as a result, note that we will be dealing with multiple labellings τ, depending on which nodes we removed before running our almost-accurate labelling algorithm). And although the final labelling we obtain is not guaranteed to be entirely correct, we show that it has very few (i.e., at most n^ϵ) errors whereas the initial labelling was only guaranteed to have $o(n)$ errors.

In order to finally produce the correct labelling, we return to the earlier idea: flipping the label of every node that has a minority. We analyze this procedure by noting that after the previous step of the algorithm, the errors were confined to a very particular set of nodes (namely, those without a very strong majority). We show that this set of nodes is small and poorly connected, which means that every node in the graph is guaranteed to only have a few neighbors in this bad set. In particular, even nodes with relatively weak majorities cannot be flipped by labelling errors in the bad set. We analyze this procedure in Section 4.3.

4.1 The initial guess

As stated in the introduction, there exist algorithms for a.a.s. correctly labelling all but $o(n)$ nodes. Assuming that $p_n + q_n = \Omega(n^{-1} \log n)$, such an algorithm is easy to describe, and we include it for completeness; indeed, the algorithm we give is essentially folklore, although a nice treatment is given in [24]. A slightly more complex algorithm that doesn't assume $p_n + q_n = \Omega(n^{-1} \log n)$ can be found in [27].

Note that the conditional expectation of the adjacency matrix given the labels is $\frac{p_n+q_n}{2}11^T + \frac{p_n-q_n}{2}\sigma\sigma^T$, where $\sigma \in \{\pm 1\}^{2n}$ is the true vector of class labels. Now, let A be the adjacency matrix of G. Then σ is the second eigenvector of $\mathbb{E}[A \mid \sigma]$, and its eigenvalue is $\frac{p_n-q_n}{2}$. In particular, if we had access to $\mathbb{E}[A \mid \sigma]$ then we could recover the labels exactly, simply by looking at its second eigenvector. Instead, we have access only to A. However, if A and $\mathbb{E}[A \mid \sigma]$ are close then we can recover the labels by rounding the second eigenvector of A.

Conditioned on σ, $A - \mathbb{E}[A \mid \sigma]$ is a symmetric matrix whose upper triangular part consists of independent entries, and so we can use results from random matrix theory [25, 26] to bound its norm:

Theorem 4.1. *If $p_n + q_n = \Omega(n^{-1}\log n)$ then there is a constant C such that*

$$\|A - \mathbb{E}[A \mid \sigma]\| \le C\sqrt{n(p_n+q_n)}$$

a.a.s. as $n \to \infty$, where $\|\cdot\|$ denotes the spectral norm.

Assuming Theorem 4.1, note that if $|p_n - q_n|/\sqrt{n(p_n+q_n)} \to \infty$ then $\|A - \mathbb{E}[A \mid \sigma]\|$ is order-wise smaller than the second eigenvalue of A. By the Davis-Kahan theorem, it is possible to recover σ up to an error of size $o(1)\|\sigma\|$. This implies that we can recover the labels of all but $o(n)$ vertices.

4.2 The replica step

Let `BBPartition` be an algorithm that is guaranteed to a.a.s. label all but $o(n)$ nodes correctly; we will use it as a black box. Note that we may assume that `BBPartition` produces an exactly balanced labelling. If not, then if its output has more $+$ labels than $-$ labels, say, we can randomly choose some $+$-labelled vertices and relabel them. The new labelling is balanced, and it is still guaranteed to have at most $o(n)$ mistakes.

For the remainder of Section 4, we will assume that $p \ge q$ in order to lighten our notation. The case $p < q$ is very similar, except that expressions like $\Pr(Y \ge X - \ell)$ should be replaced by $\Pr(X \ge Y - \ell)$. We will also assume that $p \le 2/3$; as discussed in Section 2.2, all interesting cases may be reduced to this one.

We define V_ϵ to be a set of "bad" nodes that our first step is not required to label correctly.

Definition 4.2. *Let V_ϵ be the elements of V that have a majority of size less than $\epsilon\sqrt{np\log n}$, or that have more than $100np$ neighbors.*

Proposition 4.3. *For any $\epsilon > 0$, Algorithm 1 a.a.s. correctly labels every node in $V \setminus V_\epsilon$.*

Before proving Proposition 4.3, we deal with a minor technical point. The following lemma shows that we can apply `BBPartition` to subgraphs of $G \sim \mathcal{G}(2n, p_n, q_n)$, and it will still have the required guarantees.

Lemma 4.4. *If $P(n, p_n, q_n) = o(n^{-1})$ then for any $\alpha > 0$, $P(\lfloor \alpha n \rfloor, p_n, q_n) \to 0$.*

Proof. This follows from two simple properties of the function P. First, we have $P(n_1 + n_2, p, q) \ge P(n_1, p, q)P(n_2, p, q)$ for any n_1, n_2, p, and q. Indeed, if $X_i \sim \operatorname{Binom}(n_i, p)$ and $Y_i \sim \operatorname{Binom}(n_i, q)$ are independent then

$$\begin{aligned} P(n_1 + n_2, p, q) &= \Pr(X_1 + X_2 \le Y_1 + Y_2) \\ &\ge \Pr(X_1 \le Y_1)\Pr(X_2 \le Y_2) \\ &= P(n_1, p, q)P(n_2, p, q). \end{aligned}$$

A similar coupling argument shows that for any $n_2 \ge 0$, $P(n_1, p, q) \ge \frac{1}{2}P(n_1 + n_2, p, q)$.

input : graph G, parameter $\epsilon > 0$
output : a partition W_+, W_- of $V(G)$

1 $W_+ \leftarrow \emptyset$;
2 $W_- \leftarrow \emptyset$;
3 choose $m \in \mathbb{N}$ so $(1 - 2/m)\epsilon - 80m^{-1/2} \geq \epsilon/2$;
4 partition $V(G)$ randomly into $U_1, \ldots, U_m$;
5 $U_+, U_- \leftarrow$ `BBPartitition`(G);
6 **for** $i \leftarrow 1$ **to** m **do**
7     $U_{i,+}, U_{i,-} \leftarrow$ `BBPartition`$(G \setminus U_i)$;
8   **if** $|U_{i,+} \Delta U_+| \geq n/2$ **then**
9   swap $U_{i,+}$ and $U_{i,-}$;
10   **end**
11   **for** $v \in U_i$ **do**
12   **if** $p > q$ *and* $\#\{u \in U_{i,+} : u \sim v\} > \#\{u \in U_{i,-} : u \sim v\}$ **then**
13             $W_+ \leftarrow W_+ \cup \{v\}$;
14   **else if** $p < q$ *and* $\#\{u \in U_{i,+} : u \sim v\} < \#\{u \in U_{i,-} : u \sim v\}$ **then**
15             $W_+ \leftarrow W_+ \cup \{v\}$;
16   **else**
17             $W_- \leftarrow W_- \cup \{v\}$;
18   **end**
19   **end**
20 **end**

Algorithm 1: Algorithm for initial accuracy boost

Indeed, conditioned on $X_1 + X_2 \leq Y_1 + Y_2$, the probability of $X_1 \leq Y_1$ is at least $\frac{1}{2}$. Hence,

$$\begin{aligned} P(n_1, p, q) &= \Pr(X_1 \leq Y_1) \\ &\geq \Pr(X_1 \leq Y_1 \mid X_1 + X_2 \leq Y_1 + Y_2) \Pr(X_1 + X_2 \leq Y_1 + Y_2) \\ &\geq \frac{1}{2} P(n_1 + n_2, p, q). \end{aligned}$$

Now, choose an integer k so that $\alpha \geq 1/k$. Then

$$P(n, p, q) \geq \frac{1}{2} P(2k\lfloor n/k \rfloor, p, q) \geq \frac{1}{2} P(\lfloor n/k \rfloor, p, q)^{2k} \geq \frac{1}{4} P(\lfloor \alpha n \rfloor, p, q)^{2k}.$$

Since k and α are constant as $n \to \infty$, this completes the proof. □

Proof of Proposition 4.3. First, we may assume without loss of generality that the partition U_+, U_- that was produced in line 5 is positively correlated with the true labelling σ. By our assumption on `BBPartition`, at line 8 $U_{i,+}$ either agrees with $V_+ \setminus U_i$ or $V_- \setminus U_i$, up to an error of $o(n)$. After the relabelling in line 9, then, a.a.s. $U_{i,+}$ agrees with $V_+ \setminus U_i$ up to an error of $o(n)$. Since m is a constant independent of n, this property a.a.s. holds for every i simultaneously.

Now, consider a node $v \notin V_\epsilon$ and suppose without loss of generality that $\sigma_v = +$. Conditioned on $v \in U_i$, every other node is added to U_i independently with probability $1/m$. Hence, conditioned on v having k_+ $+$-labelled neighbors and k_- $-$-labelled neighbors, it has $\mathrm{Binom}(k_+, 1/m)$ $+$-labelled neighbors in U_i and $\mathrm{Binom}(k_-, 1/m)$ $-$-labelled neighbors in U_i. Let $k_{+,i}$ denote the number of $+$-labelled neighbors that v has in U_i and let $k_{+,\neg i} = k_+ - k_{+,i}$ be the number of $+$-labelled neighbors that v has in $V \setminus U_i$ (and similarly for $-$).

By Bernstein's inequality, with probability at least $1 - 2n^{-2}$,

$$k_{+,i} \in k_+/m \pm 4\sqrt{k_+ m^{-1} \log k_+} \tag{4.1}$$

$$k_{-,i} \in k_-/m \pm 4\sqrt{k_- m^{-1} \log k_-}. \tag{4.2}$$

Recall that $v \notin V_\epsilon$ implies that $k_+ \leq 100np$, $k_- \leq 100np$ and

$$k_+ - k_- \geq \epsilon \sqrt{np \log n}.$$

Hence, (4.1) and (4.2) imply that

$$\begin{aligned} k_{+,\neg i} - k_{-,\neg i} &\geq (1 - 2/m)\epsilon\sqrt{np\log n} - 4\sqrt{k_+ m^{-1}\log k_+} - 4\sqrt{k_- m^{-1} \log k_-} \\ &\geq (1-2/m)\epsilon\sqrt{np\log n} - 80m^{-1/2}\sqrt{np \log n} \\ &\geq \frac{\epsilon}{2}\sqrt{np\log n}, \end{aligned}$$

where the last inequality follows from the definition of m. Taking a union bound over the events leading to (4.1), we see that a.a.s., for every $v \notin V_\epsilon$ with $\sigma_v = +$, if $v \in U_i$ then

$$(k_{+,\neg i} - k_{-,\neg i}) \geq \frac{\epsilon}{2}\sqrt{np \log n}. \tag{4.3}$$

In other words, every $v \notin V_\epsilon$ still has a strong majority, even if we consider only edges between v and the complement of U_i.

Let X_- be the number of $+$-valued neighbors of v that were incorrectly labelled as $-$ in line 9 (i.e. $X_- = |\{u : u \sim v, \sigma_u = +, u \in U_{i,-}\}|$), and let X_+ be the number of $-$-valued neighbors that were incorrectly labelled as $+$. Note that the quantities considered in line 12 of Algorithm 1 may be expressed in terms of k and X as

$$\begin{aligned} \#\{u \in U_{i,+} : u \sim v\} &= k_{+,\neg i} - X_- + X_+ \\ \#\{u \in U_{i,-} : u \sim v\} &= k_{-,\neg i} + X_- - X_+. \end{aligned}$$

Hence, the inequality $|X_+ - X_-| < \frac{1}{2}|k_{+,\neg i} - k_{i,\neg i}|$ will imply that v is correctly labelled in lines 12–18. For the rest of the proof, our goal will be to show that a.a.s. the above inequality holds for all $v \notin V_\epsilon$.

Let $E_- = \#\{u \in U_{i,-} : \sigma_u = +\}$ (i.e., the total number of $+$-labelled vertices that were mislabelled in line 9) and let $E_+ = \#\{u \in U_{i,+} : \sigma_u = -\}$. Note that the neighbors of v are independent of $U_{i,-}$, and so conditioned on $k_{+,\neg i}$ and $k_{-,\neg i}$,

$$\begin{aligned} X_- &\stackrel{d}{=} \mathrm{HyperGeom}(|V_+ \setminus U_i|, k_{+,\neg i}, E_-) \\ X_+ &\stackrel{d}{=} \mathrm{HyperGeom}(|V_- \setminus U_i|, k_{-,\neg i}, E_+), \end{aligned}$$

where V_+ and V_- are the set of u with $\sigma_u = +$ and $\sigma_u = -$, respectively. Now condition on $k_{+,\neg i}$ and $k_{-,\neg i}$, and on the following a.a.s. events:

$$\begin{aligned} \forall i \quad |V_+ \setminus U_i| &\in n(1-1/m) \pm \sqrt{n}\log\log n \\ \forall i \quad |V_- \setminus U_i| &\in n(1-1/m) \pm \sqrt{n}\log\log n \\ |E_- - E_+| &\leq \sqrt{n \log\log n}. \end{aligned}$$

Under the above events, and recalling that $k_+ \leq 100np$,

$$\begin{aligned} |\mathbb{E}X_- - \mathbb{E}X_+| &= \left| E_- \frac{k_{+,\neg i}}{|V_+ \setminus U_i|} - E_+ \frac{k_{-,\neg i}}{|V_- \setminus U_i|} \right| \\ &\leq \left| E_- \frac{k_{+,\neg i}}{n(1-1/m)} - E_+ \frac{k_{-,\neg i}}{n(1-1/m)} \right| + O(n^{1/2} p \log\log n) \\ &\leq O(n^{-1})|E_- - E_+| k_+ + O(n^{-1}) E_+ \big| k_{+,\neg i} - k_{-,\neg i} \big| + O(\sqrt{np}\log\log n) \\ &\leq O(\sqrt{np}\log\log n) + o(1)\big|k_{+,\neg i} - k_{-,\neg i}\big|, \end{aligned}$$

Going back to (4.3), we see that a.a.s. for all $v \notin V_\epsilon$,

$$|\mathbb{E}X_- - \mathbb{E}X_+| \le \frac{1}{8}|k_{+,\neg i} - k_{-,\neg i}|.$$

Next, we consider the deviations of X_- and X_+ around their means. By Bernstein's inequality for hypergeometric variables, there is a constant C such that with probability $1 - n^{-2}$, X_- is within

$$C\sqrt{E_- \frac{k_{+,\neg i}}{|V_+ \setminus U_i|} \log E_-} \le C' \sqrt{E_- p \log n}$$

of its expectation. Since $E_- = o(n)$, we can take n large enough so that X_- is within $\frac{\epsilon}{16}\sqrt{np\log n}$ of its expectation with probability $1 - n^{-2}$. Arguing similarly for X_+ we have

$$\begin{aligned} |X_- - X_+| &\le |\mathbb{E}X_- - \mathbb{E}X_+| + |X_- - \mathbb{E}X_-| + |X_+ - \mathbb{E}X_+| \\ &\le \frac{1}{8}|k_{+,\neg i} - k_{-,\neg i}| + \frac{\epsilon}{8}\sqrt{np\log n} \end{aligned}$$

with probability $1 - 2n^{-2}$. Taking a union bound over $v \notin V_\epsilon$ (recall that X and k both depend on v), we see that the above inequality holds a.a.s. for all $v \notin V_\epsilon$ simultaneously. By (4.3), a.a.s. for all $v \in V_\epsilon$,

$$|X_- - X_+| \le \frac{3}{8}|k_{+,\neg i} - k_{-,\neg i}|,$$

which completes the proof. □

4.3 The hill-climbing step

After running Algorithm 1, we are left with a graph in which only nodes belonging to V_ϵ could possibly be mis-labelled. Fortunately, very few nodes belong to V_ϵ, and those that do are poorly connected to the rest of the graph. This is the content of the next two propositions.

Proposition 4.5. *For every $\delta > 0$ there exists an $\epsilon > 0$ such that if $P(n,p,q) = o(n^{-1})$ then $|V_\epsilon| \le n^\delta$ a.a.s.*

Proof. Consider a single $v \in V$. By Bernstein's inequality the probability that v has $100np$ neighbors is less than n^{-2} (using $np \ge \log n$, which follows from $P(n,p,q) = o(n^{-1})$). Hence, a.a.s. every v has at most $100np$ neighbors.

It remains to show that a.a.s. at most n^δ vertices fail to have a majority of size $\epsilon\sqrt{np\log n}$. Now, if $np \ge 64\log n$ then Proposition 3.1 with $\ell = \epsilon\sqrt{np\log n}$ implies that if $Y \sim \mathrm{Binom}(n,q)$ and $X \sim \mathrm{Binom}(n-1,p)$ then

$$\Pr(Y \ge X - \epsilon\sqrt{np\log n}) \le 2n^{-2} + O(n^{-1+C\epsilon}).$$

In particular, if $C\epsilon < \delta$ then the right hand size is $o(n^{-1+\delta})$. By Markov's inequality, this implies that a.a.s. at most n^δ nodes fail to have a majority of size $\epsilon\sqrt{np\log n}$.

In the sparse case (i.e. $\frac{1}{2}\log n \le np \le 128\log n$), Proposition 3.2 with $\ell = \epsilon\sqrt{np\log n} = \Theta(\epsilon\log n)$ yields

$$\Pr(Y \ge X - \epsilon\sqrt{np\log n}) \le (2C/\epsilon)^{C\epsilon\log n} n^{-1}.$$

Since $(2/\epsilon)^\epsilon \to 1$ as $\epsilon \to 0$, we may choose ϵ so that $(2C/\epsilon)^{C\epsilon\log n} \le n^{\delta/2}$. By Markov's inequality, we see that at most n^δ nodes fail to have a majority of size $\epsilon\sqrt{np\log n}$. □

Proposition 4.6. *Suppose that $P(n,p,q) = o(n^{-1})$ and $np \le n^{1/4}$. For sufficiently small ϵ, a.a.s. no node has two or more neighbors in V_ϵ.*

Proof. Fix $u, v \in V$; let $X \sim \mathrm{Binom}(n-1, p)$ and $Y \sim \mathrm{Binom}(n, q)$. As in the proof of Proposition 4.5, a.a.s. every $v \in V$ has at most $100np$ neighbors; for the rest of the proof, we condition on this event. Moreover, we may choose ϵ small enough so that $\Pr(Y \ge X - \epsilon\sqrt{np\log n}) \le n^{-7/8}$. In particular, that means that $\Pr(u \in V_\epsilon) \le n^{-7/8}$. Now condition on the neighbors of u. If v has a majority of $2\epsilon\sqrt{np\log n}$ on all edges except for u, then it lies outside of V_ϵ regardless of whether it neighbors u. But this event is independent of whether $u \in V_\epsilon$, and if ϵ is sufficiently small then it has probability at least $1 - n^{-7/8}$. Hence, $\Pr(u, v \in V_\epsilon) \le n^{-7/4}$.

Now condition on the event that $u, v \in V_\epsilon$. Recall that u and v each have at most $100np \le 100n^{1/4}$ neighbors in V_- and at most $100n^{1/4}$ neighbors in V_+. Conditioned on the number of neighbors in V_- and V_+, the neighbors of u and v are independent and uniformly distributed. Hence, the probability that they have a common neighbor is $O(n^{-3/4-3/4+1}) = O(n^{-1/2})$. Combining this with the previous paragraph, we have

$$\Pr(u, v \in V_\epsilon \text{ and they have a common neighbor}) = O(n^{-9/4}).$$

Taking a union bound over n^2 choices of u and v completes the proof. ☐

Proposition 4.7. *Suppose that $np \le n^{1/4}$. For sufficiently small ϵ, a.a.s. no two nodes in V_ϵ are adjacent.*

Proof. Fix $u, v \in V$. The probability that they are adjacent is at most $p \le n^{-3/4}$. As in the previous proof, if ϵ is small enough then $\Pr(u \in V_\epsilon \mid u \sim v)$ and $\Pr(v \in V_\epsilon \mid u \sim v, u \in V_\epsilon)$ are both at most $n^{-7/8}$. Multiplying these conditional probabilities, we have

$$\Pr(u, v \in V_\epsilon \text{ and } u \sim v) = O(n^{-5/2}),$$

and we conclude by taking a union bound over u and v. ☐

input : graph G, an initial partition U_+, U_- of $V(G)$
output : a partition W_+, W_- of $V(G)$

$W_+ \leftarrow \{v \in V(G) : v \text{ has more neighbors in } U_+ \text{ than in } U_-\}$;
$W_- \leftarrow V(G) \setminus W_+$;

Algorithm 2: Algorithm for final labelling

Proposition 4.8. *Suppose that we initialize Algorithm 2 with a partition whose errors are restricted to V_ϵ, and suppose that $P(n, p_n, q_n) = o(n^{-1})$. Then a.a.s., Algorithm 2 returns the true partition.*

Proof. We consider two cases: the dense regime $n^{1/4} \le np \le 2n/3$, and the sparse regime $\frac{1}{2}\log n \le npn^{1/4}$.

In the dense regime, note that by Proposition 3.1, a.a.s. every node has a majority of $\Omega(\sqrt{np/\log n}) \ge \Omega(n^{1/9})$. On the other hand, if ϵ is sufficiently small then (by Proposition 4.5) $|V_\epsilon| \le n^{1/10}$, which implies that every node in V_+ will have most of its neighbors in U_+. Therefore, $W_+ = V_+$ in Algorithm 2.

In the sparse regime, let V' be the set of nodes with a majority of less than three; note that $V' \subset V_\epsilon$. By Proposition 4.6, a.a.s. every node has at most one neighbor in V_ϵ, which implies that every node in $V_+ \setminus V'$ has most of its neighbors in U_+; hence every node outside of V' will be correctly labelled. On the other hand, Proposition 4.7 shows that nodes in V' are also correctly labelled, since none of them have any neighbors in V_ϵ (recalling that $V' \subset V_\epsilon$). ☐

5 Necessary condition for strong consistency

A classical fact in Bayesian statistics says that if we are asked to produce a configuration $\hat{\sigma}$ from the graph G, then the algorithm with the highest probability of success is the *maximum a posteriori* estimator, $\hat{\sigma}$, which is defined to be any $\tau \in \{-1,1\}^{V(G)}$ satisfying $\sum_u \tau_u = 0$ that maximizes $\Pr(G \mid \sigma = \tau)$. (To see that this is the estimator with the highest probability of success, note that every τ that maximizes $\Pr(G \mid \sigma = \tau)$ also maximizes $\Pr(\sigma = \tau \mid G)$; clearly, a τ that maximizes the latter quantity is an optimal estimate.) In order to prove that $P(n, p_n, q_n) = o(n^{-1})$ is necessary for strong consistency, we relate the success probability of $\hat{\sigma}$ to the existence of nodes with minorities. Note that we say v has a majority with respect to τ if (assuming $p > q$) τ gives the same label to v as it does to most of v's neighbors.

Lemma 5.1. *If there is a unique maximal $\hat{\sigma}$ then with respect to $\hat{\sigma}$, there cannot be both a $+$-labelled node with a minority and a $-$-labelled node with a minority.*

Proof. For convenience, we will assume that $p > q$. The same proof works for $p < q$, but one needs to remember that the definition of "majority" and "minority" swap in that case (Definition 2.6).

The probability of G conditioned on the labelling τ may be written explicitly: if A_τ is the set of unordered pairs $u \neq v$ with $\tau_u = \tau_v$ and B_τ is the set of unordered pairs $u \neq v$ with $\tau_u \neq \tau_v$ then

$$\begin{aligned}\Pr(G \mid \sigma = \tau) &= p^{|E(G)\cap A_\tau|} q^{|E(G)\cap B_\tau|} (1-p)^{|A_\tau \setminus E(G)|} (1-q)^{|B_\tau \setminus E(G)|} \\ &= (1-p)^{|A_\tau|}(1-q)^{|B_\tau|} \left(\frac{p}{1-p}\right)^{|E(G)\cap A_\tau|} \left(\frac{q}{1-q}\right)^{|E(G)\cap B_\tau|}. \end{aligned} \tag{5.1}$$

Consider a labelling τ. Suppose that there exist nodes u and v with $\tau_u = +$ and $\tau_v = -$, and such that both u and v have minorities with respect to τ. We will show that τ cannot be the unique maximizer of $\Pr(G \mid \sigma = \tau)$, which will establish the lemma.

Consider the labelling τ' that is identical to τ except that $\tau'_u = -$ and $\tau'_v = +$. The fact that u and v both had minorities with respect to τ implies that

$$\begin{aligned}|E(G) \cap A_{\tau'}| &\geq |E(G) \cap A_\tau| \\ |E(G) \cap B_{\tau'}| &\geq |E(G) \cap B_\tau|\end{aligned}$$

(note that equality is possible in the inequalities above if u and v are neighbors). On the other hand, the number of $+$ and $-$ labels are the same for τ and τ'; hence $|A_\tau| = |A_{\tau'}|$ and $|B_\tau| = |B_{\tau'}|$. Looking back at (5.1), therefore, we have

$$\Pr(G \mid \sigma = \tau) \leq \Pr(G \mid \sigma = \tau').$$

Hence, τ cannot be the unique maximizer of $\Pr(G \mid \sigma = \tau)$. $\square$

In order to argue that $P(n, p_n, q_n) = o(n^{-1})$ is necessary for strong consistency, we need to show that if $P(n, p_n, q_n)$ is not $o(n^{-1})$ then $(G, \sigma) \sim \mathcal{G}(2n, p_n, q_n)$ has a non-vanishing chance of containing nodes of both labels with minorities.

Suppose that $P(n, p_n, q_n)$ is not $o(n^{-1})$. By Proposition 2.7, there is some $\epsilon > 0$ such that for infinitely many n, $\Pr(\exists u : u \text{ has a minority}) \geq \epsilon$. Since $+$-labelled nodes and $-$-labelled nodes are symmetric, there are infinitely many n such that

$$\begin{aligned}\Pr(\exists u : \sigma_u = + \text{ and } u \text{ has a minority}) &\geq \epsilon/2 \\ \Pr(\exists v : \sigma_v = - \text{ and } u \text{ has a minority}) &\geq \epsilon/2.\end{aligned}$$

By Harris's inequality [11], the two events above are non-negatively correlated because both of them are monotonic events with the same directions: both are monotonic increasing in the edges between $+$-labelled and $-$-labelled nodes and monotonic decreasing in the other edges. Hence, there are infinitely many n for which

$$\Pr(\exists u, v : \sigma_u = +, \sigma_v = -,\ u \text{ and } v \text{ have minorities}) \geq \epsilon^2/4.$$

6 Binomial approximations

In this section, we collect various technical, but not particularly enlightening, estimates for binomial variables. Specifically, we prove Propositions 2.8 and 2.9, which give explicit characterizations of the condition $P(n, p_n, q_n) = o(n^{-1})$ in the sparse and dense case respectively, and Proposition 3.1 and 3.2, which give perturbative estimates for binomial probabilities. Our main tools are Bernstein's inequality, Stirling's approximation and Taylor expansion.

6.1 Characterization of sparse strong consistency

For simplicity, in this section we write $a = a_n, b = b_n$ and $c = a + b$. If there is a constant $C > 0$ such that $C^{-1} f \leq g \leq Cf$ then we write $f \asymp g$. We recall that $a, b = \Theta(1)$ and that $pn = a \log n$ and $qn = b \log n$. Let $X \sim \mathrm{Binom}(n, p)$ and $Y \sim \mathrm{Binom}(n, q)$.

We begin with a Poisson approximation to binomials.

Lemma 6.1. *If $Z = X + Y$ then for every $k \leq 10c \log n$,*

$$\Pr(Z = k) = (1 + o(1)) n^{-c} \frac{(c \log n)^k}{k!},$$

where the sequence implicit in the $o(1)$ notation is independent of n and k.

Proof (sketch). By a direct computation, if $k \leq 10c \log n$ then

$$\Pr(X = k) = (1 + o(1)) n^{-a} \frac{(a \log n)^k}{k!}$$
$$\Pr(Y = k) = (1 + o(1)) n^{-b} \frac{(b \log n)^k}{k!},$$

and the sequences implicit in the $1 + o(1)$ notation may be taken to be independent of k. Finally, note that $\Pr(Z = k) = \sum_{\ell=0}^{k} \Pr(Y = \ell) \Pr(X = k - \ell)$. ☐

Proof of Proposition 2.8. We first note that if $a - b \leq \epsilon = \epsilon(C)$ then strong consistency does not hold. This follows because with constant probability we have that X is less than its mean $a_n \log n$ and the probability that Y is larger than $a \log n$ is at least $n^{-1/2}$ if ϵ is a sufficiently small constant.

Without loss of generality, we may assume that $c \geq 1$. Indeed, if $c < 1$ then the proposition is trivially true: on the one hand $P(n, p_n, q_n) = \Omega(n^{-1})$ because $\Pr(X = 0)$ and $\Pr(Y = 0)$ are both $\Omega(n^{-1})$; on the other hand, $(a + b - 2\sqrt{ab} - 1) \log n + \frac{1}{2} \log \log n \to -\infty$ because $a + b = c < 1$ and $\sqrt{ab}$ is bounded away from zero as $n \to \infty$.

Let $Z = X + Y$; then

$$\begin{aligned}\Pr(Y \geq X) &= \sum_{k=0}^{n} \Pr(Z = k) \Pr(Y \geq X \mid Z = k) \\ &= \sum_{k=0}^{10c \log n} \Pr(Z = k) \Pr(Y \geq X \mid Z = k) + O(n^{-2}),\end{aligned}$$

where the second equality follows from the fact that $\Pr(Z \geq 10c \log n) \leq O(n^{-2})$, recalling that $c \geq 1$.

For a fixed $k \leq 10c \log n$, we have that

$$\Pr(Y \geq X \mid Z = k) = (1 - o(1)) \Pr(\mathrm{Binom}(k, \eta) \geq k/2),$$

where $\eta = \frac{b}{a+b} \leq \frac{1}{2}(1-\epsilon)$. Recall that binomial tail probabilities decay exponentially fast; since $\eta \leq \frac{1}{2}(1-\epsilon)$, $\Pr(\mathrm{Binom}(k,\eta) \geq k/2) \asymp \Pr(\mathrm{Binom}(k,\eta) = \lceil k/2 \rceil)$.Combining this with Stirling's approximation we have

$$\Pr(Y \geq X \mid Z = k) \asymp \frac{2^k}{\sqrt{k}} \eta^{k/2} (1-\eta)^{k/2} = \frac{2^k \theta^k}{\sqrt{k}},$$

where $\theta = \sqrt{\eta(1-\eta)} = \frac{\sqrt{ab}}{a+b}$. By Lemma 6.1,

$$\Pr(Z = k) = (1 + o(1)) n^{-c} \frac{(c \log n)^k}{k!},$$

and so Stirling's approximation for $k \geq 1$ gives

$$\Pr(Z = k) \asymp \frac{n^{-c}}{\sqrt{k}} \frac{(ce \log n)^k}{k^k}$$

Thus we get that

$$\begin{aligned} \Pr(Y \geq X) &= \Pr(Y = X = 0) + \sum_{k=1}^{10c \log n} \Pr(Z = k) \Pr(Y \geq X \mid Z = k) + O(n^{-2}) \\ &\asymp n^{-c} \left(1 + \sum_{k=1}^{10c \log n} \frac{(2ce\theta \log n)^k}{k^{k+1}} \right), \end{aligned}$$

The analysis of the sum is standard, and we give a sketch. Defining $\ell(k)$ to be the logarithm of the summand, we have

$$\ell(k) = k \log(t \log n) - (k+1) \log k, \quad t = 2ce\theta.$$

Then

$$\ell'(k) = \log(t \log n) - (1 + 1/k) - \log k, \quad \ell''(k) = -1/k(1 + o(1)),$$

and so the maximum is obtained around the value

$$k^* = e^{-1} t \log n = 2c\theta \log n.$$

Moreover, the maximum value (up to a constant factor) of ℓ is

$$\frac{(2ce\theta \log n)^{k^*}}{k^* (2c\theta \log n)^{k^*}} = \frac{e^{k^*}}{k^*} \asymp \frac{n^{-c+2c\theta}}{\log n} = \frac{n^{2\sqrt{ab}}}{\log n}$$

Since ℓ is approximately quadratic around its maximum and $\ell''(k^*) \asymp -1/\log n$, we see that $\exp(\ell(k))$ varies by a constant factor on a window of length $\sqrt{\log n}$ around k^*, and then drops off geometrically fast beyond that window. Hence, the sum is given (up to a constant) by $n^{2\sqrt{ab}} \log^{-1/2} n$ and so

$$\Pr(Y \geq X) \asymp \frac{n^{2\sqrt{ab} - (a+b)}}{\sqrt{\log n}}$$

Thus $n \Pr(Y \geq X) \to 0$ if and only if

$$(1 + 2\sqrt{ab} - (a+b)) \log n - \frac{1}{2} \log \log n \to -\infty,$$

as needed. □

6.2 Characterization of dense strong consistency

Our main tool for proving Proposition 2.9 will be the following Local Central Limit Theorem. The proof is a standard application of Stirling's approximation.

Lemma 6.2. *Let $C > 0$ be an arbitrary constant and $Y \sim \mathrm{Binom}(n, q)$, where*

$$q = q_n = \omega\left(\frac{\log^3(n)}{n}\right), \quad q_n \le \frac{2}{3}.$$

Let $\sigma_q^2 = q(1-q)$ and let $\phi(x) = (2\pi)^{-1/2}e^{-x^2/2}$. Then for all integers k such that $|k - nq| \le C\sqrt{n \log n}\sigma_q$ it holds that

$$\Pr(Y = k) = (1 + o(1))\frac{1}{\sqrt{n}\sigma_q}\phi\left(\frac{k - nq}{\sqrt{n}\sigma_q}\right).$$

Moreover,

$$\Pr(Y = k) = (1 + o(1))\frac{1}{\sqrt{n}\sigma_q}\phi\left(\frac{x - nq}{\sqrt{n}\sigma_q}\right),$$

for every $k - 1 \le x \le k + 1$.

Proof. The second statement follows easily from the first one using the formula for ϕ and noting that if $\delta \le C\sqrt{n \log n}\sigma_q$ and $|\epsilon| \le 1$ then

$$\left(\frac{\delta + \epsilon}{\sigma_q\sqrt{n}}\right)^2 = \left(\frac{\delta}{\sigma_q\sqrt{n}}\right)^2 + o(1).$$

To prove the first statement, we begin with Stirling's approximation. Noting that $k \to \infty$ as $n \to \infty$, we obtain:

$$\Pr(Y = k) = \binom{n}{k}q^k(1-q)^{n-k} = (1 + o(1))\frac{1}{\sqrt{2\pi}}\sqrt{\frac{n}{k(n-k)}}\left(\frac{nq}{k}\right)^k\left(\frac{n(1-q)}{n-k}\right)^{n-k}.$$

We start by analyzing the term

$$\sqrt{\frac{n}{k(n-k)}} = \frac{1}{\sqrt{n}}\sqrt{\frac{n}{k}}\sqrt{\frac{n}{n-k}}.$$

Now

$$k/n \in [q - C\frac{\sigma_q\sqrt{\log n}}{\sqrt{n}}, q + C\frac{\sigma_q\sqrt{\log n}}{\sqrt{n}}]$$

and since $q = \omega(n^{-1}\log^3 n)$ implies $\frac{\sigma_q\sqrt{\log n}}{\sqrt{n}} = o(q/\log n)$, it follows that $n/k = (1 + o(1/\log n))\frac{1}{q}$. Similarly, $\frac{n}{n-k} = (1 + o(1/\log n))\frac{1}{1-q}$ and so

$$\sqrt{\frac{n}{k(n-k)}} = (1 + o(1/\log n))\frac{1}{\sigma_q\sqrt{n}}. \tag{6.1}$$

Next, we use Taylor expansion around $nq = k$. The first-order term vanishes and we have

$$\begin{aligned} &\log\left(\left(\frac{nq}{k}\right)^k\left(\frac{n(1-q)}{n-k}\right)^{n-k}\right) \\ &\qquad = -\frac{1}{2}(k - nq)^2\left(\frac{1}{k} + \frac{1}{n-k}\right) + O(|nq - k|^3)\left(\frac{1}{k^2} + \frac{1}{(n-k)^2}\right) \\ &\qquad = -\frac{n}{2k(n-k)}(k - nq)^2 + o(1), \end{aligned} \tag{6.2}$$

where the last equality uses the fact that

$$\frac{(nq-k)^3}{\min\{k^2,(n-k)^2\}} \to 0,$$

which follows from the assumption that $q = \omega(n^{-1}\log^3(n))$. Now, from (6.1) we have $\frac{n}{k(n-k)} = (1+o(1/\log n))\frac{1}{n\sigma_q^2}$. Since $(k-nq)^2 = O(\sigma_q^2 n \log n)$, we have

$$\frac{n}{k(n-k)}(k-nq)^2 = \frac{(k-nq)^2}{n\sigma_q^2} + o(1).$$

Going back to (6.2), we have

$$\log\left(\left(\frac{nq}{k}\right)^k \left(\frac{n(1-q)}{n-k}\right)^{n-k}\right) = -\frac{(k-nq)^2}{2n\sigma_q^2} + o(1).$$

The proof follows by combining this with (6.1) and Stirling's approximation for $\Pr(Y = k)$. $\square$

Proof of Proposition 2.9. The second and third conditions are clearly equivalent; we will show the equivalence of the first two.

Bernstein's inequality implies that

$$\Pr(|Y-\mathbb{E}Y| \geq 4\sqrt{n\log n}\sigma_q) = o(n^{-1}), \Pr(|X-\mathbb{E}X| \geq 4\sqrt{n\log n}\sigma_p) = o(n^{-1}).$$

So writing $b_q = 5\sqrt{n\log n}\sigma_q$ and $b_p = 5\sqrt{n\log n}\sigma_p$ we have:

$$\Pr(Y \geq X) = \sum_{k=\lfloor np-b_p\rfloor}^{\lceil np+b_p\rceil} \sum_{\ell=\lfloor nq-b_q\rfloor}^{\lceil nq+b_q\rceil} 1_{\{k\leq \ell\}} \Pr(X=k)\Pr(Y=\ell) + o(n^{-1})$$

Using Lemma 6.2 for every k,ℓ in the range above we have:

$$\Pr(X=k)\Pr(Y=\ell) = (1+o(1))\frac{1}{n\sigma_p\sigma_q}\int_{\Delta(k,\ell)} \phi\left(\frac{y-nq}{\sqrt{n}\sigma_q}\right)\phi\left(\frac{x-np}{\sqrt{n}\sigma_p}\right)dxdy,$$

where $\Delta(k,\ell) = (k,\ell) + \Delta$ where

$$\Delta = \{(x,y) : 0 \leq y \leq 1,\ y-1 \leq x \leq y\}$$

is a parallelogram of unit area. (In applying Lemma 6.2 note that $(x,y) \in \Delta(k,\ell)$ implies that $|x-k| \leq 1$ and $|y-\ell| \leq 1$.) Thus

$$\Pr(Y \geq X) = (1+o(1))\int_{np-b_p}^{np+b_p}\int_{nq-b_p}^{nq+b_q} 1\{x \leq y\}\phi\left(\frac{y-nq}{\sqrt{n}\sigma_q}\right)\phi\left(\frac{x-np}{\sqrt{n}\sigma_p}\right)dydx + o(n^{-1}),$$

where we use the fact that the difference between the union of $\Delta(k,\ell)$ and the integration region above is contained in the set where either $|y-nq| \geq 4\sqrt{n\log n}\sigma_q$ or $|x-np| \geq 4\sqrt{n\log n}\sigma_p$. Changing variables we see that the last expression is nothing but

$$\Pr\left(|M| \leq 5\sqrt{n\log n},\ |N| \leq 5\sqrt{n\log n},\ \sigma_q M \geq \sqrt{n}(p-q) + \sigma_p N\right),$$

Where $M,N \sim \mathcal{N}(0,1)$ are independent. The proof follows. $\square$

6.3 Perturbation estimates for dense binomials

The main approximation that we use to prove Proposition 3.1 is the following:

Lemma 6.3. *If $X \sim \mathrm{Binom}(m,p)$ then for any k and ℓ,*

$$\log \frac{\Pr(X = k+\ell)}{\Pr(X=k)} \le \ell \log \frac{mp}{k+1} + \ell \log \frac{m-k}{m-mp}.$$

Proof. We compute

$$\begin{aligned}
\log \frac{\Pr(X = k+\ell)}{\Pr(X=k)} &= \log \frac{\binom{m}{k+\ell} p^\ell}{\binom{m}{k}(1-p)^\ell} \\
&= \ell \log \frac{p}{1-p} + \sum_{i=1}^{\ell} (\log(m-k-i+1) - \log(k+i)) \\
&\le \ell \log \frac{p}{1-p} + \ell \log(m-k) - \ell \log(k+1) \\
&= \ell \log \frac{mp}{k+1} + \ell \log \frac{m-k}{m-mp}.
\end{aligned}$$ □

Proof of Proposition 3.1. Fix ℓ with $1 \le \ell \le \sqrt{mp \log m}$. We will focus on the proof of (3.2), since the proof of (3.1) is analogous. We may write

$$\Pr(Y \ge X - \ell) = \sum_{k=-\ell}^{m} \Pr(Y \ge k) \Pr(X = k+\ell).$$

Now, Bernstein's inequality implies that by incurring a cost of $2m^{-2}$, we may restrict the sum to those k for which $mp - 3\sqrt{mp\log m} \le k + \ell \le mp + 3\sqrt{mp \log m}$. Since $\ell \le \sqrt{mp \log m}$, it suffices to take $mp - 4\sqrt{mp\log m} \le k \le mp + 4\sqrt{mp\log m}$. Hence,

$$\Pr(Y \ge X - \ell) \le \sum_{k=\lfloor mp - 4\sqrt{mp\log m}\rfloor}^{\lceil mp+4\sqrt{mp \log m}\rceil} \Pr(Y \ge k)\Pr(X = k+\ell) + 2m^{-2}. \tag{6.3}$$

Now, under the assumption $mp \ge 64 \log m$, we have $mp - 4\sqrt{mp\log m} \ge mp/2$ and $mp + 4\sqrt{mp \log m} \le 3mp/2$. Consider the first term in the upper bound of Lemma 6.3:

$$\log \frac{mp}{k+1} \le \frac{|k+1-mp|}{\min\{k+1, mp\}} \le 16 \sqrt{\frac{\log m}{mp}} \tag{6.4}$$

where the last inequality used $|k - mp| \le 4\sqrt{mp \log m}$ and $k \ge mp/2$. The other term in the upper bound of Lemma 6.3 is similar:

$$\log \frac{m-k}{m-mp} \le \frac{|k-mp|}{\min\{m-mp, m-k\}} \le C\sqrt{\frac{\log m}{mp}} \tag{6.5}$$

for sufficiently large m, where the second inequality follows by lower-bounding both terms in the denominator: $p \le 2/3$ implies $m - mp \ge 2mp$ and $k \le mp + 4\sqrt{mp\log m}$ implies $m - k \ge cmp$ for some $c > 0$ and sufficiently large m (this follows by considering the cases $p \in [2^{-10}, 2/3]$ and $p \in [64m^{-1}\log m, 2^{-10}]$ separately). Combining (6.4) and (6.5) with Lemma 6.3, we obtain

$$\log \frac{\Pr(X = k+\ell)}{\Pr(X=k)} \le C\ell \sqrt{\frac{\log m}{mp}}. \tag{6.6}$$

Applying this to (6.3), we have

$$\begin{aligned}\Pr(Y \geq X - \ell) &\leq \exp\left(C\ell\sqrt{\frac{\log m}{mp}}\right) \sum_{k=\lfloor mp-4\sqrt{mp\log m}\rfloor}^{\lceil mp+4\sqrt{mp\log m}\rceil} \Pr(Y \geq k)\Pr(X = k) + 2m^{-2} \\ &\leq \Pr(Y \geq X)\exp\left(C\ell\sqrt{\frac{\log m}{mp}}\right) + 2m^{-2}.\end{aligned}$$

The lower bound (i.e. (3.1)) is essentially the same, and we give only a sketch: we write

$$\Pr(Y \geq X + \ell) \geq \sum_{k=\lfloor mp-4\sqrt{mp\log m}\rfloor}^{\lceil mp+4\sqrt{mp\log m}\rceil} \Pr(Y \geq k+\ell)\Pr(X = k).$$

We then use (6.6) to compare $\Pr(X = k)$ with $\Pr(X = k + \ell)$. This leaves us with a sum over $k \in mp \pm 4\sqrt{mp\log m}$, which we compare with the full sum using Bernstein's inequality (picking up an additive $2m^{-2}$ term). □

6.4 Perturbation estimates for sparse binomials

The sparse case needs a slightly different argument and has slightly worse bounds. We have the following analogue of Lemma 6.3:

Lemma 6.4. *If $mp \leq 128\log m$ and $k = o(m)$ then for sufficiently large m and any $\ell \geq 1$,*

$$\log\frac{\Pr(X = k+\ell)}{\Pr(X = k)} \leq \ell\log\frac{mp}{\ell} + 2\ell$$

Proof. As in the proof of Lemma 6.3, we compute

$$\begin{aligned}\log\frac{\Pr(X = k+\ell)}{\Pr(X = k)} &= \ell\log\frac{p}{1-p} + \sum_{i=1}^{\ell}(\log(m-k-i+1) - \log(k+i)) \\ &\leq \ell\log\frac{p}{1-p} + \ell\log(m-k) - \sum_{i=1}^{\ell}\log(k+i).\end{aligned}$$

This time, we will use a sharper bound on the sum: since the logarithm is an increasing function,

$$\begin{aligned}\sum_{i=1}^{\ell}\log(k+i) &\geq \int_k^{k+\ell}\log(x)\,dx \\ &= (k+\ell)\log(k+\ell) - (k+\ell) - k\log k + k \\ &\geq \ell\log(k+\ell) - \ell.\end{aligned}$$

Hence, we obtain

$$\log\frac{\Pr(X = k+\ell)}{\Pr(X = k)} \leq \ell\log\frac{mp}{k+\ell} + \ell\log\frac{m-k}{m-mp} + \ell.$$

Since k and mp are $o(m)$, $\log((m-k)/(m-mp)) = o(1)$, and so

$$\log\frac{\Pr(X = k+\ell)}{\Pr(X = k)} \leq \ell\log\frac{mp}{\ell} + 2\ell$$

for sufficiently large m. □

Proof of Proposition 3.2. This proof is similar to the proof of Proposition 3.1, but with Lemma 6.4 instead of Lemma 6.3 and some slightly different truncations: we write

$$\Pr(Y \geq X - \ell) = \Pr(X \leq \ell - 1) + \sum_{k=0}^{m} \Pr(Y \geq k) \Pr(X = k + \ell)$$

By Bernstein's inequality, we may truncate the sum at $\sqrt{m}$ at the cost of an additive $e^{-c\sqrt{m}}$ term. We apply the inequality

$$\frac{\Pr(X = k + \ell)}{\Pr(X = k)} \leq \left(\frac{e^2 mp}{\ell}\right)^{\ell} \leq \left(\frac{C \log m}{\ell}\right)^{\ell}$$

(which follows from Lemma 6.4) to each term in the sum, yielding

$$\sum_{k=0}^{m} \Pr(Y \geq k) \Pr(X = k + \ell) \leq \left(\frac{C \log m}{\ell}\right)^{\ell} \Pr(Y \geq X) + e^{-c\sqrt{m}}.$$

We may also apply Lemma 6.4 to bound the term $\Pr(X \leq \ell - 1)$, using

$$\begin{aligned}
\Pr(X \leq \ell - 1) &= \sum_{s=0}^{\ell-1} \Pr(X = s) \\
&\leq \sum_{s=0}^{\ell-1} \left(\frac{C \log m}{s}\right)^{s} \Pr(X = 0) \\
&\leq \ell \left(\frac{C \log m}{\ell}\right)^{\ell} \Pr(X = 0) \\
&\leq \left(\frac{C \log m}{\ell}\right)^{C\ell} \Pr(X = 0),
\end{aligned}$$

where the second inequality follows (assuming $C \geq e$) because $(ey/x)^x$ is an increasing function of x for $x \leq y$. Putting everything together,

$$\Pr(Y \geq X - \ell) \leq \left(\frac{C \log m}{\ell}\right)^{C\ell} \Pr(X = 0) + \left(\frac{C \log m}{\ell}\right)^{\ell} \Pr(Y \geq X) + e^{-c\sqrt{m}}.$$

Finally, note that $\Pr(X = 0) \leq \Pr(Y \geq X)$ so that the first two terms above may be combined at the cost of increasing C. For the additive term $e^{-c\sqrt{m}}$, note that $mp \leq 128 \log m$ implies that $\Pr(Y \geq X) \geq \Pr(X = 0) = \Omega(n^{-\alpha})$ for some constant α, and so $e^{-c\sqrt{m}}$ may also be absorbed into the main term at the cost of increasing C. □

References

[1] E. Abbe, A. S. Bandeira, and G. Hall. Exact recovery in the stochastic block model. arXiv:1405.3267.

[2] Arash A. Amini, Aiyou Chen, Peter J. Bickel, and Elizaveta Levina. Pseudo-likelihood methods for community detection in large sparse networks. *The Annals of Statistics*, 41(4):2097–2122, 08 2013. MR-3127859

[3] 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):21068–21073, 2009.

[4] R.B. Boppana. Eigenvalues and graph bisection: An average-case analysis. In *28th Annual Symposium on Foundations of Computer Science*, pages 280–285. IEEE, 1987.

[5] T.N. Bui, S. Chaudhuri, F.T. Leighton, and M. Sipser. Graph bisection algorithms with good average case behavior. *Combinatorica*, 7(2):171–191, 1987. MR-0905164

[6] T. Carson and R. Impagliazzo. Hill-climbing finds random planted bisections. In *Twelfth annual ACM-SIAM symposium on Discrete algorithms*, pages 903–909. Society for Industrial and Applied Mathematics, 2001. MR-1958566

[7] A. Coja-Oghlan. Graph partitioning via adaptive spectral techniques. *Combinatorics, Probability and Computing*, 19(02):227–284, 2010. MR-2593622

[8] A. Condon and R.M. Karp. Algorithms for graph partitioning on the planted partition model. *Random Structures and Algorithms*, 18(2):116–140, 2001. MR-1809718

[9] M.E. Dyer and A.M. Frieze. The solution of some random NP-hard problems in polynomial expected time. *Journal of Algorithms*, 10(4):451–489, 1989. MR-1022107

[10] Paul Erdős and Alfréd Rényi. On the strength of connectedness of a random graph. *Acta Mathematica Hungarica*, 12(1):261–267, 1961. MR-0130187

[11] Theodore E. Harris. A lower bound for the critical probability in a certain percolation process. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 56, pages 13–20. Cambridge Univ. Press, 1960. MR-0115221

[12] P.W. Holland, K.B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983. MR-0718088

[13] M. Jerrum and G.B. Sorkin. The Metropolis algorithm for graph bisection. *Discrete Applied Mathematics*, 82(1-3):155–175, 1998. MR-1609981

[14] R. Karp. Reducibility among combinatorial problems. In R. Miller and J. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, 1972. MR-0378476

[15] János Komlós and Endre Szemerédi. Limit distribution for the existence of hamiltonian cycles in a random graph. *Discrete Mathematics*, 43(1):55–63, 1983. MR-0680304

[16] Amit Kumar and Ravindran Kannan. Clustering with spectral norm and the k-means algorithm. In *51st Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2010*, pages 299–308. IEEE, 2010. MR-3025203

[17] Konstantin Makarychev, Yury Makarychev, and Aravindan 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. MR-2961518

[18] Konstantin Makarychev, Yury Makarychev, and Aravindan Vijayaraghavan. Constant factor approximation for balanced cut in the pie model. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 41–49. ACM, 2014. MR-3238929

[19] Laurent 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. MR-3238997

[20] F. McSherry. Spectral partitioning of random graphs. In *42nd IEEE Symposium on Foundations of Computer Science*, pages 529–537. IEEE, 2001. MR-1948742

[21] E. Mossel, J. Neeman, and A. Sly. Belief propagation, robust reconstruction, and optimal recovery of block models (extended abstract). *JMLR Workshop and Conference Proceedings (COLT proceedings)*, 35:1–35, 2014. Winner of best paper award at COLT 2014.

[22] E. Mossel, J. Neeman, and A. Sly. Stochastic block models and reconstruction. *Probability Theory and Related Fields*, 2014. (to appear). MR-3383334

[23] Elchanan Mossel, Joe Neeman, and Allan Sly. A proof of the block model threshold conjecture. (submitted to Combinatorica), 2014.

[24] Raj Rao Nadakuditi and Mark E.J. Newman. Graph spectra and the detectability of community structure in networks. *Physical Review Letters*, 108(18):188701, 2012.

[25] Yoav Seginer. The expected norm of random matrices. *Combinatorics, Probability and Computing*, 9:149–166, 3 2000. MR-1762786

[26] Van H. Vu. Spectral norm of random matrices. *Combinatorica*, 27(6):721–736, 2007. MR-2384414

[27] Se-Young Yun and Alexandre Proutiere. Community detection via random and adaptive sampling. arXiv:1402.3072, 2014.