Find Your Place: Simple Distributed Algorithms for Community Detection *

Luca Becchetti
Sapienza Università di Roma
Rome, Italy
becchetti@dis.uniroma1.it

Andrea Clementi
Università di Roma Tor Vergata
Rome, Italy
clementi@mat.uniroma2.it

Emanuele Natale
Max Planck Institute for Informatics
Saarbrücken, Germany
emanuele.natale@mpi-inf.mpg.de

Francesco Pasquale
Università di Roma Tor Vergata
Rome, Italy
pasquale@mat.uniroma2.it

Luca Trevisan
U.C. Berkeley
Berkeley, CA, United States
luca@berkeley.edu

Abstract

Given an underlying graph, we consider the following *dynamics*: Initially, each node locally chooses a value in $\{-1, 1\}$, uniformly at random and independently of other nodes. Then, in each consecutive round, every node updates its local value to the average of the values held by its neighbors, at the same time applying an elementary, local clustering rule that only depends on the current and the previous values held by the node.

We prove that the process resulting from this dynamics produces a clustering that exactly or approximately (depending on the graph) reflects the underlying cut in logarithmic time, under various graph models that exhibit a sparse balanced cut, including the stochastic block model. We also prove that a natural extension of this dynamics performs community detection on a regularized version of the stochastic block model with multiple communities.

Rather surprisingly, our results provide rigorous evidence for the ability of an extremely simple and natural dynamics to address a computational problem that is non-trivial even in a centralized setting.

Keywords: Distributed Algorithms, Averaging Dynamics, Community Detection, Spectral Analysis, Stochastic Block Models.

1 Introduction

Consider the following distributed algorithm on an undirected graph: At the outset, every node picks an initial value, independently and uniformly at random in $\{-1, 1\}$; then, in each synchronous round, every node updates its value to the average of those held by its neighbors. A node also tags itself "blue" if the last update increased its value, "red" otherwise.

We show that under various graph models exhibiting sparse balanced cuts, including the *stochastic block model* [28], the process resulting from the above simple local rule converges, in logarithmic time, to a coloring that exactly or approximately (depending on the model) reflects the underlying cut. We further show that our approach simply and naturally extends to more communities, providing a quantitative analysis for a regularized version of the stochastic block model with multiple communities.

Our algorithm is one of the few examples of a *dynamics* [4, 3, 21, 47] that solves a computational problem that is non-trivial in a centralized setting. By *dynamics* we here mean synchronous distributed algorithms characterized by a very simple structure, whereby the state of a node at round t depends only on its state and a symmetric function of the multiset of states of its neighbors at round $t-1$, while the update rule is the same for every graph and every node and does not change over time. Note that this definition implies that the network is *anonymous*, that is, nodes do not possess distinguished identities. Examples of dynamics include update rules in which every node updates its state to the plurality or the median of the states of its neighbors,[1] or, as is

---

*This work is partly supported by the EU FET project MULTIPLEX no. 317532 and by the National Science Foundation under Grants No. CCF 1540685 and CCF 1655215

[1] When states correspond to rational values.

the case in this paper, every node holds a value, which it updates to the average of the values held by its neighbors. In contrast, an algorithm that, say, proceeds in two phases, using averaging during the first $10 \log n$ rounds and plurality from round $1 + 10 \log n$ onward, with n the number of nodes, is not a dynamics according to our definition, since its update rule depends on the size of the graph. As another example, an algorithm that starts by having the lexicographically first vertex elected as "leader" and then propagates its state to all other nodes again does not meet our definition of dynamics, since it assigns roles to the nodes and requires them to possess distinguishable identities.

The AVERAGING dynamics, in which each node updates its value to the average of its neighbors, is perhaps one of the simplest and most interesting examples of linear dynamics and it always converges when G is connected and not bipartite: It converges to the global average of the initial values if the graph is regular and to a weighted global average if it isn't [13, 50]. Important applications of linear dynamics have been proposed in the recent past [31, 5, 53, 33], for example to perform basic tasks such as self-stabilizing *consensus* in faulty distributed systems [8, 55, 48]. The convergence time of the AVERAGING dynamics is the mixing time of a random walk on G [50]. It is logarithmic in $|V|$ if the underlying graph is a good *expander* [29], while it is slower on graphs that exhibit sparse cuts.

While previous work on applications of linear dynamics has focused on tasks that are specific to distributed computing (such as reaching consensus, or stability in the presence of faulty nodes), in this paper we show that our simple protocol based on the the AVERAGING dynamics is able to address community detection, i.e., it identifies partition (V_1, V_2) of a clustered graph $G = ((V_1, V_2), E)$, either exactly (in which case we have a *strong* reconstruction algorithm) or approximately (in which case we speak of a *weak* reconstruction algorithm).

1.1 Our contributions. Consider a graph $G = (V, E)$. We show that, if a partition (V_1, V_2) of G exists, such that $\mathbf{1}_{V_1} - \mathbf{1}_{V_2}$ is[2] (or is close to) a right-eigenvector of the second largest eigenvalue of the transition matrix of G, and the gap between the second and the third largest eigenvalues is sufficiently large, our algorithm identifies the partition (V_1, V_2), or a close approximation thereof, in a logarithmic number of rounds. Though the algorithm we propose does not explicitly perform any eigenvector computation, its exploits the spectral structure of the underlying graph, based on the intuition that the dynamics is a distributed simulation of the power method. Our analysis involves two main novelties, relating to how nodes assign themselves to clusters, and to the spectral bounds that we prove for certain classes of graphs.

A conceptual contribution is to make each node, at each round t, assign itself to a cluster ("find its place") by considering the difference between its value at time t and its value at time $t-1$. Such a criterion removes the component of the value lying in the first eigenspace without explicitly computing it. This idea has two advantages: it allows a particularly simple algorithm, and it gives a running time that depends on the third eigenvalue of the transition matrix of the graph. In graphs that have the structure of two expanders joined by a sparse cut, the running time of the algorithm depends only on the expansion of the components and it is faster than the mixing time of the overall graph. To the best of our knowledge, this is the first distributed block reconstruction algorithm converging faster than the mixing time.

Our algorithm works on any graph where (i) the right-eigenspace of the second eigenvalue of the transition matrix is correlated to the cut between the two clusters and (ii) the gap between the second and third eigenvalues is sufficiently large. While these conditions have been investigated for the spectrum of the *adjacency* matrix of the graph, our analysis requires these conditions to hold for the *transition* matrix. A technical contribution of this paper is to show that such conditions are met by a large class of graphs, that includes graphs sampled from the *stochastic block model*. Proving spectral properties of the transition matrix of a random graph is more challenging than proving such properties for the adjacency matrix, because the entries of the transition matrix are not independent.

Strong reconstruction for regular clustered graphs. A $(2n, d, b)$-clustered regular graph $G = ((V_1, V_2), E)$ is a connected graph over vertex set $V_1 \cup V_2$, with $|V_1| = |V_2| = n$, adjacency matrix A, and such that every node has degree d and it has (exactly) b neighbors outside its cluster. If the two subgraphs induced by V_1 and V_2 are good expanders and b is sufficiently small, the second and third eigenvalues of the graph's transition matrix $P = (1/d) \cdot A$ are separated by a large gap. In this case, we can prove that the following happens with high

---

[2] As explained further, $\mathbf{1}_{V_i}$, is the vector with $|V|$ components, such that the j-th component is 1 if $j \in V_i$, it is 0 otherwise.

probability (for short *w.h.p*[3]): If the AVERAGING dynamics is initialized by having every node choose a value uniformly and independently at random in $\{-1, 1\}$, within a logarithmic number of rounds the system enters a regime in which nodes' values are monotonically increasing or decreasing, depending on the community they belong to. As a consequence, every node can apply a simple and completely local clustering rule in each round, which eventually results in a strong reconstruction (Theorem 3.1).

We then show that, under mild assumptions, a graph selected from the *regular stochastic block model* [14] is a $(2n, d, b)$-clustered regular graph that satisfies the above spectral gap hypothesis, w.h.p. We thus obtain a fast and extremely simple dynamics for strong reconstruction, over the full range of parameters of the regular stochastic block model for which this is known to be possible using centralized algorithms [45, 14] (Section 1.2 and Corollary 3.1).

We further show that a natural extension of our algorithm, in which nodes maintain an array of values and an array of colors, correctly identifies a hidden balanced k-partition in a regular clustered graph with a gap between λ_k and λ_{k+1}. Graphs sampled from the regular stochastic block model with k communities satisfy such conditions, w.h.p. (Theorem 5.1).

Weak reconstruction for non-regular clustered graphs. As a main technical contribution, we extend the above analysis to show that our dynamics also ensures weak reconstruction in clustered graphs having two clusters that satisfy an approximate regularity condition and a gap between second and third eigenvalues of the transition matrix P (Theorem 4.1). As an application, we then prove that these conditions are met by the *stochastic block model* [1, 18, 19, 22, 28, 30, 41], a random graph model that offers a popular framework for the probabilistic modelling of graphs that exhibit good clustering or community properties. We here consider its simplest version, i.e., the random graph $\mathcal{G}_{2n,p,q}$ consisting of $2n$ nodes and an edge probability distribution defined as follows: The node set is partitioned into two subsets V_1 and V_2, each of size n; edges linking nodes belonging to the same partition appear in E independently at random with probability $p = p(n)$, while edges connecting nodes from different partitions appear with probability $q = q(n) < p$. Calling $a = pn$ and $b = qn$, we prove that graphs sampled from $\mathcal{G}_{2n,p,q}$ satisfy the above approximate regularity and spectral gap conditions, w.h.p., whenever $a - b > \sqrt{(a+b) \cdot \log n}$ (Lemma 4.2).

We remark that the latter result for the stochastic block model follows from an analysis that applies to general *non-random* clustered graphs and hence does not exploit crucial properties of random graphs. A further technical contribution of this paper is a refined, ad-hoc analysis of the AVERAGING dynamics for the $\mathcal{G}_{2n,p,q}$ model, showing that this protocol achieves weak-reconstruction, correctly classifying a $1 - \varepsilon$ fraction of vertices, in logarithmic time whenever $a - b > \Omega_\varepsilon(\sqrt{(a+b)})$ and the expected degree $d = a + b$ grows at least logarithmically (Theorem 4.2). This refined analysis requires a deeper understanding of the eigenvectors of the *transition matrix* of G. Coja-Oghlan [18] defined certain graph properties that guarantee that a near-optimal bisection can be found based on eigenvector computations of the *adjacency matrix*. Similarly, we show simple sufficient conditions under which a right eigenvector of the second largest eigenvalue of the transition matrix of a graph approximately identifies the hidden partition. We give a tight analysis of the spectrum of the transition matrix of graphs sampled from the stochastic block model in Section D.2. Notice that the analysis of the transition matrix is somewhat harder than that of the adjacency matrix, since the entries are not independent of each other; we were not able to find comparable results in the existing literature.

1.2 Related work and additional remarks

Dynamics for block reconstruction. Dynamics received considerable attention in the recent past across different research communities, both as efficient distributed algorithms [4, 8, 48, 42] and as abstract models of *natural* interaction mechanisms inducing emergent behavior in complex systems [3, 15, 21, 24, 47]. For instance, simple averaging dynamics have been considered to model opinion formation mechanisms [20, 25], while a number of other dynamics have been proposed to describe different social phenomena [23]. *Label propagation algorithms* [49] are dynamics based on majority updating rules [4] and have been applied to some computational problems including clustering. Several papers present experimental results for such protocols on specific classes of clustered graphs [6, 38, 49]. The only available rigorous analysis of label propagation algorithms on planted partition graphs is the one presented in [34], where the authors propose and analyze a label propagation protocol on $\mathcal{G}_{2n,p,q}$ for dense topologies. In particular, their analysis considers the case where $p = \Omega(1/n^{1/4-\varepsilon})$ and $q = \mathcal{O}(p^2)$, a parameter range in which very dense clusters of constant di-

---

[3] We say that a sequence of events $\mathcal{E}_n$, $n = 1, 2, \ldots$ holds *with high probability* if $\mathbf{P}(\mathcal{E}_n) = 1 - \mathcal{O}(1/n^\gamma)$ for some positive constant $\gamma > 0$.

ameter separated by a sparse cut occur w.h.p. In this setting, characterized by a polynomial gap between p and q, simple combinatorial and concentration arguments show that the protocol converges in constant expected time. They also conjecture a logarithmic bound for sparser topologies.

Because of their relevance for the reconstruction problem, we also mention another class of algorithms, *belief propagation algorithms*, whose simplicity is close to that of dynamics. *Belief propagation algorithms* are best known as message-passing algorithms for performing inference in graphical models [39]. Belief propagation cannot be considered a dynamics: At each round, each node sends a different message to each neighbors, thus the update rule is not symmetric w.r.t. the neighbors, requiring thus port numbering [52], and the required local memory grows linearly in the degree of the node. Non-rigorous methods have given strong evidence that some *belief propagation algorithms* are optimal for the reconstruction problem [19]. Its rigorous analysis is a major challenge; in particular, the convergence to the correct value of belief propagation is far from being fully-understood on graphs which are not trees [54, 43]. As we discuss in the next subsection, more complex algorithms, inspired by belief propagation, have been rigorously shown to perform reconstruction optimally.

General algorithms for block reconstruction. While an important goal, improving performance of spectral clustering algorithms and testing their limits to the purpose of block reconstruction is not the main driver behind this work. Still, for the sake of completeness, we next compare our dynamics to previous general algorithms for block reconstruction.

Several algorithms for community detection are *spectral*: They typically consider the eigenvector associated to the second eigenvalue of the adjacency matrix A of G, or the eigenvector corresponding to the largest eigenvalue of the matrix $A - \frac{d}{n}J$ [9, 17, 18, 41][4], since these are correlated with the hidden partition. More recently spectral algorithms have been proposed [2, 18, 44, 35, 12] that find a weak reconstruction even in the sparse, tight regime.

Even though the above mentioned algorithms have been presented in a centralized setting, spectral algorithms turn out to be a feasible approach also for distributed models. Indeed, Kempe and McSherry [32] show that eigenvalue computations can be performed in a distributed fashion, yielding distributed algorithms for community detection in various models, including the stochastic block model. However, the algorithm of Kempe and McSherry as well as any distributed version of the above mentioned centralized algorithms are not dynamics. Actually, adopting the effective concept from Hassin and Peleg in [27], such algorithms are even not *light-weight*: Different and not-simple operations are executed at different rounds, nodes have identities, messages are treated differently depending on the originator, and so on. Moreover, a crucial aspect is convergence time: The mixing time of the simple random walk on the graph is a bottleneck for the distributed algorithm of Kempe and McSherry and for any algorithm that performs community detection in a graph G by employing the power method or the Lanczos method [36] as a subroutine to compute the eigenvector associated to the second eigenvalue of the adjacency matrix of G. Notice that the mixing time of graphs sampled from $\mathcal{G}_{2n,p,q}$ is at least of the order of $\frac{a+b}{2b}$: hence, it can be super-logarithmic and even $n^{\Omega(1)}$.

In general, the reconstruction problem has been studied extensively using a multiplicity of techniques, which include combinatorial algorithms [22], belief propagation [19] and variants of it [46], spectral-based techniques [41, 18], Metropolis approaches [30], and semidefinite programming [1], among others. Stochastic Block Models have been studied in a number of areas, including computer science [9, 41, 40], probability theory [45], statistical physics [19], and social sciences [28]. Unlike the distributed setting, where the existence of *light-weight protocols* [27] is the main issue (even in non-sparse regimes), in centralized setting strong attention has been devoted to establishing sharp thresholds for weak and strong reconstruction. Define $a = np$ as the expected *internal degree* (the number of neighbors that each node has on the same side of the partition) and $b = nq$ as the expected *external degree* (the number of neighbors that each node has on the opposite side of the partition). Decelle et al. [19] conjectured that weak reconstruction is possible if and only if $a - b > 2\sqrt{a+b}$. This was proved by Massoulie and Mossel et al. [44, 40, 45]. Strong recovery is instead possible if and only if $a - b > 2\sqrt{a+b} + \log n$ [1].

Versions of the stochastic block model in which the random graph is regular have also been considered [45, 14]. In particular Brito et al. [14] show that strong reconstruction is possible in polynomial-time when $a - b > 2\sqrt{a+b-1}$.

---

[4] A is the adjacency matrix of G, J is the matrix having all entries equal to 1, d is the average degree and $2n$ is the number of vertices.

2 Preliminaries

Distributed block reconstruction. Let $G = ((V_1, V_2), E)$ be a graph with $V_1 \cap V_2 = \emptyset$. A weak (block) reconstruction is a two-coloring of the nodes that separates V_1 and V_2 up to a small fraction of the nodes. Formally, we define an *ε-weak reconstruction* as a map $f : V_1 \cup V_2 \to \{\texttt{red}, \texttt{blue}\}$ such that there are two subsets $W_1 \subseteq V_1$ and $W_2 \subseteq V_2$ with $|W_1 \cup W_2| \geqslant (1-\varepsilon)|V_1 \cup V_2|$ and $f(W_1) \cap f(W_2) = \emptyset$. When $\varepsilon = 0$ we say that f is a *strong reconstruction*. Given a graph $G = ((V_1, V_2), E)$, the block reconstruction problem requires computing an ε-reconstruction of G.

In this paper, we propose the following distributed protocol. It is based on the averaging dynamics and produces a coloring of the nodes at the end of every round. In the next two sections we show that, within $\mathcal{O}(\log n)$ rounds, the coloring computed by the algorithm we propose achieves *strong reconstruction* of the two blocks in the case of clustered regular graphs and *weak reconstruction* in the case of clustered non-regular graphs.

Averaging protocol:

Rademacher initialization: At round $t = 0$ every node $v \in V$ independently samples its value from $\{-1, +1\}$ uniformly at random;

Updating rule: At each subsequent round $t \geqslant 1$, every node $v \in V$

1. (Averaging dynamics) Updates its value $\mathbf{x}^{(t)}(v)$ to the average of the values of its neighbors at the end of the previous round,
2. (Coloring) If $\mathbf{x}^{(t)}(v) \geqslant \mathbf{x}^{(t-1)}(v)$ then v sets $\texttt{color}^{(t)}(v) = \texttt{blue}$ otherwise v sets $\texttt{color}^{(t)}(v) = \texttt{red}$.

The choice of the above coloring rule will be clarified in the next section, just before Theorem 3.1. We give here two remarks. First of all, the algorithm is completely oblivious to time, being a dynamics in the strictest sense. Namely, after initialization the protocol iterates over and over at every node. Convergence to a (possibly weak) reconstruction is a property of the protocol, of which nodes are not aware, it is something that eventually occurs. Second, the clustering criterion is completely *local*, in the sense that a decision is individually and independently made by each node in each round, only on the basis of its state in the current and previous rounds. This may seem counterintuitive at first, but it is only superficially so. Despite being local, the clustering criterion uses information that reflects the global structure of the network, since nodes' values are related to the second eigenvector of the network's transition matrix.

The Averaging dynamics and random walks on G. The analysis of the Averaging dynamics on a graph G is closely related to the behavior of random walks in G, which are best studied using tools from linear algebra that we briefly summarize below.

Let $G = (V, E)$ be an undirected graph (possibly with multiple edges and self loops), A its adjacency matrix and d_i the degree of node i. The *transition matrix* of (the random walk on) G is the matrix $P = D^{-1}A$, where D is the diagonal matrix such that $D_{i,i} = d_i$. $P_{i,j} = (1/d_i) \cdot A_{i,j}$ is thus the probability of going from i to j in one-step of the random walk on G. P operates as the random walk process on G by left multiplication, and as the Averaging dynamics by right multiplication. For $i = 1, 2$, define $\mathbf{1}_{V_i}$, as the $|V|$-dimensional vector, whose j-th component is 1 if $j \in V_i$, it is 0 otherwise. If (V_1, V_2) is a bipartition of the nodes with $|V_1| = |V_2| = n$, we define the *partition indicator vector* $\boldsymbol{\chi} = \mathbf{1}_{V_1} - \mathbf{1}_{V_2}$. If $\mathbf{x}$ is the initial vector of values, after t rounds of the Averaging dynamics the vector of values at time t is $\mathbf{x}^{(t)} = P^t\mathbf{x}$. The product of the power of a matrix times a vector is best understood in terms of the spectrum of the matrix, which is what we explore in the next section.

In what follows we always denote by $\lambda_1 \geqslant \ldots \geqslant \lambda_{2n}$ the eigenvalues of P. Recall that, since P is a stochastic matrix we have $\lambda_1 = 1$ and $\lambda_{2n} \geqslant -1$, moreover for all graphs that are connected and not bipartite it holds that $\lambda_2 < 1$ and $\lambda_{2n} > -1$. We denote by λ the largest, in absolute value, among all but the first two eigenvalues, namely $\lambda = \max\{|\lambda_i| : i = 3, 4, \ldots, 2n\}$. Unless otherwise specified, the norm of a vector $\mathbf{x}$ is the ℓ_2 norm $\|\mathbf{x}\| := \sqrt{\sum_i (\mathbf{x}(i))^2}$ and the norm of a matrix A is the spectral norm $\|A\| := \sup_{\mathbf{x}:\|\mathbf{x}\|=1} \|A\mathbf{x}\|$. For a diagonal matrix, this is the largest diagonal entry in absolute value.

3 Strong reconstruction for regular graphs

If G is d-regular then $P = (1/d)A$ is a real symmetric matrix and P and A have the same set of eigenvectors. We denote by $\mathbf{v}_1 = (1/\sqrt{2n})\mathbf{1}, \mathbf{v}_2, \ldots, \mathbf{v}_{2n}$ a basis of orthonormal eigenvectors, where each $\mathbf{v}_i$ is the eigenvector associated to eigenvalue λ_i. Then, we can write a vector $\mathbf{x}$ as a linear combination $\mathbf{x} = \sum_i \alpha_i \mathbf{v}_i$ and we have:

$$P^t\mathbf{x} = \sum_i \lambda_i^t \alpha_i \mathbf{v}_i = \frac{1}{2n}\left(\sum_i \mathbf{x}(i)\right)\mathbf{1} + \sum_{i=2}^{2n} \lambda_i^t \alpha_i \mathbf{v}_i,$$

which implies that $\mathbf{x}^{(t)} = P^t \mathbf{x}$ tends to $\alpha_1 \mathbf{v}_1$ as t tends to infinity, i.e., it converges to the vector that has the average of $\mathbf{x}$ in every coordinate.

We next show that, if the regular graph is "well" clustered, then the AVERAGING protocol produces a strong reconstruction of the two clusters w.h.p.

DEFINITION 1. (CLUSTERED REGULAR GRAPH) *A $(2n, d, b)$-clustered regular graph $G = ((V_1, V_2), E)$ is a graph over vertex set $V_1 \cup V_2$, with $|V_1| = |V_2| = n$ and such that: (i) Every node has degree d and (ii) Every node in cluster V_1 has b neighbors in cluster V_2 and every node in V_2 has b neighbors in V_1.*

We know that $\mathbf{1}$ is an eigenvector of P with eigenvalue 1, and it is easy to see that the partition indicator vector $\boldsymbol{\chi}$ is an eigenvector of P with eigenvalue $1 - 2b/d$ (see Observation 2 in Appendix A). We first show that, if $1 - 2b/d$ happens to be the second eigenvalue, after t rounds of the AVERAGING dynamics, the configuration $\mathbf{x}^{(t)}$ is close to a linear combination of $\mathbf{1}$ and $\boldsymbol{\chi}$. Formally, if $\lambda < 1 - 2b/d$ we prove (see Lemma C.1 in Appendix C) that there are reals α_1, α_2 such that for every t

$$\mathbf{x}^{(t)} = \alpha_1 \mathbf{1} + \alpha_2 \lambda_2^t \boldsymbol{\chi} + \mathbf{e}^{(t)}, \tag{3.1}$$

where $\left\| \mathbf{e}^{(t)} \right\|_\infty \leqslant \lambda^t \sqrt{2n}$.

Informally speaking, the equation above naturally "suggested" the choice of the coloring rule in the AVERAGING protocol, once we considered the difference of two consecutive values of any node u, i.e.,

$$\begin{aligned} & \mathbf{x}^{(t-1)}(u) - \mathbf{x}^{(t)}(u) \\ & = \alpha_2 \lambda_2^{t-1} (1 - \lambda_2) \boldsymbol{\chi}(u) + \mathbf{e}^{(t-1)}(u) - \mathbf{e}^{(t)}(u) . \end{aligned} \tag{3.2}$$

Intuitively, if λ is sufficiently small, we can exploit the bound on $\left\| \mathbf{e}^{(t)} \right\|_\infty$ in (3.1) to show that, after a short initial phase, the sign of $\mathbf{x}^{(t-1)}(u) - \mathbf{x}^{(t)}(u)$ is essentially determined by $\boldsymbol{\chi}(u)$, thus by the community u belongs to, w.h.p. The following theorem and its proof provide formal statements of the above fact.

THEOREM 3.1. (STRONG RECONSTRUCTION) *Let $G = ((V_1, V_2), E)$ be a connected $(2n, d, b)$-clustered regular graph with $1 - 2b/d > (1 + \delta)\lambda$ for an arbitrarily-small constant $\delta > 0$. Then the* AVERAGING *protocol produces a strong reconstruction within $\mathcal{O}(\log n)$ rounds, w.h.p.*

Outline of Proof. From (3.2), we have that $\operatorname{sgn}\left(\mathbf{x}^{(t-1)}(u) - \mathbf{x}^{(t)}(u)\right) = \operatorname{sgn}\left(\alpha_2 \boldsymbol{\chi}(u)\right)$ whenever

$$\left| \alpha_2 \lambda_2^{t-1} (1 - \lambda_2) \right| > \left| \mathbf{e}^{(t-1)}(u) - \mathbf{e}^{(t)}(u) \right| \tag{3.3}$$

From (3.1) we have that $\left| \mathbf{e}^{(t)}(u) \right| \leqslant \lambda^t \sqrt{2n}$, thus (3.3) is satisfied for all t such that

$$t - 1 \geqslant \log \left(\frac{2\sqrt{2n}}{|\alpha_2|(1 - \lambda_2)} \right) \cdot \frac{1}{\log\left(\lambda_2 / \lambda\right)} .$$

The second key-step of the proof relies on the randomness of the initial vector. Indeed, since $\mathbf{x}$ is a vector of independent and uniformly distributed random variables in $\{-1, 1\}$, the absolute difference between the two partial averages in the two communities, i.e. $|\alpha_2|$, is "sufficiently" large, w.h.p. More precisely, from Lemma B.1 we have that is the sum of $2n$ Rademacher random variables, we have

$$\mathbf{P}\left(|R| \leqslant \delta \sqrt{2n} \right) \leqslant \mathcal{O}(\delta).$$

Since $\alpha_2 = \frac{1}{2n} \langle \boldsymbol{\chi}, \mathbf{x} \rangle$ and $\mathbf{x}$ is a vector of Rademacher random variables, the previous inequality implies that

$$|\alpha_2| = \frac{1}{2n} \langle \boldsymbol{\chi}, \mathbf{x} \rangle \geqslant n^{-\gamma},$$

for some positive constant γ, w.h.p. The theorem thus follows from the above bound on $|\alpha_2|$ and from the hypothesis $\lambda_2 \geqslant (1 + \delta)\lambda$. □

Remark. Graphs to which Theorem 3.1 apply are those consisting of two regular expanders connected by a regular sparse cut. Indeed, let $G = ((V_1, V_2), E)$ be a $(2n, d, b)$-clustered regular graph, and let $\lambda_A = \max\{\lambda_2(A_1), \lambda_2(A_2)\}$ and $\lambda_B = \lambda_2(B)$, where A_1, A_2 and B are the adjacency matrices of the subgraphs induced by V_1, V_2 and the cut between V_1 and V_2, respectively. Since $\lambda = \frac{a}{d}\lambda_A + \frac{b}{d}\lambda_B$, if $a - b > (1 + \varepsilon)(a\lambda_A + b\lambda_B)$, G satisfies the hypothesis of Theorem 3.1.

Regular stochastic block model. We can use Theorem 3.1 to prove that the AVERAGING protocol achieves strong reconstruction in the regular stochastic block model. In the case of two communities, a graph on $2n$ vertices is obtained as follows: Given two parameters $a(n)$ and $b(n)$ (*internal* and *external* degrees, respectively), partition vertices into two equal-sized subsets V_1 and V_2 and then sample a random $a(n)$-regular graph over each of V_1 and V_2 and a random $b(n)$-regular graph between V_1 and V_2. This model can be instantiated in different ways depending on how one samples the random regular graphs (for example, via the uniform distribution over regular graphs, or by taking the disjoint union of random matchings) [45, 14].

If G is a graph sampled from the regular stochastic block model with internal and external degrees a and b respectively, then it is a $(2n, d, b)$-clustered graph with largest eigenvalue of the transition matrix 1 and corresponding eigenvector $\mathbf{1}$, while $\boldsymbol{\chi}$ is also an eigenvector, with eigenvalue $1 - 2b/d$, where $d := a + b$. Furthermore, we can derive the following upper bound on the maximal absolute value achieved by the other $2n - 2$ eigenvalues corresponding to eigenvectors orthogonal to $\mathbf{1}$ and $\boldsymbol{\chi}$:

$$\lambda \leqslant \frac{2}{a+b}(\sqrt{a+b-1} + o_n(1)) \tag{3.4}$$

This bound can be proved using some general result of Friedman and Kohler [26] on *random degree k lifts* of a graph. (see Lemma D.1 in Appendix D). Since $\lambda_2 = \frac{a-b}{a+b}$, using (3.4) in Theorem 3.1, we get a strong reconstruction for the regular stochastic block model:

Corollary 3.1. *Let G be a random graph sampled from the regular stochastic block model with $a - b > 2(1 + \eta)\sqrt{a + b}$ for any constant $\eta > 0$, then the* Averaging *protocol produces a strong reconstruction in $\mathcal{O}(\log n)$ rounds, w.h.p.*

4 Weak reconstruction for non-regular graphs

The results of Section 3 rely on very clear spectral properties of regular, clustered graphs, immediately reflecting their underlying topological structure. Intuition suggests that these properties should be approximately preserved if we suitably relax the notion of regularity. With this simple intuition in mind, we generalize our approach for regular graphs to a large class of non-regular clustered graphs.

Definition 2. (Clustered γ-regular graphs) *A $(2n, d, b, \gamma)$-clustered graph $G = ((V_1, V_2), E)$ is a graph over vertex set $V_1 \cup V_2$, where $|V_1| = |V_2| = n$ such that: i) Every node has degree $d \pm \gamma d$, and ii) Every node in V_1 has $b \pm \gamma d$ neighbors in V_2 and every node in V_2 has $b \pm \gamma d$ neighbors in V_1.*

If G is not regular then matrix $P = D^{-1}A$ is not symmetric in general, however it is possible to relate its eigenvalues and eigenvectors to those of a symmetric matrix as follows. Denote the *normalized adjacency matrix* of G as $N := D^{-1/2}AD^{-1/2} = D^{1/2}PD^{-1/2}$. Notice that N is symmetric, P and N have the same eigenvalues $\lambda_1, \ldots, \lambda_{2n}$, and $\mathbf{x}$ is an eigenvector of P if and only if $D^{1/2}\mathbf{x}$ is an eigenvector of N (if G is regular then P and N are the same matrix). Let $\mathbf{w}_1, \ldots, \mathbf{w}_{2n}$ be a basis of orthonormal eigenvectors of N, with $\mathbf{w}_i$ the eigenvector associated to eigenvalue λ_i, for every i. We have that $\mathbf{w}_1 = \frac{D^{1/2}\mathbf{1}}{\|D^{1/2}\mathbf{1}\|}$. If we set $\mathbf{v}_i := D^{-1/2}\mathbf{w}_i$, we obtain a set of eigenvectors for P and we can write $\mathbf{x} = \sum_i \alpha_i \mathbf{v}_i$ as a linear combination of them. Then, the averaging process can again be described as

$$P^t\mathbf{x} = \sum_i \lambda_i^t \alpha_i \mathbf{v}_i = \alpha_1 \mathbf{v}_1 + \sum_{i=2}^{2n} \lambda_i^t \alpha_i \mathbf{v}_i .$$

So, if G is connected and not bipartite, the Averaging dynamics converges to $\alpha_1 \mathbf{v}_1$. In general, it is easy to see that $\alpha_i = \mathbf{w}_i^T D^{1/2}\mathbf{x}$ (see the first lines in the proof of Lemma 4.1) and $\alpha_1 \mathbf{v}_1$ is the vector

$$(\mathbf{w}_1^T D^{1/2}\mathbf{x}) \cdot D^{-1/2}\mathbf{w}_1 = \frac{\mathbf{1}^T D\mathbf{x}}{\|D^{1/2}\mathbf{1}\|^2}\mathbf{1} = \frac{\sum_i d_i \mathbf{x}(i)}{\sum_i d_i} \cdot \mathbf{1} .$$

As in the regular case, if the transition matrix P of a clustered γ-regular graph has λ_2 close to 1 and $|\lambda_3|, \ldots, |\lambda_{2n}|$ small, the Averaging dynamics has a long phase in which $\mathbf{x}^{(t)} = P^t\mathbf{x}$ is close to $\alpha_1\mathbf{1} + \alpha_2\mathbf{v}_2$.

However, providing an argument similar to the regular case is considerably harder, since the partition indicator vector $\boldsymbol{\chi}$ is no longer an eigenvector of P. In order to fix this issue, we generalize (3.1), proving in Lemma 4.1 that $\mathbf{x}^{(t)}$ is still close to a linear combination of $\mathbf{1}$ and $\boldsymbol{\chi}$. We set $\nu = 1 - \frac{2b}{d}$, since this value occurs frequently in this section.

Lemma 4.1. *Let G be a connected $(2n, d, b, \gamma)$-clustered graph with $\gamma \leqslant 1/10$, and assume the* Averaging *dynamics is run on G with initial vector $\mathbf{x}$. If $\lambda < \nu$ we have:*

$$\mathbf{x}^{(t)} = \alpha_1\mathbf{1} + \alpha_2\lambda_2^t\boldsymbol{\chi} + \alpha_2\lambda_2^t\mathbf{z} + \mathbf{e}^{(t)} ,$$

for some vectors $\mathbf{z}$ and $\mathbf{e}^{(t)}$ with $\|\mathbf{z}\| \leqslant \frac{88\,\gamma}{\nu - \lambda_3}\sqrt{2n}$ and $\|\mathbf{e}^{(t)}\| \leqslant 4\lambda^t\|\mathbf{x}\|$. Coefficients α_1 and α_2 are $\alpha_1 = \frac{\mathbf{1}^\intercal D\mathbf{x}}{\|D^{\frac{1}{2}}\mathbf{1}\|^2}$ and $\alpha_2 = \frac{\mathbf{w}_2^\intercal D^{\frac{1}{2}}\mathbf{x}}{\mathbf{w}_2^\intercal D^{\frac{1}{2}}\boldsymbol{\chi}}$.

Outline of Proof. We prove the following two key-facts: (i) the second eigenvalue of the transition matrix of G is not much smaller than $1 - 2b/d$, and (ii) $D^{1/2}\boldsymbol{\chi}$ is close, in norm, to its projection on the second eigenvector of the normalized adjacency matrix N. Namely, in Lemma C.2 we prove that if $\lambda_3 < \nu$ then

$$\begin{aligned} &\lambda_2 \geqslant \nu - 10\gamma \ \text{ and} \\ &\left\|D^{1/2}\boldsymbol{\chi} - \beta_2\mathbf{w}_2\right\| \leqslant \tfrac{44\,\gamma}{\nu - \lambda_3}\sqrt{2nd} , \end{aligned} \tag{4.5}$$

where $\beta_2 = \boldsymbol{\chi}^\intercal D^{1/2}\mathbf{w}_2$. Now, we can use the above bounds to analyze $\mathbf{x}^{(t)} = P^t\mathbf{x}$. To begin, note that $N = D^{-1/2}AD^{-1/2}$ and $P = D^{-1}A$ imply that $P = D^{-1/2}ND^{1/2}$ and $P^t = D^{-1/2}N^tD^{1/2}$. Thus, for any vector $\mathbf{x}$, if we write $D^{1/2}\mathbf{x}$ as a linear combination of an orthonormal basis of N, $D^{1/2}\mathbf{x} = \sum_{i=1}^{2n} a_i\mathbf{w}_i$, we get

$$P^t\mathbf{x} = D^{-1/2}N^tD^{1/2}\mathbf{x}$$

$$= D^{-1/2}\sum_{i=1}^{2n} a_i\lambda_i^t \mathbf{w}_i = \sum_{i=1}^{2n} a_i \lambda_i^t D^{-1/2}\mathbf{w}_i.$$

We next estimate the first term, the second term, and the sum of the remaining terms:

- We have $\mathbf{w}_1 = \frac{D^{1/2}\mathbf{1}}{\|D^{1/2}\mathbf{1}\|}$, so the first term can be written as $\alpha_1 \mathbf{1}$ with

$$\alpha_1 = \frac{a_1}{\left\|D^{1/2}\mathbf{1}\right\|} = \frac{\mathbf{w}_1^\intercal D^{1/2}\mathbf{x}}{\left\|D^{1/2}\mathbf{1}\right\|} = \frac{\mathbf{1}^\intercal D\mathbf{x}}{\left\|D^{1/2}\mathbf{1}\right\|^2}.$$

- If we write $D^{1/2}\boldsymbol{\chi} = \beta_2\mathbf{w}_2 + \mathbf{y}$, with $\beta_2 = \mathbf{w}_2^\intercal D^{1/2}\boldsymbol{\chi}$, (4.5) implies that $\|\mathbf{y}\| \leqslant \frac{44\,\gamma}{\nu-\lambda_3}\sqrt{2nd}$. Hence the second term can be written as

$$\begin{aligned} a_2\lambda_2^t D^{-1/2}\mathbf{w}_2 &= a_2\lambda_2^t D^{-1/2}\left(\frac{D^{1/2}\boldsymbol{\chi} - \mathbf{y}}{\beta_2}\right) \\ &= \frac{a_2}{\beta_2}\lambda_2^t\boldsymbol{\chi} - \frac{a_2}{\beta_2}\lambda_2^t\mathbf{z} = \alpha_2\lambda_2^t\boldsymbol{\chi} - \alpha_2\lambda_2^t\mathbf{z}, \end{aligned}$$

where

$$\begin{aligned} \|\mathbf{z}\| &= \left\|D^{-1/2}\mathbf{y}\right\| \leqslant \left\|D^{-1/2}\right\| \, \|\mathbf{y}\| \\ &\leqslant \frac{2}{\sqrt{d}} \cdot \frac{44\,\gamma}{\nu-\lambda_3}\sqrt{2nd} = \frac{88\,\gamma}{\nu-\lambda_3}\sqrt{2n}\,, \end{aligned}$$

and

$$\alpha_2 = a_2/\beta_2 = \frac{\mathbf{w}_2^\intercal D^{1/2}\mathbf{x}}{\mathbf{w}_2 D^{1/2}\boldsymbol{\chi}}.$$

- As for all other terms, observe that

$$\begin{aligned} \|\mathbf{e}^{(t)}\|^2 &= \left\|\sum_{i=3}^{2n} a_i\lambda_i^t D^{-1/2}\mathbf{w}_i\right\|^2 \\ &\leqslant \left\|D^{-1/2}\right\|^2 \left\|\sum_{i=3}^{2n} a_i\lambda_i^t\mathbf{w}_i\right\|^2 = \left\|D^{-1/2}\right\|^2 \sum_{i=3}^{2n} a_i^2\lambda_i^{2t} \\ &\leqslant \left\|D^{-1/2}\right\|^2 \lambda^{2t}\sum_{i=3}^{2n} a_i^2 \leqslant \left\|D^{-1/2}\right\|^2 \lambda^{2t}\left\|D^{1/2}\mathbf{x}\right\|^2 \\ &\leqslant \left\|D^{-1/2}\right\|^2 \left\|D^{1/2}\right\|^2 \lambda^{2t}\|\mathbf{x}\|^2 \leqslant 16\lambda^{2t}\|\mathbf{x}\|^2. \end{aligned}$$

□

The above lemma allows us to generalize our approach to achieve efficient, weak reconstruction in non-regular clustered graphs. The full proof of the following theorem is given in appendix C.1.

THEOREM 4.1. (WEAK RECONSTRUCTION) *Let G be a connected $(2n, d, b, \gamma)$-clustered graph with $\gamma \leqslant c(\nu - \lambda_3)$ for a suitable constant $c > 0$. If $\lambda < \nu$ and $\lambda_2 \geqslant (1+\delta)\lambda$ for an arbitrarily-small positive constant δ, then the* AVERAGING *protocol produces an $\mathcal{O}(\gamma^2/(\nu-\lambda_3)^2)$-weak reconstruction within $\mathcal{O}(\log n)$ rounds w.h.p.*[5]

Outline of Proof. Lemma 4.1 implies that for every node u at any round t we have

$$\begin{aligned} &\mathbf{x}^{(t-1)}(u) - \mathbf{x}^{(t)}(u) = \\ &= \alpha_2\lambda_2^{t-1}(1-\lambda_2)\left(\boldsymbol{\chi}(u) + \mathbf{z}(u)\right) + \mathbf{e}^{(t-1)}(u) - \mathbf{e}^{(t)}(u) \end{aligned}$$

Hence, for every node u such that $|\mathbf{z}(u)| < 1/2$,[6] we have $\mathtt{sgn}\left(\mathbf{x}^{(t-1)}(u) - \mathbf{x}^{(t)}(u)\right) = \mathtt{sgn}\left(\alpha_2\boldsymbol{\chi}(u)\right)$ whenever

$$(4.6) \qquad \left|\frac{1}{2}\alpha_2\lambda_2^{t-1}(1-\lambda_2)\right| > \left|\mathbf{e}^{(t-1)}(u) - \mathbf{e}^{(t)}(u)\right|.$$

From Lemma 4.1 we have $\left|\mathbf{e}^{(t)}(u)\right| \leqslant 4\lambda^t\sqrt{2n}$, thus (4.6) is satisfied for any t such that

$$t - 1 \geqslant \log\left(\frac{16\sqrt{2n}}{|\alpha_2|(1-\lambda_2)}\right) \cdot \frac{1}{\log\left(\lambda_2/\lambda\right)}\,.$$

The right-hand side of the above formula is $\mathcal{O}(\log n)$ w.h.p., because of the following three points: i) $\lambda_2 \geqslant (1+\delta)\lambda$ by hypothesis; ii) $1 - \lambda_2 \geqslant 1/(2n^4)$ from Cheeger's inequality (see e.g. [16]) and the fact that the graph is connected; iii) using similar (although harder - see Lemma B.2) arguments as in the proof of Theorem 3.1, we can prove that Rademacher initialization of $\mathbf{x}$ w.h.p. implies $|\alpha_2| \geqslant n^{-c}$ for some large enough positive constant c. Finally, from Lemma 4.1 we have $\|\mathbf{z}\| \leqslant \frac{88\,\gamma}{\nu-\lambda_3}\sqrt{2n}$. Thus, the number of nodes u with $\mathbf{z}(u) \geqslant 1/2$ is $\mathcal{O}(n\gamma^2/(\nu-\lambda_3)^2)$. □

Roughly speaking, the above theorem states that the quality of block reconstruction depends on the regularity of the graph (through parameter γ) and conductance within each community (here represented by the difference $|\nu - \lambda_3|$). Interestingly enough, as long as $|\nu - \lambda_3| = \Theta(1)$, the protocol achieves $\mathcal{O}(\gamma^2)$-weak reconstruction on $(2n, d, b, \gamma)$-clustered graphs.

Stochastic block model. Below we prove that the stochastic block model $\mathcal{G}_{2n,p,q}$ satisfies the hypotheses of Theorem 4.1, w.h.p., and, thus, the AVERAGING protocol efficiently produces a good reconstruction. In what follows, we will often use the following parameters of the model: expected internal degree $a = pn$, expected external degree $b = qn$, and $d = a + b$.

---

[5] Consistently, Theorem 3.1 is a special case of this one when $\gamma = 0$.

[6] The value 1/2 is chosen here only for readability sake, any constant smaller than 1 will do.

Lemma 4.2. *Let $G \sim \mathcal{G}_{2n,p,q}$. If $a - b > \sqrt{(a+b)\log n}$ then a positive constant δ exists such that the following hold w.h.p.: i) G is $(2n, d, b, 6\sqrt{\log n/d})$-clustered* and *ii)* $\lambda \leqslant \min\left\{\lambda_2/(1+\delta)\,,\, 24\sqrt{(\log n)/d}\right\}$.

Outline of Proof. Claim (i) follows (with probability $1 - n^{-1}$) from an easy application of the Chernoff bound. As for Claim (ii), since G is not regular and random, we derive spectral properties on its adjacency matrix A by considering a "more-tractable" matrix, namely the expected matrix

$$B := \mathbf{E}\left[A\right] = \begin{pmatrix} pJ, & qJ \\ qJ, & pJ \end{pmatrix}$$

where $B_{i,j}$ is the probability that the edge (i, j) exists in a random graph $G \sim \mathcal{G}_{2n,p,q}$. In Lemma D.2 we will prove that such a G is likely to have an adjacency matrix A close to B in spectral norm. Then, in Lemma D.3 we will show that every clustered graph whose adjacency matrix is close to B has the properties required in the analysis of the Averaging dynamics, thus getting Claim (ii). □

By combining Lemma 4.2 and Theorem 4.1, we achieve weak reconstruction for the stochastic block model.

Corollary 4.1. *Let $G \sim \mathcal{G}_{2n,p,q}$. If $a - b > 25\sqrt{d\log n}$ and $b = \Omega(\log n/n^2)$ then the* Averaging *protocol produces an $\mathcal{O}(d\log n/(a-b)^2)$-weak reconstruction in $\mathcal{O}(\log n)$ rounds w.h.p.*

Outline of Proof. From Lemma 4.2 we get that w.h.p. G is $(2n, d, b, \gamma)$-clustered with $\gamma \leqslant 6\sqrt{\log n/d}$, $|\lambda_i| \leqslant 4\gamma$ for all $i = 3, \ldots, 2n$ and $\lambda_2 \geqslant (1+\delta)\lambda_3$ for some constant $\delta > 0$. Given the hypotheses on a and b, we also have that the graph is connected w.h.p. Moreover, since $d\nu = (a - b) > 25\sqrt{d\log n}$, then

$$\frac{\gamma}{\nu - \lambda_3} = \frac{d\gamma}{d\nu - d\lambda_3} \leqslant \frac{6\sqrt{d\log n}}{(a-b) - 24\sqrt{d\log n}} = \mathcal{O}\left(\frac{\sqrt{d\log n}}{(a-b)}\right).$$

Theorem 4.1 then guarantees that the Averaging protocol finds an $\mathcal{O}\left(d\log n/(a-b)^2\right)$-weak reconstruction w.h.p. □

4.1 Tight analysis for the stochastic block model In Lemma 4.2 we have shown that, when $(a - b) > \sqrt{(a+b)\log n}$, a graph sampled according to $\mathcal{G}_{2n,p,q}$ satisfies the hypothesis of Theorem 4.1 w.h.p.: The simple Averaging protocol thus gets weak-reconstruction in $\mathcal{O}(\log n)$ rounds. As for the parameters' range of $\mathcal{G}_{2n,p,q}$, we know that the above result is still off by a factor $\sqrt{\log n}$ from the threshold $(a - b) > 2\sqrt{(a+b)}$ [44, 40, 45], the latter being a necessary condition for any (centralized or not) non-trivial weak reconstruction. Essentially, the reason behind this gap is that, while Theorem 4.1 holds for *any* (i.e. "worst-case") $(2n, d, b, \gamma)$-clustered graph, in order to apply it to $\mathcal{G}_{2n,p,q}$ we need to choose parameters a and b in a way that γd bounds the variation of the degree of *any* node w.r.t. the regular case w.h.p.

On the other hand, since the degrees in $\mathcal{G}_{2n,p,q}$ are distributed according to a sum of Bernoulli random variables, the rare event that some degrees are much higher than the average does not affect too much the eigenvalues and eigenvectors of the graph. Indeed, by adopting ad-hoc arguments for $\mathcal{G}_{2n,p,q}$, we prove that the Averaging protocol actually achieves an $\mathcal{O}(d/(a-b)^2)$-weak reconstruction w.h.p., provided that $(a - b)^2 > c_1(a + b) > 5\log n$, thus matching the weak-reconstruction threshold up to a constant factor for graphs of logarithmic degree. The main argument relies on the spectral properties of $\mathcal{G}_{2n,p,q}$ stated in the following lemma, whose complete proof is given in Appendix D.

Lemma 4.3. *Let $G \sim \mathcal{G}_{2n,p,q}$. If $(a-b)^2 > c_1(a+b) > 5\log n$ and*[6] *$a+b < n^{\frac{1}{3}-c_5}$ for some positive constants c_1 and c_5, then the following claims hold w.h.p.:*

1. *$\lambda_2 \geqslant 1 - 2b/d - c_2/\sqrt{d}$ for some constant $c_2 > 0$,*
2. *$\lambda_2 \geqslant (1+\delta)\lambda$ for some constant $\delta > 0$ (where as usual $\lambda = \max\{|\lambda_3|, \ldots, |\lambda_{2n}|\}$),*
3. *$|\sqrt{2nd}(D^{-1/2}\mathbf{w}_2)(i) - \boldsymbol{\chi}(i)| \leqslant \frac{1}{100}$ for each $i \in V \setminus S$, for some subset S with $|S| = \mathcal{O}(nd/(a-b)^2)$.*

Idea of the proof. The key-steps of the proof are two probability-concentration results. In Lemma D.5, we prove a tight bound on the deviation of the Laplacian $\mathcal{L}(A) = I - N$ of $\mathcal{G}_{2n,p,q}$ from the Laplacian of the expected matrix $\mathcal{L}(B) = I - \frac{1}{d}B$. As one may expect from previous results on the Erdős-Rényi model and from Le and Vershynin's recent concentration results for inhomogeneous Erdős-Rényi graph (see [37]), we can prove that w.h.p. $\|\mathcal{L}(A) - \mathcal{L}(B)\| = \mathcal{O}(\sqrt{d})$, even when $d = \Theta(\log n)$. To derive the latter result, we leverage on the aforementioned Le and Vershynin's bound on the spectral norm of inhomogeneous Erdős-Rényi graphs; in $\mathcal{G}_{2n,p,q}$ this bound implies that if $d = \Omega(\log n)$ then w.h.p. $\|A - B\| = \mathcal{O}(\sqrt{d})$. Then, while Le and Vershynin replace the Laplacian matrix with regularized versions of it, we are able to bound $\|\mathcal{L}(A) - \mathcal{L}(B)\|$ directly by upper bounding it with $\|A - B\|$ and an additional factor $\|B - d^{-1}\, D^{1/2} B D^{1/2}\|$. We then bound from above the latter additional factor thanks to our second result: In Lemma D.6, we prove that w.h.p. $\sum(\sqrt{d_i}-\sqrt{d})^2 \leqslant 2n$ and $\sum(d_i-d)^2 \leqslant 2nd$. We can then prove the first two claims of Lemma 4.3 by bounding the distance of the eigenvalues of N from those of $d^{-1}\, B$ via Lemma A.2. As for the third claim of the lemma, we prove it by upper bounding the components of $D^{-1/2}\mathbf{w}$ orthogonal to $\boldsymbol{\chi}$. In particular, we can limit the projection $\mathbf{w_1}$ of $D^{-1/2}\mathbf{w}$ on $\mathbf{1}$ by using Lemma D.6. Then, we can upper bound the projection $\mathbf{w}_\perp$ of $D^{-1/2}\mathbf{w}$ on the space orthogonal to both $\boldsymbol{\chi}$ and $\mathbf{1}$ with Lemma D.5: We look at N as a perturbed version of B and apply the Davis-Kahan theorem. Finally, we conclude the proof observing that $\|\mathbf{w}_2 - (2n)^{-1/2}\| \leqslant 2(\|\mathbf{w_1}\| + \|\mathbf{w}_\perp\|)$. ☐

Once we have Lemma 4.3 we can prove the main theorem on $\mathcal{G}_{2n,p,q}$ with the same argument used for Theorem 4.1 (the full proof is given in Appendix D).

Theorem 4.2. *Let $G \sim \mathcal{G}_{2n,p,q}$. If $(a-b)^2 > c_1(a+b) > 5\log n$ and*[7] *$a+b < n^{\frac{1}{3}-c_5}$ for some positive constants c_1 and c_5, then the* Averaging *protocol produces an $\mathcal{O}(d/(a-b)^2)$-weak reconstruction within $\mathcal{O}(\log n)$ rounds w.h.p.*

5 Moving beyond two communities: An outlook

The Averaging protocol can be naturally extended to address the case of more communities. One way to achieve this is by performing a suitable number of independent, parallel runs of the protocol. We next outline the analysis for a natural generalization of the regular block model. This allows us to easily present the main ideas and to provide an intuition of how and why the protocol works.

Let $G = (V, E)$ be a d-regular graph in which V is partitioned into k equal-size communities $V_1, \ldots, V_k$, while every node in V_i has exactly a neighbors within V_i and exactly b neighbors in each V_j, for $j \neq i$. Note that $d = a + (k-1) \cdot b$. It is easy to see that the transition matrix P of the random walk on G has an eigenvalue $(a-b)/d$ with multiplicity $k-1$. The eigenspace of $(a-b)/d$ consists of all stepwise vectors that are constant within each community V_i and whose entries sum to zero. If $\max\{|\lambda_{2n}|, \lambda_{k+1}\} < (1-\varepsilon)\cdot(a-b)/d$, P has eigenvalues $\lambda_1 = 1$, $\lambda_2 = \cdots = \lambda_k = (a-b)/d$, with all other eigenvalues strictly smaller by a $(1-\varepsilon)$ factor.

Let T be a large enough threshold such that, for all $t \geqslant T$, $\lambda_2^t > n^2 \lambda_{k+1}^t$ and note that T is in the order of $(1/\varepsilon)\log n$. Let $\mathbf{x} \in \mathbb{R}^V$ be a vector. We say that a vertex v is of *negative type* with respect to $\mathbf{x}$ if, for all $t > T$, the value $(P^t\mathbf{x})_v$ decreases with t. We say that a vertex v is of *positive type* with respect to $\mathbf{x}$ if, for all $t > T$, the value $(P^t\mathbf{x})_v$ increases with t. Note that a vertex might have neither type, because $(P^t\mathbf{x})_v$ might not be strictly monotone in t for all $t > T$.

In Appendix E we prove the following: If we pick ℓ random vectors $\mathbf{x}^1, \ldots, \mathbf{x}^\ell$, each in $\{-1, 1\}^V$ then, with high probability, i) every vertex is either of positive or negative type for each $\mathbf{x}^i$;[8] ii) furthermore, if we associate a "signature" to each vertex, namely, the sequence of ℓ types, then vertices within the same V_i exhibit the same signature, while vertices in different V_i, V_j have different signatures. These are the basic intuitions that allow us to prove the following theorem.

Theorem 5.1. (More communities) *Let $G = (V, E)$ be a k-clustered d-regular graph defined as above and assume that $\lambda = \max\{|\lambda_{2n}|, \lambda_{k+1}\} < (1-\varepsilon)\frac{a-b}{d}$, for a suitable constant $\varepsilon > 0$. Then, for $\ell = \Theta(\log n)$, the* Averaging *protocol with ℓ parallel runs produces a strong reconstruction within $\mathcal{O}(\log n)$ rounds, w.h.p.*

References

[1] Emmanuel Abbe, Afonso S Bandeira, and Georgina Hall. Exact recovery in the stochastic block model. *IEEE Trans. on Information Theory*, 62(1):471–487, 2014.

[2] Emmanuel Abbe and Colin Sandon. Detection in the stochastic block model with multiple clusters: proof of the achievability conjectures, acyclic bp, and the information-computation gap. *arXiv preprint arXiv:1512.09080*, 2015.

[3] Yehuda Afek, Noga Alon, Omer Barad, Eran Hornstein, Naama Barkai, and Ziv Bar-Joseph. A biological solution to a fundamental distributed computing problem. *science*, 331(6014):183–185, 2011.

[4] Dana Angluin, James Aspnes, and David Eisenstat. A Simple Population Protocol for Fast Robust Approximate Majority. *Distributed Computing*, 21(2):87–102, 2008. (Preliminary version in DISC'07).

[5] Tuncer Can Aysal, Mehmet Ercan Yildiz, Anand D Sarwate, and Anna Scaglione. Broadcast gossip algorithms for consensus. *Signal Processing, IEEE Transactions on*, 57(7):2748–2761, 2009.

---

[7] It should be possible to weaken the condition $d < n^{\frac{1}{3}-c_5}$ via some stronger concentration argument; see the proof of Lemma D.6 in [7] for details.

[8] I.e., for every $t > T$, $(P^t\mathbf{x})_v$ monotonically increases (or decreases) with t.

[6] Michael J Barber and John W Clark. Detecting network communities by propagating labels under constraints. *Physical Review E*, 80(2):026129, 2009.

[7] Luca Becchetti, Andrea E. F. Clementi, Emanuele Natale, Francesco Pasquale, and Luca Trevisan. Find your place: Simple distributed algorithms for community detection. *CoRR*, abs/1511.03927, 2015.

[8] Florence Bénézit, Patrick Thiran, and Martin Vetterli. Interval consensus: from quantized gossip to voting. In *Proc. of the 34th IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP '09)*, pages 3661–3664, 2009.

[9] Ravi B Boppana. Eigenvalues and graph bisection: An average-case analysis. In *Proc. of the 28th Ann. IEEE Symp. on Foundations of Computer Science (FOCS'87)*, pages 280–285. IEEE, 1987.

[10] Charles Bordenave. Personal communication, 2015.

[11] Charles Bordenave. A new proof of Friedman's second eigenvalue Theorem and its extension to random lifts. *arXiv preprint arXiv:1502.04482*, 2015.

[12] Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Non-backtracking spectrum of random graphs: community detection and non-regular ramanujan graphs. In *Proc. of 56rd Ann. IEEE Symp. on Foundations of Computer Science (FOCS'15)*, pages 1347–1357. IEEE, 2015.

[13] Stephen Boyd, Arpita Ghosh, Balaji Prabhakar, and Devavrat Shah. Randomized gossip algorithms. *IEEE/ACM Transactions on Networking (TON)*, 14:2508–2530, 2006.

[14] Gerandy Brito, Ioana Dumitriu, Shirshendu Ganguly, Christopher Hoffman, and Linh V. Tran. Recovery and rigidity in a regular stochastic block model. In *Proc. of the 26th Ann. ACM-SIAM Symp. on Discrete Algorithms (SODA'15)*, pages 371–390, 2015.

[15] Luca Cardelli and Attila Csikász-Nagy. The Cell Cycle Switch Computes Approximate Majority. *Scientific Reports*, Vol. 2(656), 2012.

[16] F. R. K. Chung. Laplacians of graphs and Cheeger's inequalities. In *Combinatorics, Paul Erdös is eighty*, volume 2, pages 157–172. János Bolyai Math. Soc., 1996.

[17] Amin Coja-Oghlan. *Spectral techniques, semidefinite programs, and random graphs.* Habilitation thesis, Humboldt University Berlin, 2005.

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

[19] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84(6):066106, 2011.

[20] Morris H DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.

[21] David Doty. Timing in chemical reaction networks. In *Proc. of 25th Ann. ACM-SIAM Symp. on Discrete Algorithms (SODA'14)*, pages 772–784. SIAM, 2014.

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

[23] David Easley and Jon Kleinberg. *Networks, crowds, and markets: Reasoning about a highly connected world.* Cambridge University Press, 2010.

[24] Nigel R. Franks, Stephen C. Pratt, Eamonn B. Mallon, Nicholas F. Britton, and David J.T. Sumpter. Information flow, opinion polling and collective intelligence in house–hunting social insects. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 357(1427):1567–1583, 2002.

[25] Noah E. Friedkin and Eugene C. Johnsen. Social influence and opinions. *The Journal of Mathematical Sociology*, 15(3-4):193–206, 1990.

[26] Joel Friedman and David-Emmanuel Kohler. The relativized second eigenvalue conjecture of alon. *arXiv preprint arXiv:1403.3462*, 2014.

[27] Yehuda Hassin and David Peleg. Distributed probabilistic polling and applications to proportionate agreement. *Information and Computation*, 171(2):248–268, 2001.

[28] Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

[29] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.

[30] Mark Jerrum and Gregory B. Sorkin. The metropolis algorithm for graph bisection. *Discrete Applied Mathematics*, 82(1):155–175, 1998.

[31] David Kempe, Alin Dobra, and Johannes Gehrke. Gossip-based computation of aggregate information. In *Proc. of 43rd Ann. IEEE Symp. on Foundations of Computer Science (FOCS'03)*, pages 482–491, 2003.

[32] David Kempe and Frank McSherry. A decentralized algorithm for spectral analysis. In *Proc. of the 36th Ann. ACM Symp. on Theory of Computing (STOC'04)*, pages 561–568, 2004.

[33] Jon M Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999.

[34] Kishore Kothapalli, Sriram V Pemmaraju, and Vivek Sardeshmukh. On the analysis of a label propagation algorithm for community detection. In *Proc. of the 14th Int. Conf. on Distributed Computing and Networking (ICDCN'13)*, pages 255–269, 2013.

[35] Florent Krzakala, Cristopher Moore, Elchanan Mossel, Joe Neeman, Allan Sly, Lenka Zdeborová, and Pan Zhang. Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences*, 110(52):20935–20940, 2013.

[36] Cornelius Lanczos. *An iteration method for the so-*

lution of the eigenvalue problem of linear differential and integral operators. United States Governm. Press Office Los Angeles, CA, 1950.

[37] Can M. Le and Roman Vershynin. Concentration and regularization of random graphs. *arXiv:1506.00669 [cs, math, stat]*, 2015. arXiv: 1506.00669.

[38] Xin Liu and Tsuyoshi Murata. Advanced modularity-specialized label propagation algorithm for detecting communities in networks. *Physica A: Statistical Mechanics and its Applications*, 389(7):1493–1500, 2010.

[39] David JC MacKay. *Information theory, inference and learning algorithms.* Cambridge university press, 2003.

[40] Laurent Massoulie. Community Detection Thresholds and the Weak Ramanujan Property. In *Proc. of the 46th Ann. ACM Symp. on Theory of Computing (STOC'14)*, pages 694–703, 2014.

[41] Frank McSherry. Spectral partitioning of random graphs. In *Proc. of the 42nd IEEE Symp. on Foundations of Computer Science (FOCS'01)*, pages 529–537, 2001.

[42] Yves Métivier, John Michael Robson, Nasser Saheb-Djahromi, and Akka Zemmari. An optimal bit complexity randomized distributed mis algorithm. *Distributed Computing*, 23(5-6):331–340, 2011.

[43] Joris M Mooij and Hilbert J Kappen. Sufficient conditions for convergence of the sum–product algorithm. *IEEE Transactions on Information Theory*, 53(12):4422–4437, 2007.

[44] Elchanan Mossel, Joe Neeman, and Allan Sly. A proof of the block model threshold conjecture. *arXiv preprint arXiv:1311.4115*, 2013.

[45] Elchanan Mossel, Joe Neeman, and Allan Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162(3-4):431–461, 2014.

[46] Elchanan Mossel, Joe Neeman, and Allan Sly. Belief propagation, robust reconstruction and optimal recovery of block models. *Ann. Appl. Probab.*, 26(4):2211–2256, 08 2016.

[47] Elchanan Mossel, Joe Neeman, and Omer Tamuz. Majority dynamics and aggregation of information in social networks. *Autonomous Agents and Multi-Agent Systems*, 28(3):408–429, 2014.

[48] Alex Olshevsky and John N Tsitsiklis. Convergence speed in distributed consensus and averaging. *SIAM Journal on Control and Optimization*, 48(1):33–55, 2009.

[49] Usha Nandini Raghavan, Réka Albert, and Soundar Kumara. Near linear time algorithm to detect community structures in large-scale networks. *Physical Review E*, 76(3):036106, 2007.

[50] Devavrat Shah. *Gossip algorithms.* Now Publishers Inc, 2009.

[51] Pete Stewart and Ji-guang Sun. *Matrix Perturbation Theory.* Academic Press, 1990.

[52] Jukka Suomela. Survey of local algorithms. *ACM Comput. Surv.*, 45(2):24:1–24:40, March 2013.

[53] John Nikolas Tsitsiklis. Problems in decentralized decision making and computation. Technical report, DTIC Document, 1984.

[54] Yair Weiss. Correctness of local probability propagation in graphical models with loops. *Neural computation*, 12(1):1–41, 2000.

[55] Lin Xiao, Stephen Boyd, and Seung-Jean Kim. Distributed average consensus with least-mean-square deviation. *Journal of Parallel and Distributed Computing*, 67(1):33–46, 2007.

Appendix

A Linear algebra toolkit

If $M \in \mathbb{R}^{n\times n}$ is a real symmetric matrix, then it has n real eigenvalues (counted with repetitions), $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$, and we can find a corresponding collection of orthonormal real eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ such that $M\mathbf{v}_i = \lambda_i \mathbf{v}_i$. Thus, if $\mathbf{x} \in \mathbb{R}^n$ is any vector, then we can write it as a linear combination $\mathbf{x} = \sum_i \alpha_i \mathbf{v}_i$ of eigenvectors, where the coefficients of the linear combination are $\alpha_i = \langle \mathbf{x}, \mathbf{v}_i \rangle$. In this notation, we can see that

$$M\mathbf{x} = \sum_i \lambda_i \alpha_i \mathbf{v}_i, \quad \text{and so} \quad M^t\mathbf{x} = \sum_i \lambda_i^t \alpha_i \mathbf{v}_i.$$

Lemma A.1. (Cauchy-Schwarz inequality) *For any pair of vectors* $\mathbf{x}$ *and* $\mathbf{y}$

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leqslant \|\mathbf{x}\| \cdot \|\mathbf{y}\|.$$

Observation 1. *For any matrix* A *and any vector* $\mathbf{x}$

$$\|A\mathbf{x}\| \leqslant \|A\| \cdot \|\mathbf{x}\|, \quad \text{and} \quad \|A \cdot B\| \leqslant \|A\| \cdot \|B\|.$$

Observation 2. *If* G *is a* $(2n, d, b)$*-clustered regular graph with clusters* V_1 *and* V_2 *and* $\boldsymbol{\chi} = \mathbf{1}_{V_1} - \mathbf{1}_{V_2}$ *is the partition indicator vector, then* $\boldsymbol{\chi}$ *is an eigenvector of the transition matrix* P *of* G *with eigenvalue* $1 - 2b/d$.

Proof. Every node i has b neighbors j on the opposite side of the partition, for which $\boldsymbol{\chi}(j) = -\boldsymbol{\chi}(i)$, and $d - b$ neighbors j on the same side, for which $\boldsymbol{\chi}(j) = \boldsymbol{\chi}(i)$, so

$$(P\boldsymbol{\chi})_i = \frac{1}{d}\left((d-b)\boldsymbol{\chi}(i) - b\boldsymbol{\chi}(i)\right) = \left(1 - \frac{2b}{d}\right)\boldsymbol{\chi}(i).$$

Theorem A.1. (Matrix Bernstein Inequality) *Let* $X_1, \ldots, X_N$ *be a sequence of independent* $n \times n$ *symmetric random matrices, such that* $\mathbf{E}[X_i] = \mathbf{0}$ *for every* i *and such that* $\|X_i\| \leqslant L$ *with probability 1 for some* L. *Call* $\sigma := \|\mathbf{E}\left[\sum_i X_i^2\right]\|$. *Then, for every* t, *we have*

$$\mathbf{P}\left(\left\|\sum_i X_i\right\| \geqslant t\right) \leqslant 2ne^{\frac{-t^2}{2\sigma + \frac{2}{3}Lt}}.$$

THEOREM A.2. *(Corollary 4.10 in [51]) Let M_1 and M_2 be two Hermitian matrices, let $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$ be the eigenvalues of M_1 with multiplicities in non-increasing order, and let $\lambda'_1 \geqslant \lambda'_2 \geqslant \cdots \geqslant \lambda'_n$ be the eigenvalues of M_2 with multiplicities in non-increasing order. Then, for every i,*

$$|\lambda_i - \lambda'_i| \leqslant \|M_1 - M_2\|.$$

THEOREM A.3. (DAVIS AND KAHAN, 1970) *Let M_1 and M_2 be two symmetric real matrices, let $\mathbf{x}$ be a unit length eigenvector of M_1 of eigenvalue t, and let $\mathbf{x}_p$ be the projection of $\mathbf{x}$ on the eigenspace of the eigenvectors of M_2 corresponding to eigenvalues $\leqslant t - \delta$. Then*

$$\|\mathbf{x}_p\| \leqslant \frac{2}{\delta\pi}\|M_1 - M_2\|.$$

B Length of the projection of x

For the analysis of the AVERAGING dynamics on both regular and non-regular graphs, it is important to understand the distribution of the projection of $\mathbf{x}$ on $\mathbf{1}$ and $\boldsymbol{\chi}$, that is (up to scaling) the distribution of the inner products $\langle \mathbf{x}, \mathbf{1}\rangle$ and $\langle \mathbf{x}, \boldsymbol{\chi}\rangle$. In particular we are going to use the following bound.

LEMMA B.1. *If we pick $\mathbf{x}$ uniformly at random in $\{-1,1\}^{2n}$ then, for any $\delta > 0$ and any fixed vector $\mathbf{w} \in \{-1,1\}^{2n}$ with ± 1 entries, it holds*

$$\mathbf{P}\left(\left|\langle (1/\sqrt{2n})\,\mathbf{w},\, \mathbf{x}\rangle\right| \leqslant \delta\right) \leqslant \mathcal{O}(\delta).$$

Proof. Since $\mathbf{x}$ is a vector of independent and uniformly distributed random variables in $\{-1,1\}$, both $\langle \mathbf{x}, \boldsymbol{\chi}\rangle$ and $\langle \mathbf{x}, \mathbf{1}\rangle$ have the distribution of a sum of $2n$ Rademacher random variables. Such a sum takes the value $2k - 2n$ with probability $\frac{1}{2^n}\binom{2n}{k}$, and so every possible value has probability at most $\frac{1}{2^n}\binom{2n}{n} \approx \frac{1}{\sqrt{2\pi n}}$. Consequently, if R is the sum of $2n$ Rademacher random variables, we have $\mathbf{P}\left(|R| \leqslant \delta\sqrt{2n}\right) \leqslant \mathcal{O}(\delta)$.

Although it is possible to argue that a Rademacher vector has $\Omega(1)$ probability of having inner product $\Omega(\|\mathbf{w}\|)$ with every vector $\mathbf{w}$, such a statement does not hold w.h.p. We do have, however, estimates of the inner product of a vector $\mathbf{w}$ with a Rademacher vector $\mathbf{x}$ provided that $\mathbf{w}$ is close to a vector in $\{-1,1\}^{2n}$.

LEMMA B.2. *Let k be a positive integer. For every nk-dimensional vector $\mathbf{w}$ such that $|\{i \mid |\mathbf{w}(i)| \geqslant c\}| \geqslant n$ for some positive constant c, if we pick $\mathbf{x}$ uniformly at random in $\{-1,1\}^{kn}$, then*

$$\mathbf{P}\left(\left|\langle (1/\sqrt{kn})\,\mathbf{w},\, \mathbf{x}\rangle\right| \leqslant \delta\right) \leqslant \mathcal{O}(k\delta) + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right).$$

Proof. Let $S \subset \{1, \ldots, kn\}$ be the set of coordinates i of $\mathbf{w}$ such that $|\mathbf{w}(i)| \geqslant c$. By hypothesis, we have $|S| \geqslant n$. Let $T := \{1, \ldots, kn\} - S$. Now, for every assignment $\mathbf{a} \in \{-1,1\}^{kn}$, we will show that

$$\mathbf{P}\left(|\langle \mathbf{w}, \mathbf{x}\rangle| \leqslant \delta\sqrt{kn} \;\middle|\; \forall i \in T,\, \mathbf{x}(i) = \mathbf{a}(i)\right) \leqslant \mathcal{O}(\delta),$$

and then the lemma will follow. Call $t := \sum_{i \in T} a_i z_i$. We need to show

$$\mathbf{P}\left(|\sum_{i \in S} \mathbf{x}(i)\mathbf{w}(i) + t| \leqslant \delta\sqrt{kn}\right) \leqslant \mathcal{O}(\delta).$$

From the Berry-Esseen theorem,

$$\mathbf{P}\left(|\sum_{i \in S} \mathbf{x}(i)\mathbf{w}(i) + t| \leqslant \delta\sqrt{kn}\right) \leqslant$$

$$\leqslant \mathbf{P}\left(|g + t| \leqslant \delta\sqrt{kn}\right) + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right),$$

where g is a Gaussian random variable of mean 0 and variance $\sigma^2 = \sum_{i \in S}(\mathbf{w}(i))^2 \geqslant c^2|S| \geqslant c^2\, n$, so

$$\mathbf{P}\left(|g + t| \leqslant \delta\sqrt{kn}\right) = \frac{1}{\sqrt{2\sigma^2\pi}} \int_{-t-\delta\sqrt{kn}}^{-t+\delta\sqrt{kn}} e^{-\frac{s^2}{2\sigma^2}}\, ds$$

$$\leqslant \frac{2\delta\sqrt{kn}}{\sqrt{2\pi c^2\, n}} = \frac{\sqrt{2k}\delta}{\sqrt{\pi}c},$$

where we used the fact that $e^{-s^2/2} \leqslant 1$ for all s.

C Clustered Graphs

LEMMA C.1. *Assume we run the* AVERAGING *dynamics in a $(2n, d, b)$-clustered regular graph G (see Definition 1) with any initial vector $\mathbf{x} \in \{-1,1\}^{2n}$. If $\lambda < 1 - 2b/d$ then there are reals α_1, α_2 such that at every round t we have*

$$\mathbf{x}^{(t)} = \alpha_1 \mathbf{1} + \alpha_2 \lambda_2^t \boldsymbol{\chi} + \mathbf{e}^{(t)} \quad \textit{where} \quad \left\|\mathbf{e}^{(t)}\right\|_\infty \leqslant \lambda^t \sqrt{2n}\,.$$

Proof. Since $\mathbf{x}^{(t)} = P^t \mathbf{x}$ we can write

$$P^t \mathbf{x} = \sum_i \lambda_i^t \langle \mathbf{x}, \mathbf{v}_i\rangle \mathbf{v}_i,$$

where $1 = \lambda_1 > \lambda_2 = 1 - 2b/d > \lambda_3 \geqslant \cdots \geqslant \lambda_{2n}$ are the eigenvalues of P and $\mathbf{v}_1 = \frac{1}{\sqrt{2n}}\mathbf{1}$, $\mathbf{v}_2 = \frac{1}{\sqrt{2n}}\boldsymbol{\chi}$, $\mathbf{v}_3$, $\ldots$, $\mathbf{v}_{2n}$ are a corresponding sequence of orthonormal eigenvectors. Hence,

$$\mathbf{x}^{(t)} = \frac{1}{2n}\langle \mathbf{x}, \mathbf{1}\rangle \cdot \mathbf{1} + \lambda_2^t \frac{1}{2n}\langle \mathbf{x}, \boldsymbol{\chi}\rangle \cdot \boldsymbol{\chi} + \sum_{i=3}^{2n} \lambda_i^t \alpha_i \mathbf{v}_i$$

$$= \alpha_1 \mathbf{1} + \alpha_2 \lambda_2^t \cdot \boldsymbol{\chi} + \sum_{i=3}^{2n} \lambda_i^t \alpha_i \mathbf{v}_i,$$

where we set $\alpha_1 = \frac{1}{2n}\langle \mathbf{1}, \mathbf{x} \rangle$ and $\alpha_2 = \frac{1}{2n}\langle \boldsymbol{\chi}, \mathbf{x} \rangle$. We bound the ℓ_∞ norm of the last term as

$$\left\| \sum_{i=3}^{2n} \lambda_i^t \alpha_i \mathbf{v}_i \right\|_\infty \leqslant \left\| \sum_{i=3}^{2n} \lambda_i^t \alpha_i \mathbf{v}_i \right\|_2 = \sqrt{\sum_{i=3}^{2n} \lambda_i^{2t} \alpha_i^2}$$
$$\leqslant \lambda^t \sqrt{\sum_{i=1}^{2n} \alpha_i^2} = \lambda^t \|\mathbf{x}\| = \lambda^t \sqrt{2n}.$$

Lemma C.2. *Let G be a connected $(2n, d, b, \gamma)$-clustered graph (see Definition 2) with $\gamma \leqslant 1/10$. If $\lambda_3 < \nu$ then*

$$\lambda_2 \geqslant \nu - 10\gamma \quad \textit{and}$$
$$\left\| D^{1/2}\boldsymbol{\chi} - \beta_2 \mathbf{w}_2 \right\| \leqslant \tfrac{44\,\gamma}{\nu - \lambda_3}\sqrt{2nd},$$

where $\beta_2 = \boldsymbol{\chi}^\intercal D^{1/2} \mathbf{w}_2$.

Proof. For every node v, let us name a_v and b_v the numbers of neighbors of v in its own cluster and in the other cluster, respectively, and $d_v = a_v + b_v$ its degree. Since from the definition of $(2n, d, b, \gamma)$-clustered graph it holds that $(1-\gamma)d \leqslant d_v \leqslant (1+\gamma)d$ and $b - \gamma d \leqslant b_v \leqslant b + \gamma d$, it is easy to check that

$$|a_v - b_v - \nu d_v| \leqslant 4d\,\gamma$$

for any node v. Hence,

$$\begin{aligned}
&\|A\boldsymbol{\chi} - \nu D\boldsymbol{\chi}\|^2 \\
&= \sum_{v \in [2n]} \left(\sum_{w \in \mathrm{Neigh}(v)} \boldsymbol{\chi}(w) - \nu d_v \boldsymbol{\chi}(v) \right)^2 \\
&= \sum_{v \in [2n]} \left(a_v \boldsymbol{\chi}(v) - b_v \boldsymbol{\chi}(v) - \nu d_v \boldsymbol{\chi}(v) \right)^2 \\
&= \sum_{v \in [2n]} \left(a_v - b_v - \nu d_v \right)^2 \leqslant 32 n d^2 \gamma^2.
\end{aligned}$$

Thus,

$$\begin{aligned}
\text{(C.1)} \quad &\left\| N D^{1/2}\boldsymbol{\chi} - \nu D^{1/2}\boldsymbol{\chi} \right\| = \\
&= \left\| D^{-1/2} A\boldsymbol{\chi} - \nu D^{1/2}\boldsymbol{\chi} \right\| \\
&= \left\| D^{-1/2} \left(A\boldsymbol{\chi} - \nu D\boldsymbol{\chi} \right) \right\| \\
&\leqslant \left\| D^{-1/2} \right\| \cdot \|A\boldsymbol{\chi} - \nu D\boldsymbol{\chi}\| \\
&\leqslant \frac{2}{\sqrt{d}} \cdot \sqrt{2n} 4d\,\gamma = 8\sqrt{2nd}\,\gamma.
\end{aligned}$$

Observe that $\mathbf{w}_1$ is parallel to $D^{1/2}\mathbf{1}$ and we have that

$$\begin{aligned}
\text{(C.2)} \quad |\mathbf{1}^\intercal D \boldsymbol{\chi}| &= \left| \sum_{v \in [2n]} \boldsymbol{\chi}(v) d_v \right| \\
&\leqslant (1+\gamma)dn - (1-\gamma)dn = 2nd\,\gamma.
\end{aligned}$$

Hence, if we name $\mathbf{y}$ the component of $D^{1/2}\boldsymbol{\chi}$ orthogonal to the first eigenvector, we can write it as

$$\text{(C.3)} \qquad D^{1/2}\boldsymbol{\chi} = \frac{\mathbf{1}^\intercal D \boldsymbol{\chi}}{\|D^{1/2}\mathbf{1}\|^2} D^{1/2}\mathbf{1} + \mathbf{y}.$$

Thus,

$$\begin{aligned}
\text{(C.4)} \quad &\|N\mathbf{y} - \nu \mathbf{y}\| = \\
&= \Bigg\| N \left(D^{1/2}\boldsymbol{\chi} - \frac{\mathbf{1}^\intercal D \boldsymbol{\chi}}{\|D^{1/2}\mathbf{1}\|^2} D^{1/2}\mathbf{1} \right) - \\
&\quad - \nu \left(D^{1/2}\boldsymbol{\chi} - \frac{\mathbf{1}^\intercal D \boldsymbol{\chi}}{\|D^{1/2}\mathbf{1}\|^2} D^{1/2}\mathbf{1} \right) \Bigg\| \leqslant \\
&\leqslant \left\| N D^{1/2}\boldsymbol{\chi} - \nu D^{1/2}\boldsymbol{\chi} \right\| + \\
&\quad + \frac{|\mathbf{1}^\intercal D \boldsymbol{\chi}|}{\|D^{1/2}\mathbf{1}\|^2} \left\| N D^{1/2}\mathbf{1} - \nu D^{1/2}\mathbf{1} \right\| = \\
&= \left\| N D^{1/2}\boldsymbol{\chi} - \nu D^{1/2}\boldsymbol{\chi} \right\| + \frac{|\mathbf{1}^\intercal D \boldsymbol{\chi}|}{\|D^{1/2}\mathbf{1}\|} \frac{2b}{d} \leqslant \\
&\leqslant 8\sqrt{2nd}\,\gamma + 4\sqrt{2nd}\,\gamma,
\end{aligned}$$

where in the last inequality we used (C.1) and (C.2) and the facts that $b \leqslant d/2$ and $\left\| D^{1/2}\mathbf{1} \right\| \geqslant (1/2)\sqrt{2nd}$. From (C.3) it follows that

$$\begin{aligned}
\text{(C.5)} \quad \|\mathbf{y}\| &\geqslant \left\| D^{1/2}\boldsymbol{\chi} \right\| - \frac{\mathbf{1}^\intercal D \boldsymbol{\chi}}{\left\| D^{1/2}\mathbf{1} \right\|} \\
&\geqslant (1-\gamma)\sqrt{2nd} - 4\gamma\sqrt{2nd} \\
&= (1-5\gamma)\sqrt{2nd} \geqslant (1/2)\sqrt{2nd}.
\end{aligned}$$

Now, let us we write $\mathbf{y}$ as a linear combination of the orthonormal eigenvectors of N, $\mathbf{y} = \beta_2 \mathbf{w}_2 + \cdots + \beta_n \mathbf{w}_n$ (recall that $\mathbf{y}^\intercal \mathbf{w}_1 = 0$ by definition of $\mathbf{y}$ in (C.3)). From (C.4) and (C.5), it follows that

$$\begin{aligned}
\text{(C.6)} \quad 100\gamma^2 \|\mathbf{y}\|^2 &\geqslant \|N\mathbf{y} - \nu \mathbf{y}\|^2 \\
&= \left\| \sum_{i=2}^{n} (\lambda_i - \nu)\beta_i \mathbf{w}_i \right\|^2 = \sum_{i=2}^{n} (\lambda_i - \nu)^2 \beta_i^2.
\end{aligned}$$

Moreover, from hypothesis $\lambda_3 < \nu$ we have that

$$\begin{aligned}
\text{(C.7)} \quad \sum_{i=2}^{n} (\lambda_i - \nu)^2 \beta_i^2 &\geqslant \sum_{i=3}^{n} (\lambda_i - \nu)^2 \beta_i^2 \\
&\geqslant (\lambda_3 - \nu)^2 \sum_{i=3}^{n} \beta_i^2 = (\lambda_3 - \nu)^2 \|\mathbf{y} - \beta_2 \mathbf{w}_2\|^2.
\end{aligned}$$

Thus, by combining together (C.6) and (C.7) we get

$$\|\mathbf{y} - \beta_2 \mathbf{w}_2\| \leqslant \frac{10\,\gamma}{\nu - \lambda_3}\|\mathbf{y}\|$$

where $\beta_2 = \mathbf{y}^\intercal \mathbf{w}_2 = \left(D^{1/2}\boldsymbol{\chi}\right)^\intercal \mathbf{w}_2$.
As for the first thesis of the lemma, observe that if $\lambda_2 \geqslant \nu$ then the first thesis is obvious. Otherwise, if $\lambda_2 < \nu$, then $(\lambda_2 - \nu)^2 \leqslant (\lambda_3 - \nu)^2 \leqslant \cdots \leqslant (\lambda_n - \nu)^2$. Thus, the first thesis follows from (C.6) and the fact that

$$\sum_{i=2}^{n}(\lambda_i - \nu)^2 \beta_i^2 \geqslant (\lambda_2 - \nu)^2 \sum_{i=2}^{n} \beta_i^2 = (\lambda_2 - \nu)^2 \|\mathbf{y}\|^2.$$

As for the second thesis of the lemma, we have

$$\begin{aligned}
&\left\|D^{1/2}\boldsymbol{\chi} - \beta_2 \mathbf{w}_2\right\| = \left\| \frac{\mathbf{1}^\intercal D \boldsymbol{\chi}}{\|D^{1/2}\mathbf{1}\|^2} D^{1/2}\mathbf{1} + \mathbf{y} - \beta_2 \mathbf{w}_2 \right\| \\
&\leqslant \frac{|\mathbf{1}^\intercal D \boldsymbol{\chi}|}{\|D^{1/2}\mathbf{1}\|} + \|\mathbf{y} - \beta_2 \mathbf{w}_2\| \leqslant 4\,\gamma\sqrt{2nd} + \frac{10\,\gamma}{\nu - \lambda_3}\|\mathbf{y}\| \\
&\overset{(a)}{\leqslant} 4\,\gamma\sqrt{2nd} + \frac{20\,\gamma}{\nu - \lambda_3}\sqrt{2nd} \leqslant \frac{44\,\gamma}{\nu - \lambda_3}\sqrt{2nd},
\end{aligned}$$

where in (a) we used that $\mathbf{y}$ is the projection of $D^{\frac{1}{2}}\boldsymbol{\chi}$ on $D^{\frac{1}{2}}\mathbf{1}$, and thus $\|\mathbf{y}\| \leqslant \|D^{\frac{1}{2}}\boldsymbol{\chi}\| \leqslant 2\sqrt{2nd}$.

C.1 Proof of Theorem 4.3 From Lemma 4.1 it follows that for every node u at any round t we have

$$\begin{aligned}
&\mathbf{x}^{(t-1)}(u) - \mathbf{x}^{(t)}(u) = \\
&= \alpha_2 \lambda_2^{t-1}(1-\lambda_2)\left(\boldsymbol{\chi}(u) + \mathbf{z}(u)\right) + \mathbf{e}^{(t-1)}(u) - \mathbf{e}^{(t)}(u)\,.
\end{aligned}$$

Hence, for every node u such that $|\mathbf{z}(u)| < 1/2$ (we choose $1/2$ here for readability sake, however any other constant smaller than 1 works as well) it holds that $\mathbf{sgn}\left(\mathbf{x}^{(t-1)}(u) - \mathbf{x}^{(t)}(u)\right) = \mathbf{sgn}\left(\alpha_2 \boldsymbol{\chi}(u)\right)$ whenever

$$\left|\frac{1}{2}\alpha_2 \lambda_2^{t-1}(1-\lambda_2)\right| > \left|\mathbf{e}^{(t-1)}(u) - \mathbf{e}^{(t)}(u)\right|. \tag{C.8}$$

From Lemma 4.1 we have that $\left|\mathbf{e}^{(t)}(u)\right| \leqslant 4\lambda^t \sqrt{2n}$, thus (C.8) is satisfied for all

$$t - 1 \geqslant \frac{\log\left(\frac{16\sqrt{2n}}{|\alpha_2|(1-\lambda_2)}\right)}{\log\left(\lambda_2/\lambda\right)}. \tag{C.9}$$

The right-hand side in the above formula is $\mathcal{O}(\log n)$ w.h.p., because of the following three points:

- From Cheeger's inequality (see e.g. [16]) and the fact that the graph is connected it follows that $1 - \lambda_2 \geqslant 1/(2n^4)$;
- $\lambda_2 \geqslant (1+\delta)\lambda$ by hypothesis;
- It holds $|\alpha_2| \geqslant n^{-c}$ for some large enough positive constant c w.h.p., as a consequence of the following equations that we prove below:

$$\begin{aligned}
\mathbf{P}\left(|\alpha_2| \leqslant \frac{1}{n^c}\right) &= \mathbf{P}\left(\frac{\left|\mathbf{w}_2^\intercal D^{\frac{1}{2}}\mathbf{x}\right|}{\left|\mathbf{w}_2^\intercal D^{\frac{1}{2}}\boldsymbol{\chi}\right|} \leqslant \frac{1}{n^c}\right) \\
&\leqslant \mathbf{P}\left(\left|\mathbf{w}_2^\intercal D^{1/2}\mathbf{x}\right| \leqslant \frac{2\sqrt{d}}{n^{c-1/2}}\right) \\
&\leqslant \mathcal{O}\left(\frac{1}{\sqrt{n}}\right).
\end{aligned} \tag{C.10}$$

In the first equality of (C.10) we used that, by definition, $|\alpha_2| = |\mathbf{w}_2^\intercal D^{\frac{1}{2}}\mathbf{x}|/|\mathbf{w}_2^\intercal D^{\frac{1}{2}}\boldsymbol{\chi}|$. In the first inequality we used that, by the Cauchy-Schwarz inequality, $|\mathbf{w}_2^\intercal D^{\frac{1}{2}}\boldsymbol{\chi}| \leqslant \|D^{\frac{1}{2}}\boldsymbol{\chi}\| \leqslant 2\sqrt{dn}$. In order to prove the last inequality of (C.10), we use that from Lemma C.2 it holds

$$\begin{aligned}
&\left\|D^{1/2}\boldsymbol{\chi} - \beta_2 \mathbf{w}_2\right\|^2 = \\
&= \left\|D^{1/2}\boldsymbol{\chi}\right\|^2 + \|\beta_2 \mathbf{w}_2\|^2 - 2\langle D^{1/2}\boldsymbol{\chi}, \beta_2 \mathbf{w}_2\rangle \leqslant \\
&\leqslant 2\frac{44^2\,\gamma^2}{(\nu - \lambda_3)^2}\,nd,
\end{aligned}$$

that is

$$\begin{aligned}
&\langle D^{1/2}\boldsymbol{\chi}, \beta_2 \mathbf{w}_2\rangle = \langle D^{1/2}\boldsymbol{\chi}, \mathbf{w}_2\rangle^2 \\
&\geqslant \frac{1}{2}\left(\left\|D^{1/2}\boldsymbol{\chi}\right\|^2 - 2\frac{44^2\,\gamma^2}{(\nu - \lambda_3)^2}\,nd\right) \geqslant \frac{nd}{3}.
\end{aligned} \tag{C.11}$$

Since $\mathbf{w}_2$ is normalized the absolute value of its entries is at most 1, which toghether with (C.11) implies that at least a fraction $12/13$ of its entries have an absolute value greater than $1/12$. Thus, we can apply Lemma B.2 and prove the last inequality of (C.10) and, consequently, the fact that (C.9) is $\mathcal{O}(\log n)$.

Finally, from Lemma 4.1 we have

$$\|\mathbf{z}\| \leqslant \frac{88\,\gamma}{\nu - \lambda_3}\sqrt{2n}\,.$$

Thus the number of nodes u with $\mathbf{z}(u) \geqslant 1/2$ is $\mathcal{O}(n\gamma^2/(\nu - \lambda_3)^2)$.

D Stochastic Block Models

D.1 Regular stochastic block model

LEMMA D.1. *Let G be a graph sampled from the regular stochastic block model with internal and external degrees a and b respectively. W.h.p., it holds that*

$$\lambda \leqslant \frac{2}{a+b}\left(\sqrt{a+b-1} + o_n(1)\right)$$

Proof. The lemma follows from the general results of Friedman and Kohler [26], recently simplified by Bordenave [11]. If G is a multigraph on n vertices, then a *random degree k lift* of G is a distribution over graphs G' on kn vertices sampled as follows: every vertex v of G is replaced by k vertices $v_1, \ldots, v_k$ in G', every edge (u, v) in G is replaced by a random bipartite matching between $u_1, \ldots, u_k$ and $v_1, \ldots, v_k$ (if there are multiple edges, each edge is replaced by an independently sampled matching) and every self loop over u is replaced by a random degree-2 graph over $u_1, \ldots, u_k$ which is sampled by taking a random permutation $\pi : \{1, \ldots, k\} \to \{1, \ldots, k\}$ and connecting u_i to $u_\pi(i)$ for every i.

For every lift of any d-regular graph, the lifted graph is still d-regular, and every eigenvalue of the adjacency matrix of the base graph is still an eigenvalue of the lifted graph. Friedman and Kohler [26] prove that, if $d \geqslant 3$, then with probability $1 - \mathcal{O}(1/k)$ over the choice of a random lift of degree k, the new eigenvalues of the adjacency matrix of the lifted graph are at most $2\sqrt{d-1} + o_k(1)$ in absolute value. Bordenave [11, Corollary 20] has considerably simplified the proof of Friedman and Kohler; although he does not explicitly state the probability of the above event, his argument also bound the failure probability by $1/k^{\Omega(1)}$ [10].

The lemma now follows by observing that the regular stochastic block model is a random lift of degree n of the graph that has only two vertices v_1 and v_2, it has b parallel edges between v_1 and v_2, and it has $a/2$ self-loops on v_1 and $a/2$ self-loops on v_2.

D.2 Proof of Lemma 4.2

Lemma D.2. *Let A be the adjacency matrix of G. If $a(n), b(n)$ are such that $d := a + b > \log n$, then w.h.p. (over the choice of $G \sim \mathcal{G}_{2n, \frac{a}{n}, \frac{b}{n}}$) it holds $\|A - B\| \leqslant \mathcal{O}(\sqrt{d \log n})$.*

Proof. We can write $A - B$ as $\sum_{\{i,j\}} X^{\{i,j\}}$, where the matrix $X^{\{i,j\}}$ is zero in all coordinates except (i, j) and (j, i), and, in those coordinates, it is equal to $A - B$. Then we see that the matrices $X^{\{i,j\}}$ are independent, that $\mathbf{E}\left[X^{\{i,j\}}\right] = 0$, that $\|X^{\{i,j\}}\| \leqslant 1$, because every row contains at most one non-zero element, and that element is at most 1 in absolute value, and that $\mathbf{E}[\sum_{\{i,j\}} (X^{\{i,j\}})^2]$ is the matrix that is zero everywhere except for the diagonal entries (i, i) and (j, j), in which we have $B_{i,i} - B_{i,i}^2$ and $B_{j,j} - B_{j,j}^2$ respectively. It follows that

$$\|\mathbf{E}[\sum_{\{i,j\}} (X^{\{i,j\}})^2]\| \leqslant d.$$

Putting these facts together, and applying the Matrix Bernstein Inequality (see Theorem A.1 in Appendix A) with $t = \sqrt{6d \log n}$, we have

$$\begin{aligned} \mathbf{P}\left(\|A - B\| \geqslant \sqrt{9d \log n}\right) &\leqslant 2ne^{-\frac{9d \log n}{2d + \frac{2}{3}\sqrt{9d \log n}}} \\ &\leqslant 2ne^{-\frac{9d \log n}{4d}} \leqslant 2n^{-1}, \end{aligned}$$

where we used $d > \log n$.

Lemma D.3. *Let G be a $(2n, d, b, \gamma)$-clustered graph such that $\nu = 1 - \frac{2b}{d} > 12\gamma$ and such that its adjacency matrix A satisfies $\|A - B\| \leqslant \gamma d$. Then for every $i \in \{3, \ldots, 2n\}$, $|\lambda_i| \leqslant 4\gamma$ and $\lambda_2 \geqslant (1 + \delta)\lambda_3$ for some constant $\delta > 0$.*

Proof. The matrix B has a very simple spectral structure: $\mathbf{1}$ is an eigenvector of eigenvalue d, $\boldsymbol{\chi}$ is an eigenvector of eigenvalue $a - b$, and all vectors orthogonal to $\mathbf{1}$ and to $\boldsymbol{\chi}$ are eigenvectors of eigenvalue 0. In order to understand the eigenvalues and eigenvectors of N, and hence the eigenvalues and eigenvectors of P, we first prove that A approximates B and that N approximates $(1/d)A$, namely $\|dN - A\| \leqslant 3\gamma d$.

To show that dN approximates A we need to show that D approximates dI. The condition on the degrees immediately gives us $\|D - dI\| \leqslant \gamma d$. Since every vertex has degree d_i in the range $d \pm \gamma d$, then the square root $\sqrt{d_i}$ of each vertex must be in the range $[\sqrt{d} - \gamma\sqrt{d}, \sqrt{d} + \gamma\sqrt{d}]$, so we also have the spectral bound:

$$\|D^{1/2} - \sqrt{d}I\| \leqslant \gamma\sqrt{d}. \tag{D.12}$$

We know that $\|D\| \leqslant d + \gamma d < 2d$ and that $\|N\| = 1$, so from (D.12) we get

$$\begin{aligned} &\|A - dN\| = \|D^{1/2}ND^{1/2} - dN\| \\ &\leqslant \|D^{1/2}ND^{1/2} - \sqrt{d}ND^{1/2}\| + \|\sqrt{d}ND^{1/2} - dN\| \\ &= \|(D^{1/2} - \sqrt{d}I) \cdot ND^{1/2}\| + \|\sqrt{d}N \cdot (D^{1/2} - \sqrt{d}I)\| \\ &\leqslant \|D^{1/2} - \sqrt{d}I\| \cdot \|N\| \cdot \|D^{1/2}\| \\ &\quad + \sqrt{d} \cdot \|N\| \cdot \|D^{1/2} - \sqrt{d}I\| \leqslant 3\gamma d. \end{aligned} \tag{D.13}$$

By using the triangle inequality and (D.13) we get

$$\begin{aligned} &\|N - (1/d)B\| \\ &\leqslant \|N - (1/d)A\| + (1/d) \cdot \|A - B\| \leqslant 4\gamma. \end{aligned} \tag{D.14}$$

Finally, we use Theorem A.2 (See Appendix A), which is a standard fact in matrix approximation theory: if two real symmetric matrices are close in spectral norm then their eigenvalues are close. From (D.14) and the fact that all eigenvalues of $(1/d)B$ except for the first and second one are 0, for each $i \in \{3, \dots, 2n\}$ we have

$$|\lambda_i| = |\lambda_i - 0| \leqslant \|N - \frac{1}{d}B\| \leqslant 4\gamma. \tag{D.15}$$

Similarly, from the fact that the second eigenvalue of $(1/d)B$ is $1 - 2b/d$ we get

$$|\lambda_2 - (1 - 2b/d)| \leqslant \|N - \frac{1}{d}B\| \leqslant 4\gamma,$$

that is, from hypothesis $\nu > 12\gamma$ and (D.15), $\lambda_2 \geqslant (1 + \delta)\lambda_3$ for some constant $\delta > 0$. This concludes the proofs of Lemma D.3 and Theorem 4.2.

D.3 Proof of Lemma 4.3

Let G be a randomly-generated graph according to $\mathcal{G}_{2n,p,q}$ with $a = pn$, $b = qn$ and $d = a + b$. Recall the definitions of A, D, N, P, λ_i and $\mathbf{w}_i$ ($i \in \{1, \dots, 2n\}$) in Section 2, and let B be defined as in Section D.2. Let us denote with A_i ($i \in \{1, 2\}$) the adjacency matrix of the subgraph of G induced by community V_i, with $A_B = \{A_{u,v-n}\}_{u \in V_1, v \in V_2}$ the matrix whose entry (i, j) is 1 iff there is an edge between the i-th node of V_1 and the j-th node of V_2, then

$$A = \begin{pmatrix} A_1 & A_B \\ A_B^{\intercal} & A_2 \end{pmatrix}.$$

We need the following technical lemmas.

Lemma D.4. *If $d > 5 \log n$ then for some positive constant c_3 it holds $\|A - B\| \leqslant c_3\sqrt{d}$ w.h.p.*

Proof. The lemma directly follows from Theorem 2.1 in [37] with $d' = 2d$ and the observation that, from the Chernoff bounds, all degrees are smaller than $2d$ w.h.p.

Lemma D.5. *If $d > 5 \log n$ then for some constant $c_4 > 0$ it holds w.h.p.*

$$\|dN - B\| \leqslant c_4\sqrt{d}.$$

The idea for proving Lemma D.5 is to use the triangle inequality to upper bound $\|dN - B\|$ in terms of $\|A - B\|$, which we can bound with Lemma D.4, and $\|B - 1/dD^{1/2}BD^{1/2}\|$, which we can upper bound by bounding $\|\sqrt{d}\mathbf{1} - D^{1/2}\mathbf{1}\|$ and $\|\sqrt{d}\boldsymbol{\chi} - D^{1/2}\boldsymbol{\chi}\|$ where $\mathbf{1}$ and $\boldsymbol{\chi}$ are the eigenvector corresponding to the only two non-zero eigenvalues of B. The complete proof of Lemma D.5 is deferred to Section D.4. As for the required bound on $\|\sqrt{d}\mathbf{1} - D^{1/2}\mathbf{1}\|^2 = \|\sqrt{d}\boldsymbol{\chi} - D^{1/2}\boldsymbol{\chi}\|^2 = \sum_{j \in V} |\sqrt{d} - \sqrt{d_j}|^2$, we provide it in the following lemma, whose proof is given in the full version [7].

Lemma D.6. *If $5 \log n < d < n^{\frac{1}{3} - c_5}$ for any constant $c_5 > 0$, it holds w.h.p.*

$$\sum_{j \in V} |\sqrt{d} - \sqrt{d_j}|^2 \leqslant 2n \text{ and}$$
$$\sum_{j \in V} |d - d_j|^2 \leqslant 2dn.$$

By combining Lemma D.5 and Theorem A.2 we have $|\lambda_i - \lambda_i'| \leqslant \|N - d^{-1}B\| = \mathcal{O}(1/\sqrt{d})$, where $\lambda_1' = 1$, $\lambda_2' = 1 - 2b/d$ and $\lambda_i' = 0$ for $i \in \{3, \dots, 2n\}$ are the eigenvalues of $d^{-1}B$. This proves the first two part of Lemma 4.3.

As for the third part, let us write $\mathbf{w}_2 = \mathbf{w}_{\mathbf{1}} + \mathbf{w}_{\boldsymbol{\chi}} + \mathbf{w}_{\perp}$ where $\mathbf{w}_{\mathbf{1}}$ and $\mathbf{w}_{\boldsymbol{\chi}}$ are the projection of $\mathbf{w}_2$ on $\mathbf{1}$ and $\boldsymbol{\chi}$ respectively, and $\mathbf{w}_{\perp}$ is the projection of $\mathbf{w}_2$ on the space orthogonal to $\mathbf{1}$ and $\boldsymbol{\chi}$.

Observe that the only non-zero eigenvalues of $(1/d)B$ are 1 and $(a - b)/d$. Thus, from Lemma D.5 and the Davis-Kahan theorem (Theorem A.3) with $M_1 = N$, $M_2 = \frac{1}{d}B$, $t = \lambda_2$, $\mathbf{x} = \mathbf{w}_2$ and $\delta = \lambda_2/2$, we get

$$\begin{aligned} \|\mathbf{w}_{\perp}\| &\leqslant \frac{4}{\lambda_2 \pi}\left\|N - \frac{1}{d}B\right\| \\ &\leqslant \mathcal{O}\left(\frac{1}{\sqrt{d}\lambda_2}\right) = \mathcal{O}\left(\frac{\sqrt{d}}{a - b}\right). \end{aligned} \tag{D.16}$$

As for $\mathbf{w}_{\mathbf{1}}$, we know that $\langle \mathbf{w}_2, D^{-1/2}\mathbf{1} \rangle = 0$, thus

$$\begin{aligned} \|\mathbf{w}_{\mathbf{1}}\| &= \frac{1}{\sqrt{2n}}\langle \mathbf{w}_2, \mathbf{1} - d^{-\frac{1}{2}}D^{\frac{1}{2}}\mathbf{1} \rangle \\ &\leqslant \frac{1}{\sqrt{2n}}\|\mathbf{w}_2\|\|\mathbf{1} - d^{-\frac{1}{2}}D^{\frac{1}{2}}\mathbf{1}\| \leqslant \frac{1}{\sqrt{d}}, \end{aligned} \tag{D.17}$$

where in the last inequality we used Lemma D.6.

By the law of cosines and the fact that $\sqrt{1 - x} \geqslant 1 - x$ for $x \in [0, 1]$ we have that

$$\begin{aligned} &\left\|\mathbf{w}_2 - \frac{1}{\sqrt{2n}}\boldsymbol{\chi}\right\|^2 \\ &= \|\mathbf{w}_2\|^2 + \left\|\frac{1}{\sqrt{2n}}\boldsymbol{\chi}\right\|^2 - 2\langle \mathbf{w}_2, \frac{1}{\sqrt{2n}}\boldsymbol{\chi} \rangle \\ &= 2 - 2\|\mathbf{w}_{\boldsymbol{\chi}}\| \\ &= 2 - 2\sqrt{1 - \|\mathbf{w}_{\mathbf{1}}\|^2 - \|\mathbf{w}_{\perp}\|^2} \\ &\leqslant 2\left(\|\mathbf{w}_{\mathbf{1}}\|^2 + \|\mathbf{w}_{\perp}\|^2\right) = \mathcal{O}\left(\frac{d}{(a - b)^2}\right), \end{aligned} \tag{D.18}$$

where in the last inequality we used (D.16) and (D.17). (D.18) implies that, with the exception of a set S of at most $\mathcal{O}(nd/(a - b)^2)$ nodes, we have

$$\left|\sqrt{2n}\mathbf{w}_2(i) - \boldsymbol{\chi}(i)\right| \leqslant \frac{1}{201}, \tag{D.19}$$

for each $i \in V/S$. From the Chernoff bound, we also have that w.h.p. $\sqrt{d/d_i} = 1 \pm 1/201$. Thus, (D.19) and the last fact imply that for each $i \in V/S$ it holds w.h.p.

$$\left|\sqrt{2nd}D^{-\frac{1}{2}}\mathbf{w}_2(i) - \boldsymbol{\chi}(i)\right| \leqslant \frac{1}{100},$$

concluding the proof. □

REMARK 1. *After looking at Lemma 4.3, one may wonder whether it could be enough to generalize Definition 2 to include "quasi-$(2n, d, b, \gamma)$-clustered graph", i.e. graphs that are $(2n, d, b, \gamma)$-clustered except for a small number of nodes which may have a much higher degree. In fact, this would be rather surprising: This higher-degree nodes may connect to the other nodes in such a way that would greatly perturb the eigenvalues and eigenvectors of the graph. In $\mathcal{G}_{2n,p,q}$, besides the fact that the nodes with degree much larger than d are few, it is also crucial that they are connected in a* non-adversarial *way, i.e. randomly.*

D.4 Proof of Lemma D.5 A simple application of the Chernoff bound and the union bound shows that w.h.p.

$$\sqrt{d}\|D^{-1/2}\| \leqslant 1 + \mathcal{O}\left(\sqrt{\frac{\log n}{d}}\right), \tag{D.20}$$

hence

$$\begin{aligned}
&\|dN - B\| \\
&= \|(\sqrt{d}D^{-1/2})A(\sqrt{d}D^{-1/2}) - B\| \\
&\leqslant \|\sqrt{d}D^{-1/2}\| \left\|A - \frac{1}{\sqrt{d}}D^{1/2}B\frac{1}{\sqrt{d}}D^{1/2}\right\| \|\sqrt{d}D^{-1/2}\| \\
&\leqslant \left\|A - \frac{1}{d}D^{1/2}BD^{1/2}\right\| \|\sqrt{d}D^{-1/2}\|^2 \\
&\leqslant \left(\|A - B\| + \|B - \tfrac{1}{d}D^{1/2}BD^{1/2}\|\right)\left(1 + \mathcal{O}\left(\sqrt{\frac{\log n}{d}}\right)\right). \qquad \text{(D.21)}
\end{aligned}$$

Thanks to Lemma D.4, it holds $\|A - B\| = \mathcal{O}(\sqrt{d})$. Hence, in order to conclude the proof, it remains to show that $\|B - d^{-1}D^{1/2}BD^{1/2}\| = \mathcal{O}(\sqrt{d})$. We do that by observing that

$$\begin{aligned}
\left\|B - \frac{1}{d}D^{1/2}BD^{1/2}\right\| &\leqslant \left\|B - \frac{1}{\sqrt{d}}BD^{1/2}\right\| \\
&+ \left\|\frac{1}{\sqrt{d}}BD^{1/2} - \frac{1}{d}D^{1/2}BD^{1/2}\right\|, \qquad \text{(D.22)}
\end{aligned}$$

and by upper-bounding the two terms on the right hand side. The two only non-zero eigenvalues of B are $a + b$ and $a - b$, with corresponding eigenvectors $(2n)^{-1/2}\,\mathbf{1}$ and $(2n)^{-1/2}\,\boldsymbol{\chi}$, therefore we can write $B = d/(2n)\,\mathbf{1}\mathbf{1}^\intercal + (a - b)/(2n)\,\boldsymbol{\chi}\boldsymbol{\chi}^\intercal$, which implies that

$$\begin{aligned}
B - \frac{1}{\sqrt{d}}BD^{1/2} = \frac{\sqrt{d}}{2n}\,&\mathbf{1}\,(\sqrt{d}\,\mathbf{1} - D^{1/2}\,\mathbf{1})^\intercal \\
&+ \frac{a - b}{\sqrt{d}\,2n}\,\boldsymbol{\chi}\,(\sqrt{d}\,\boldsymbol{\chi} - D^{1/2}\,\boldsymbol{\chi})^\intercal.
\end{aligned}$$

It follows that, for an arbitrary unitary vector $\mathbf{x}$ it holds

$$\begin{aligned}
&\left\|\left(B - \frac{1}{\sqrt{d}}BD^{1/2}\right)\mathbf{x}\right\| \\
&\leqslant \left\|\frac{\sqrt{d}}{2n}\,\mathbf{1}\,(\sqrt{d}\,\mathbf{1} - D^{1/2}\,\mathbf{1})^\intercal\mathbf{x}\right\| \\
&\qquad + \left\|\frac{a - b}{\sqrt{d}\,2n}\,\boldsymbol{\chi}\,(\sqrt{d}\,\boldsymbol{\chi} - D^{1/2}\,\boldsymbol{\chi})^\intercal\mathbf{x}\right\| \\
&= \frac{\sqrt{d}}{2n}\,\|\mathbf{1}\|\,|(\sqrt{d}\,\mathbf{1} - D^{1/2}\,\mathbf{1})^\intercal\mathbf{x}| \\
&\qquad + \frac{a - b}{\sqrt{d}\,2n}\,\|\boldsymbol{\chi}\|\,|(\sqrt{d}\,\boldsymbol{\chi} - D^{1/2}\,\boldsymbol{\chi})^\intercal\mathbf{x}| \\
&\leqslant \frac{\sqrt{d}}{\sqrt{2n}}\,\left\|\sqrt{d}\,\mathbf{1} - D^{1/2}\,\mathbf{1}\right\| \cdot \|\mathbf{x}\| \\
&\qquad + \frac{a - b}{\sqrt{2dn}}\,\left\|\sqrt{d}\,\boldsymbol{\chi} - D^{1/2}\,\boldsymbol{\chi}\right\| \cdot \|\mathbf{x}\| \leqslant 2\sqrt{d}, \qquad \text{(D.23)}
\end{aligned}$$

where we used the triangle inequality, the fact that $\|\mathbf{1}\| = \|\boldsymbol{\chi}\| = \sqrt{2n}$, the Cauchy-Schwartz inequality, Lemma D.6 and $a - b < d$. As for the other term on the r.h.s. of (D.22), we have that w.h.p.

$$\begin{aligned}
&\left\|\frac{1}{\sqrt{d}}BD^{1/2} - \frac{1}{d}D^{1/2}BD^{1/2}\right\| \\
&\leqslant \left\|B - \frac{1}{\sqrt{d}}D^{1/2}B\right\| \frac{1}{\sqrt{d}}\|D^{1/2}\| \\
&\leqslant 2\sqrt{d}\left(1 + \mathcal{O}\left(\sqrt{\frac{\log n}{d}}\right)\right), \qquad \text{(D.24)}
\end{aligned}$$

where in the last inequality we used (D.20) and that for any matrix M it holds $\|M\| = \|M^\intercal\|$. Finally, (D.23) and (D.24) togheter implies the desired upper bound on (D.22) and thus (D.21), concluding the proof. □

D.5 Proof of Theorem 4.2 For any vector $\mathbf{x}$, we can write

$$\mathbf{x}^{(t)} = P^t\mathbf{x} = \sum_{i=1}^{2n} a_i\lambda_i^t D^{-1/2}\mathbf{w}_i = \alpha_1\mathbf{1} + a_2\lambda_2^t D^{-1/2}\mathbf{w}_2 + \mathbf{e}^{(t)},$$

where $\alpha_1 = \mathbf{1}^\intercal D\mathbf{x}/\|D^{1/2}\mathbf{1}\|$ and $\|\mathbf{e}^{(t)}\| \leqslant 4\lambda^t\|\mathbf{x}\|$.

From Lemma 4.3 (Claim 3) we have that for at least $2n - \mathcal{O}(nd/(a-b)^2)$ entries i of $D^{-1/2}\mathbf{w}_2$, we get $|\sqrt{2nd}(D^{-1/2}\mathbf{w}_2)(i) - \boldsymbol{\chi}(i)| \leqslant \frac{1}{100}$, that is

$$\begin{aligned}(D^{-1/2}\mathbf{w}_2)(i) &\geqslant \frac{99}{100\sqrt{2nd}} \quad \text{if } i \in V_1 \cap S \text{ and}\\ (D^{-1/2}\mathbf{w}_2)(i) &\leqslant -\frac{99}{100\sqrt{2nd}} \quad \text{if } i \in V_2 \cap S.\end{aligned}$$

Thus, we get

$$\begin{aligned}&\left|\mathbf{x}^{(t)} - \mathbf{x}^{(t-1)}\right|\\ &= \left|a_2\lambda_2^{t-1}(\lambda_2 - 1)D^{-1/2}\mathbf{w}_2 + \mathbf{e}^{(t)} + \mathbf{e}^{(t-1)}\right|\\ &\leqslant \left|a_2\lambda_2^{t-1}(\lambda_2 - 1)D^{-1/2}\mathbf{w}_2\right| + \left|\mathbf{e}^{(t)} - \mathbf{e}^{(t-1)}\right| \tag{D.25}\end{aligned}$$

and, when $t - 1 \geqslant \log\left(\frac{16\sqrt{2n}}{|a_2|(1-\lambda_2)}\right)/\log\left(\frac{\lambda_2}{\lambda}\right)$, from (D.25) it follows that

$$\begin{aligned}(\mathbf{x}^{(t)} - \mathbf{x}^{(t-1)})(i) &\geqslant \frac{99}{200\sqrt{2nd}}\,a_2\lambda_2^{t-1}(\lambda_2 - 1)\\ &\quad \text{if } i \in V_j \cap S \text{ and}\\ (\mathbf{x}^{(t)} - \mathbf{x}^{(t-1)})(i) &\leqslant -\frac{99}{200\sqrt{2nd}}\,a_2\lambda_2^{t-1}(\lambda_2 - 1)\\ &\quad \text{if } i \in V_{3-j} \cap S.\end{aligned}$$

either for $j = 1$ or for $j = 2$. Since $|S| > n - \mathcal{O}(nd/(a-b)^2)$, we thus get a $\mathcal{O}(d/(a-b)^2)$-weak reconstruction. □

E More communities

Recall the definition of negative and positive type in Section 5. In this section we prove Theorem 5.1. The proof is divided in the following two lemmas.

LEMMA E.1. *Pick $\mathbf{x} \sim \{-1,1\}^{kn}$ u.a.r. Then, with high probability, the vertices of V_1 are either all of positive type or all of negative type. Furthermore, the two events have equal probability.*

Proof. We will write

$$\mathbf{x} = \mathbf{x_1} + \mathbf{x}_{V_1} + \mathbf{x}_{\perp_1} + \mathbf{x}_\perp,$$

where $\mathbf{x_1}$ is the component of $\mathbf{x}$ parallel to $\mathbf{1}$, $\mathbf{x}_{V_1}$ is the component parallel to the vector $\mathbf{1}_{V_1} - k^{-1}\mathbf{1}_V$, $\mathbf{x}_{\perp_1}$ is the component in the eigenspace of λ_2 and orthogonal to $\mathbf{1}_{V_1} - k^{-1}\mathbf{1}_V$, and $\mathbf{x}_\perp$ is the component orthogonal to $\mathbf{1}$ and to the eigenspace of λ_2.

For the above the make sense, $\mathbf{1}_{V_1} - k^{-1}\mathbf{1}_V$ must be an eigenvector of λ_2, which is easily verified because its entries sum to zero and they are constant within components.

An important observation, and the reason for picking the above decomposition, is that $\mathbf{x}_{\perp_1}$ is zero in V_1. The reason is that $\mathbf{x}_{\perp_1}$ has to be orthogonal to $\mathbf{1}_V$ and to $\mathbf{1}_{V_1} - k^{-1}\mathbf{1}_V$ so from

$$\langle \mathbf{x}_{\perp_1}, \mathbf{1}_V\rangle = \langle \mathbf{x}_{\perp_1}, \mathbf{1}_{V_1} - k^{-1}\mathbf{1}_V\rangle = 0,$$

we deduce

$$\langle \mathbf{x}_{\perp_1}, \mathbf{1}_{V_1}\rangle = 0.$$

Thus, the entries of $\mathbf{x}_{\perp_1}$ sum to zero within V_1, but, being in the eigenspace of λ_2, the entries of $\mathbf{x}_{\perp_1}$ are constant within components, and so they must be all zero within V_1.

Now we have

$$P^t\mathbf{x} = \mathbf{x_1} + \lambda_2^t\mathbf{x}_{V_1} + \lambda_2^t\mathbf{x}_{\perp_1} + P^t\mathbf{x}_\perp,$$

and so, for each $v \in V_1$ it holds

$$\begin{aligned}&(P^{t+1}\mathbf{x})_v - (P^t\mathbf{x})_v\\ &= \lambda_2^t \cdot (1-\lambda_2)(\mathbf{x}_{V_1})_v + ((P^{t+1} - P^t)\mathbf{x}_\perp)_v. \tag{E.26}\end{aligned}$$

For $t > T$, the hypothesis $\lambda < (1-\varepsilon)\lambda_2$ implies that

$$\begin{aligned}&|(P^t\mathbf{x}_\perp)_v| \leqslant ||P^t\mathbf{x}_\perp||_\infty\\ &\leqslant ||P^t\mathbf{x}_\perp|| \leqslant \lambda^t||\mathbf{x}_\perp|| \leqslant \sqrt{n}\cdot\lambda^t \leqslant \frac{1}{n^{1.5}}\lambda_2^t. \tag{E.27}\end{aligned}$$

Moreover, for each $v \in V_1$ we have

$$\begin{aligned}|(\mathbf{x}_{V_1})_v| &= \frac{\langle \mathbf{x}, \mathbf{1}_{V_1} - k^{-1}\mathbf{1}_V\rangle}{\|\mathbf{1}_{V_1} - k^{-1}\mathbf{1}_V\|^2}\,(1 - k^{-1})\\ &= \frac{k}{(k-1)n}\left(\sum_{i\in V_1} x_i - \sum_{i\in V}\frac{x_i}{k}\right)\left(\frac{k-1}{k}\right)\\ &= \frac{1}{n}\left(\sum_{i\in V_1} x_i - \sum_{i\in V}\frac{x_i}{k}\right),\end{aligned}$$

and

$$\begin{aligned}||\mathbf{x}_{V_1}|| &= \frac{\langle \mathbf{x}, \mathbf{1}_{V_1} - k^{-1}\mathbf{1}_V\rangle}{\|\mathbf{1}_{V_1} - k^{-1}\mathbf{1}_V\|}\\ &= \sqrt{\frac{k}{(k-1)n}}\left(\sum_{i\in V_1} x_i - \sum_{i\in V}\frac{x_i}{k}\right),\end{aligned}$$

which imply that

$$|(\mathbf{x}_{V_1})_v| = \sqrt{(1-1/k)/n}\,\|\mathbf{x}_{V_1}\|. \tag{E.28}$$

Finally, note that by Lemma B.2 it holds w.h.p. $||\mathbf{x}_{V_1}|| \geqslant \frac{1}{n}||\mathbf{x}|| \geqslant \sqrt{k/n}$.

The latter fact together with (E.27) and (E.28) imply that w.h.p. the sign of (E.26) is the same as the sign of $(\mathbf{x}_{V_1})_v$, which is the same for all elements of V_1 and is equally likely to be positive or negative.

Of course the same statement is true if we replace V_1 by V_i for any $i = 1, \dots, k$; by a union bound, it is also true for all i simultaneously with high probability.

Lemma E.2. *Pick $\mathbf{x} \sim \{-1,1\}^{kn}$ u.a.r. There is an absolute constant p (e.g., $p = \frac{1}{100}$) such that, with probability at least p, all vertices of V_1 have the same type, all vertices of V_2 have the same type, and the types are different.*

Proof. This time we write

$$\mathbf{x} = \mathbf{x_1} + \mathbf{x}_{V_{1\oplus 2}} + \mathbf{x}_{V_{1\ominus 2}} + \mathbf{x}_{\perp_{1,2}} + \mathbf{x}_{\perp}$$

where

- $\mathbf{x_1}$ is the component parallel to $\mathbf{1}_V$,
- $\mathbf{x}_{V_{1\oplus 2}}$ is the component parallel to $\mathbf{1}_{V_1} + \mathbf{1}_{V_2} - \frac{2}{k}\mathbf{1}_V$,
- $\mathbf{x}_{V_{1\ominus 2}}$ is the component parallel to $\mathbf{1}_{V_1} - \mathbf{1}_{V_2}$,
- $\mathbf{x}_{\perp_{1,2}}$ is the component in the eigenspace of λ_2 and orthogonal to $\mathbf{x}_{V_{1\oplus 2}}$ and $\mathbf{x}_{V_{1\ominus 2}}$,
- $\mathbf{x}_{\perp}$ is the rest.

Similarly to the proof of Lemma E.1, the important observations are that $\mathbf{x}_{V_{1\oplus 2}}$ and $\mathbf{x}_{V_{1\ominus 2}}$ are in the eigenspace of λ_2, and that $\mathbf{x}_{\perp_{1,2}}$ is zero in all the coordinates of V_1 and of V_2.

Thus, for each $v \in V_1 \cup V_2$ we have

$$\begin{aligned}(P^{t+1}\mathbf{x})_v - (P^t\mathbf{x})_v &= \lambda_2^t(1-\lambda_2)(\mathbf{x}_{V_{1\oplus 2}} + \mathbf{x}_{V_{1\ominus 2}})_v \\ &\quad + ((P^{t+1} - P^t)\mathbf{x}_{\perp})_v.\end{aligned} \tag{E.29}$$

From (E.29) it is easy to see that if $\mathbf{x}$ is such that, for every $v \in V_1 \cup V_2$, we have the two conditions

$$|(\mathbf{x}_{V_{1\oplus 2}})_v| \leqslant \frac{3}{4}|(\mathbf{x}_{V_{1\ominus 2}})_v| \quad \text{and} \tag{E.30}$$

$$\begin{aligned}&|((P^{t+1} - P^t)\mathbf{x}_{\perp})_v| \\ &\qquad \leqslant \frac{1}{8}\lambda_2^t \cdot (1-\lambda_2) \cdot |(\mathbf{x}_{V_{1\ominus 2}})_v|,\end{aligned} \tag{E.31}$$

then such an $\mathbf{x}$ satisfies the conditions of the Lemma, that is all the elements in V_1 have the same type, all the elements of V_2 have the same type, and the types are different. Now note that, since

$$|(\mathbf{x}_{V_{1\oplus 2}})_v| = \frac{1}{2n}\left(\sum_{i \in V_1} x_i + \sum_{i \in V_1} x_i - \frac{2}{k}\sum_{i \in V} x_i\right) \quad \text{and}$$

$$|(\mathbf{x}_{V_{1\ominus 2}})_v| = \frac{1}{2n}\left(\sum_{i \in V_1} x_i - \sum_{i \in V_2} x_i\right),$$

if $\mathbf{x}$ satisfies

$$2\sqrt{n} \leqslant \sum_{v \in V_1} x_v \leqslant 3\sqrt{n}, \tag{E.32}$$

$$-2\sqrt{n} \leqslant \sum_{v \in V_2} x_v \leqslant -\sqrt{n} \quad \text{and} \tag{E.33}$$

$$0 \leqslant \sum_{v \in V/(V_1 \cup V_2)} x_v \leqslant \frac{1}{10}\sqrt{kn}, \tag{E.34}$$

then (E.30) is satisfied, and note that (E.32), (E.33) and (E.34) are independent and each happens with constant probability.

Finally, observe that if (E.30) holds then (E.31) is satisfied with high probability when $t > T$.

It is enough to pick $\ell = \log(3n)$ to have, with high probability, that the signatures are well defined and they are the same within each community and different between communities. The first lemma guarantees that, with high probability, for all ℓ vectors, all vertices within each community have the same type. The second lemma guarantees that, with high probability, the signatures are different between communities.