Average whenever you meet: Opportunistic protocols for community detection

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

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

Pasin Manurangsi
U.C. Berkeley
Berkeley, California
pasin@berkeley.edu

Emanuele Natale
Max Planck Institute for Informatics
Saarbrcken, Germany
enatale@mpi-inf.mpg.de

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

Prasad Raghavendra
U.C. Berkeley
Berkeley, CA, United States
raghavendra@berkeley.edu

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

Abstract

Consider the following asynchronous, opportunistic communication model over a graph G: in each round, one edge is activated uniformly and independently at random and (only) its two endpoints can exchange messages and perform local computations. Under this model, we study the following random process: *The first time a vertex is an endpoint of an active edge, it chooses a random number, say* ± 1 *with probability* $1/2$*; then, in each round, the two endpoints of the currently active edge update their values to their average.* We show that, if G exhibits a two-community structure (for example, two expanders connected by a sparse cut), the values held by the nodes will collectively reflect the underlying community structure over a suitable phase of the above process, allowing efficient and effective recovery in important cases.

In more detail, we first provide a first-moment analysis showing that, for a large class of almost-regular clustered graphs that includes the *stochastic block model*, the expected values held by all but a negligible fraction of the nodes eventually reflect the underlying cut signal. We prove this property emerges after a "mixing" period of length $\mathcal{O}(n \log n)$. We further provide a second-moment analysis for a more restricted class of regular clustered graphs that includes the *regular stochastic block model*. For this case, we are able to show that most nodes can efficiently and locally identify their community of reference over a suitable time window. This results in the first opportunistic protocols that approximately recover community structure using only logarithmic (or polylogarithmic, depending on the sparsity of the cut) work per node. Even for the above class of regular graphs, our second moment analysis requires new concentration bounds on the product of certain random matrices that are technically challenging and possibly of independent interest.

Keywords: Distributed Community Detection, Asynchronous Protocols, Random Processes, Spectral Analysis.

1 Introduction

Consider the following, elementary distributed process on an undirected graph $G = (V, E)$ with $|V| = n$ nodes and $|E| = m$ edges. Each node v holds a real number x_v (which we call the *state* of node v); at each time step, one random edge $\{u, v\}$ becomes active and its endpoints u and v update their states to their average.

Viewed as a protocol, the above process is consistent with asynchronous, opportunistic communication models, such as those considered in [AAER07] for *population-protocols*; here, in every round, one edge is activated uniformly and independently at random and (only) its two endpoints can exchange messages and perform local computations in that round[1]. We further assume no global clock is available (nodes can at most count the number of local activations) and that the network is *anonymous*, i.e., nodes are not aware of theirs or their neighbors' identities and all nodes run the same process at all times.

The long-term behavior of the process outlined above is well-understood: assuming G is connected, for each initial global state $\mathbf{x} \in \mathbb{R}^V$ the system converges to a global state in which all nodes share a common value, namely, the average of their initial states. A variant of an argument of Boyd et al. [BGPS06] shows that convergence occurs in $\mathcal{O}\left(\frac{1}{\lambda_2} n \log n\right)$ steps, where λ_2 is the second smallest eigenvalue of the normalized Laplacian of G.

Suppose now that G is *well-clustered*, i.e. it exhibits a *community structure* which in the simplest case consists of two equal-sized expanders, connected by a sparse cut: This structure arises, for instance, when the graph is sampled from the popular *stochastic block model* [MNS16] $\mathcal{G}_{n,p,q}$ for $p \gg q$ and $p \geqslant \log n/n$. If we let the averaging process unfold on such a graph, for example starting from an initial ± 1 random global state, one might reasonably expect a faster, transient convergence toward some local average within each community, accompanied by a slower, global convergence toward the average taken over the entire graph. If, as is likely the case, a gap exists between the local averages of the two communities, the global state during the transient phase would reflect the graph's underlying community structure. This intuition suggests the main questions we address in this paper:

Is there *a phase in which the global state carries information about community structure*? If so, *how strong is the corresponding "signal"*? Finally, *can nodes leverage local history to recover this information*?

The idea of using averaging local rules to perform distributed community detection is not new: In [BCN+17], Becchetti et al. consider a deterministic dynamics in which, at every round, each node updates its local state to the average of its neighbors. The authors show that this results in a fast clustering algorithm with provable accuracy on a wide class of almost-regular graphs that includes the stochastic block model. We remark that the algorithm in [BCN+17] works in a synchronous, parallel communication model where every node exchanges data with all its neighbors in each round. This implies considerable work and communication costs, especially when the graph is dense. On the other hand, each step of the process is described by the same matrix and its evolution unfolds according to the power of this matrix applied to the initial state. In contrast, the averaging process we consider in this paper is considerably harder to analyze than the one in [BCN+17], since each step is described by a random, possibly different averaging matrix.

Differently from [BCN+17], our goal here is the design of simple, lightweight protocols for

[1] In an essentially equivalent continuous-time model, each edge has a clock that ticks at random intervals with a Poisson distribution of average 1; when the clock ticks, then the edge endpoints become activated. For t larger than $n \log n$, the behavior of the continuous time process for t/n units of time and the behavior of the discrete-time process for t steps are roughly equivalent.

fully-decentralized community detection which fit the asynchronous, opportunistic communication model, in which a (random) link activation represents an opportunistic meeting that the endpoints can exploit to exchange one-to-one messages. More specifically, by "lightweight" we mean protocols that require minimalistic assumptions as to network capabilities, while performing their task with minimal work, storage and communication per node (at most logarithmic or polylogarithmic in our case). In this respect, any clustering strategies (like the one in [SZ17]) which construct (and then work over) some static, sparse subgraph of the underlying graph are unfeasible in the opportunistic model we consider here. This restrictive setting is motivated by network scenarios in which individual agents need to autonomously and locally uncover underlying, implicit communities of which they are members. This has widespread applicability, for example in communication systems where lightweight data can be locally shared via wireless opportunistic meetings when agents come within close range [WWA12].

We next discuss what it means to recover the "underlying community structure" in a distributed setting, a notion that can come in stronger or weaker flavors. Ideally, we would like the protocol to reach a state in which, at least with high probability, each node can use a simple rule to assign itself one of two possible labels, so that labelling within each community is consistent and nodes in different communities are assigned different labels. Achieving this corresponds to *exact (block) reconstruction*. The next best guarantee is *weak (block) reconstruction* (see Definition 2.3). In this case, with high probability the above property is true for all but a small fraction of misclassified nodes. In this paper, we introduce a third notion, which we call *community-sensitive labeling* (CSL for short): in this case, there is a predicate that can be applied to pairs of labels so that, for all but a small fraction of outliers, the labels of any two nodes within the same community satisfy the predicate, whereas the converse occurs when they belong to different communities[2]. In this paper, informally speaking, nodes are labelled with binary signatures of logarithmic length, while two labels satisfy the predicate whenever their Hamming distance is below a certain threshold. This introduces a notion of similarity between nodes of the graph, with labels behaving like profiles that reflect community membership[3]. Note that this weaker notion of community-detection allows nodes to locally tell "friends" in their community from "foes" in the other community, which is the main application of distributed community detection in the opportunistic setting we consider here.

1.1 Our results

First moment analysis. Our first contribution is an analysis of the expected evolution of the averaging process over a wide class of almost-regular graphs (see Definition 2.1) that possess a hidden and balanced partition of the nodes with the following properties: (i) The cut separating the two communities is sparse, i.e., it contains $o(m)$ edges; (ii) the subgraphs induced by the two communities are expanders, i.e., the gap $\lambda_3 - \lambda_2$ between the third and the second eigenvalues of the normalized Laplacian matrix $\mathcal{L}$ of the graph is constant. The above conditions on the underlying graph are satisfied, for instance, by graphs sampled from the stochastic block model[4] $\mathcal{G}_{n,p,q}$ for $q = o(p)$ and $p \geqslant \log n/n$.

Let $L = D - A$ be the Laplacian matrix of G. The first moment analysis considers the *deterministic* process described by the linear equation $\mathbf{x}^{(t+1)} = \overline{W}^t \cdot \mathbf{x}^{(0)}$ $(t \geqslant 1)$, where $\mathbf{x}^{(0)} = \mathbf{x}$ is the vector with components the nodes' initial random values and $\overline{W} := \mathbb{E}[W] = I - \frac{1}{2m}L$ is the expectation of the random matrix that describes a single step of the averaging process. While a formal proof of the above equation can be found in Section 3, our analysis reveals that

[2]Note that a weak reconstruction protocol entails a community-sensitive labeling. In this case, the predicate is true if two labels are the same.

[3]Hence the phrase *community-sensitive Labeling* we use to refer to our approach.

[4]See Subsection 1.2 for the definition of $\mathcal{G}_{n,p,q}$ and for more details about our results for $\mathcal{G}_{n,p,q}$.

the expected values held by the nodes are correlated with the underlying cut. This phenomenon follows from structural connections between the underlying graph's community structure and some spectral properties of $\overline{W}$. This allows us to show that, after an initial *"mixing"* phase of $\Theta(n \log n)$ rounds and for all but $o(n)$ nodes, the following properties hold: (i) There exists a relatively large time window in which the signs of the expected values of nodes are correlated with the community they belong to. (ii) The expected values of nodes belonging to one of the communities increase in each round, while those of nodes in the other community decrease.

The formal statements of the above claims can be found in Theorem 3.1. Here, we note that these results suggest two different local criteria for community-sensitive labeling: (i) According to the first one, every node uses the sign of its own state within the aforementioned time window to set the generic component of its binary label (we in fact use independent copies of the averaging process to get binary labels of logarithmic size - see Protocol SIGN-LABELING in Section 4.2). (ii) According to the second criterion, every node uses the signs of fluctuations of its own value along consecutive rounds to set the generic component of its binary label (see Protocol JUMP-LABELING in Section 5.2) [5].

The above analysis describes the "expected" behaviour of the averaging process over a large class of well-clustered graphs, at the same time showing that our approach might lead to efficient, opportunistic protocols for block reconstruction. Yet, designing and analyzing protocols with provable, high-probability guarantees, requires addressing the following questions:

1. *Do realizations of the averaging process approximately follow its expected behavior with high, or even constant, probability?*
2. *If this is the case, how can nodes locally and asynchronously recover the cut signal, let alone guess the "right" global time window?*

Second moment analysis. The first question above essentially requires a characterization of the *variance* of the process over time, which turns out to be an extremely challenging task. The main reason for this is that a realization of the averaging process is now described by an equation of the form $\mathbf{x}^{(t)} = W_t \cdot \ldots\ldots \cdot W_1\mathbf{x}$, where the W_i's are sampled independently and uniformly from some matrix distribution (see Eq. (1) in Section 3). Here, matrix W_i "encodes" both the i-th edge selected for activation and the averaging of the values held by its endpoints at the end of the $(i-1)$-th step.

Not much is known about concentration of the products of identically distributed random matrices, but we are able to accurately characterize the class of regular clustered graphs. We point out that many of the technical results and tools we develop to this purpose apply to far more general settings than the regular case and may be of independent interest. In more detail, we are able to provide accurate upper bounds on the norm of $\mathbf{x}^{(t)}$'s projection onto the subspace spanned by the first and second eigenvector of $\overline{W}$ (see the proof's outline of Theorem 3.1 and Lemma B.2) for a class of regular clustered graphs that includes the *regular stochastic block model*[6] [BCN+17, BDG+15, MNS14] - see Definition 2.2.

These bounds are derived separately for two different regimes, defined by the sparseness of the cut separating the two communities. Assuming a good inner expansion of the communities, the first concentration result concerns cuts of size $o(m/\log^2 n)$ and it is given in Subsection 4.1 while, for the case of cuts of size up to αm for any $\alpha < 1$, the obtained concentration results are described in Subsection D.1.

[5] Having a node set its label within the correct time window is technically challenging in the asynchronous, opportunistic communication model we consider. This issue is briefly discussed in this and the following sections and formally addressed in Appendix C.4).

[6] See Subsection 1.2 for more details about our results for regular stochastic block models.

These bounds alone are not sufficient to prove accuracy of the clustering criteria, due to the asynchronous nature of the communication model we consider, whereby every node only updates its value once every m/d rounds in expectation, where d is the degree of the nodes.

Desynchronization. The variance analysis outlined in the previous paragraph ensures that, for any fixed step t, the actual states of a large fraction of the nodes are "close" to their expectations, with high probability. Unfortunately, the asynchrony of the model we consider does not allow to easily apply this result (e.g., using a union bound) to prove that most nodes will eventually label themselves within the right global window and in a way that is consistent with the graph's community structure. Rather, we show that there exists a large fraction of *non-ephemeral* "good nodes" whose states remain close to their expectations over a suitable time-window. The technical form of the concentration bound and the relative time window again depend on the sparsity of the cut: See Definition 4.1 and Lemma 4.2 for sparse cuts and Theorem 5.1 for dense cuts, respectively.

Distributed community detection. We exploit the second moment analysis and the desynchronization above to devise two different opportunistic protocols for community detection on regular clustered graphs.
- In the case of sparse cuts (i.e. of size $o(m/\log^2 n)$), the obtained bound on the variance of non-ephemeral nodes (see Lemma 4.2) holds over a time window that essentially equals the one "suggested" by our first moment analysis. This allows us to give rigorous bounds on the performance of the opportunistic Protocol SIGN-LABELING based on the sign criterion (see Section 4.2). This "good" time-window begins after $\mathcal{O}(n \log n)$ rounds: So, if the underlying graph has dense communities and a sparse cut, nodes can collectively compute an accurare labeling *before* the global mixing time of the graph. For instance, if the cut is $\mathcal{O}(m/n^\gamma)$, for some constant $\gamma < 2$, our protocol is polynomially faster than the global mixing time. In more detail, we prove that, given any regular clustered graph with cut of size $o(m/\log^2 n)$, Protocol SIGN-LABELING performs community-sensitive labeling for $n - o(n)$ nodes within global time[7] $\mathcal{O}(n \log^2 n)$ and with work per node $\mathcal{O}(\log^2 n)$, with high probability (see Theorem 4.3 and its Corollary 4.2 for formal and more general statements about the performances of the protocol). Importantly enough, the costs of our first protocol do not depend on the cardinality of the edge set E.
- The bound on the variance that allows us to adopt the sign-based criterion above does not hold when the cut is not sparse, i.e., whenever it is $\omega(m/\log^2 n)$. For such dense cuts, we use a different bound on the variance of nodes' values given in Theorem 5.1, which starts to hold after the global mixing time of the underlying graph and over a time window of length $\Theta(n^2)$. In this case, the specific form of the concentration bound leads to adoption of the second clustering criterion suggested by our first moment analysis, i.e., the one based on monotonicity of the values of non-ephemeral nodes. To this aim, we consider a "lazy" version of the averaging process equipped with a local clustering criterion, whereby nodes use the signs of fluctuations of their own values along consecutive rounds to label themselves (see Protocol JUMP-LABELING in Section 5.2). A restricted but relevant version of the algorithmic result we achieve in this setting can be stated as follows (see Theorem 5.3 and its corollaries for more general statements): Given any regular clustered graph consisting of two expanders as communities and a cut of size up to αm (for any $\alpha < 1$), the opportunistic protocol JUMP-LABELING achieves weak reconstruction for a fraction $(1-\varepsilon)n$ nodes (where ε is an arbitrary positive constant), with high probability. The protocol converges within $\mathcal{O}(n \log n\, (\log^2 n + m/m_{1,2}))$ rounds (where we named $m_{1,2} = |E(V_1, V_2)|$ the size of the cut) and every node performs $\mathcal{O}(\log^2 n + (m/m_{1,2}) \log n)$ work,

[7] The extra logarithmic factor is needed to let every node update each component of its $\Theta(\log n)$-size binary label, independently (see Subsection 4.2 for details).

with high probability. Notice that this second protocol achieves a stronger form of community detection than the first one, but it is less efficient, especially when the underlying graph is dense and exhibits a sparse cut (it requires to "wait" for the global mixing time of the graph). On the other hand, the two protocols have comparable costs in the parameter ranges they were designed for.[8]

1.2 Comparison to previous work

We earlier compared our results to those of [BCN$^+$17]. The advantage of [BCN$^+$17] is that their analysis achieves concentration over a class of graphs that are almost-regular and extends to the case of more than two communities. Furthermore, in graphs in which the indicator of the cut is an eigenvector of λ_2, the algorithm of [BCN$^+$17] achieves exact reconstruction. On the other hand, as previously remarked, the advantage of our work over [BCN$^+$17] is that, for the first time, it applies to the asynchronous opportunistic model and the communication cost per node does not depend on the degree of the graph, so it is much more efficient in dense graphs.

If we do not restrict to asynchronous and/or opportunistic protocols, recently, in [SZ17], Sun and Zanetti[9] introduced a synchronous, averaging-based protocol that first computes a fixed random subgraph of the underlying graph and then, working on this sparse subgraph, returns an efficient block-reconstruction for a wide class of almost-regular clustered graphs including the stochastic block model. We remark that, besides having no desynchronization issue to deal with, their second moment analysis uses Chernoff-like concentration bounds on some random submatrix (rather than a product of them, as in our setting) which essentially show that, under reasonable hypothesis, the signal of the cut can still be recovered from the corresponding sparse subgraph the algorithm works on.

Further techniques for community detection and spectral clustering exist, which are not based on averaging. In particular, Kempe and McSherry showed that the top k eigenvectors of the adjacency matrix of the underlying graph can be computed in a distributed fashion [KM04]. These eigenvectors can then be used to partition the graph; in our settings, since we assume that the indicator of the cut is the second eigenvector of the graph, applying Kempe and McSherry's algorithm with $k = 2$ immediately reveals the underlying partition. Again, we note here that the downside of this algorithm is that it is synchronous and quite complex. In particular, the algorithm requires a computation of $Q_1\mathbf{x}$, for which $\lambda_2^{-1} \log n$ work per node is a bottleneck, while our first algorithm only requires $\lambda_3^{-1} \log n$ work per node, a difference that can become significant for very sparse cuts.

At a technical level, we note that our analysis establishes concentration results for products of certain i.d.d. random matrices; concentrations of such products have been studied in the ergodic theory literature [CPV93, Ips15], but under assumptions that are not met in our setting, and with convergence rates that are not suitable for our applications.

While we only focused on decentralized settings so far, we note that the question of community detection, especially in stochastic block models, has been extensively studied in centralized computational models [ABH14, CO10, DKMZ11, DF89, HLL83, JS98, McS01]. The stochastic block model offers a popular framework for the probabilistic modelling of graphs that exhibit good clustering or community properties. In its simplest version, the random graph $\mathcal{G}_{n,p,q}$ consists of n 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 $n/2$; edges linking nodes belonging to the same

[8] As for the fraction of outliers guaranteed by the two protocols, please see the technical discussion after Corollary 5.5.

[9] Before the technical report in [SZ17], the same authors in [SZ16] presented a synchronous distributed algorithm able to perform approximate reconstruction with multiple communities. Sun and Zanetti then discovered a gap in their analysis (personal communication), and they retracted the claims in [SZ16].

partition appear in E independently at random with probability $p = p(n)$, while edges connecting nodes from different partitions appear with probability $q = q(n) < p$. In the centralized setting, the focus of most studies on stochastic block models is on determining the threshold at which weak recovery becomes possible, rather than simplicity or running time of the algorithm (as most algorithms are already reasonably simple and efficient). After a remarkable line of work [DKMZ11, MNS14, Mas14, MNS13], such a threshold has now been precisely determined.

Calling $a = pn$ and $b = qn$, it is known [BCN+17] that graphs sampled from $\mathcal{G}_{n,p,q}$ satisfy (w.h.p.) the approximate regularity and spectral gap conditions required by our first moment analysis (i.e. Theorem 3.1) whenever $b = o(a)$ and $a = \Omega(\log n)$. Versions of the stochastic block model in which the random graph is regular have also been considered [MNS14, BDG+15]. In particular Brito et al. [BDG+15] show that strong reconstruction is possible in polynomial-time when $a - b > 2\sqrt{a + b - 1}$. As for these regular random graphs, we remark that our opportunistic protocol for sparse cut works whenever $a/b \geqslant \log^2 n$ (see also [Bor15, BDG+15]), while our protocol for dense cuts works for any parameter a and b such that $a - b > 2(1+\rho)\sqrt{a+b}$, where ρ is any positive constant. Since it is (information-theoretically) impossible to reconstruct the graph when $a - b \leqslant \mathcal{O}(\sqrt{a+b})$ [MNS14], our result comes within a constant factor of this threshold.

1.3 Roadmap of the paper

After presenting some preliminaries in Section 2, the first moment analysis for almost-regular graphs is given in Section 3. The analysis of the variance of the averaging process in regular graphs for the case of sparse cuts and the analysis of the resulting sign-based protocol are described in Section 4. In Section 5, we address the case of dense cuts: Similarly to the previous section, we first give a second moment analysis and then show how to apply it to devise a suitable opportunistic protocol for this regime.

Due to the considerable length of this paper, most of the technical results are given in a separate appendix.

2 Preliminaries

We study the weighted version of the Averaging process described in the introduction. In each round, one edge of the graph is sampled uniformly at random and the two endpoints of the sampled edge execute the following algorithm.

> AVERAGING(δ) (for a node u that is one of the two endpoints of an active edge)
>
> **Initialization:** If it is the first time u is active, then pick $\mathbf{x}_u \in \{-1, +1\}$ u.a.r.
>
> **Update:** Send $\mathbf{x}_u$ to the other endpoint of the active edge
> and then update $\mathbf{x}_u := (1 - \delta)\mathbf{x}_u + \delta r$, where r is the
> value received from the other endpoint.

Algorithm 1: Updating rule for a node u of an active edge, where $\delta \in (0, 1)$ is the parameter measuring the weight given to the neighbor's value

For a graph G with n nodes and adjacency matrix A, let $0 = \lambda_1 \leqslant \cdots \leqslant \lambda_n$ be the eigenvalues of the normalized Laplacian $\mathcal{L} = I - D^{-1/2} A D^{-1/2}$, where D is the diagonal matrix with the degrees of the nodes. We consider the following classes of graphs.

Definition 2.1 (Almost-regular graphs)**.** *An (n, d, β)-almost-regular graph*[10] $G = (V, E)$ *is a connected, non-bipartite graph over vertex set V, such that every node has degree $d \pm \beta d$.*

Definition 2.2 (Clustered regular graphs)**.** *Let $n \geqslant 2$ be an even integer and d and b two positive integers such that $2b < d < n$. An (n, d, b)-*clustered regular graph *$G = ((V_1, V_2), E)$ is a graph over node set $V = V_1 \cup V_2$, with $|V_1| = |V_2| = n/2$ and such that: (i) Every node has degree d and (ii) Every node in V_1 has b neighbors in V_2 and every node in V_2 has b neighbors in V_1.*

We remark that if a graph is clustered regular then we easily get that the indicator vector $\boldsymbol{\chi}$ of the cut (V_1, V_2) is an eigenvector of $\mathcal{L}$ with eigenvalue $\frac{2b}{d}$; If we further assume that $\lambda_3 > \frac{2b}{d}$, then $\boldsymbol{\chi}$ is an eigenvector of λ_2. We next recall the notion of *weak reconstruction* [BCN+17].

Definition 2.3 (Weak Reconstruction)**.** *A function $f : V \to \{\pm 1\}$ is said to be an ε-*weak reconstruction *of G if subsets $W_1 \subseteq V_1$ and $W_2 \subseteq V_2$ exist, each of size at least $(1-\varepsilon)n/2$, such that $f(W_1) \cap f(W_2) = \emptyset$.*

In this paper, we introduce a weaker notion of distributed community detection. Namely, let $\Delta(\mathbf{x}, \mathbf{y})$ denote the *Hamming distance* between two binary strings $\mathbf{x}$ and $\mathbf{y}$.

Definition 2.4 (Community-sensitive labeling)**.** *Let $G = (V, E)$ be a graph, let (V_1, V_2) be a partition of V and let $\gamma \in (0, 1]$. For some $m \in \mathbb{N}$, a function $\mathbf{h} : V_1 \cup V_2 \to \{0, 1\}^m$ is a γ-*community-sensitive labeling *for (V_1, V_2) if a subset $\tilde{V} \subseteq V$ with size $|\tilde{V}| \geqslant (1-\gamma)|V|$ and two constants $0 \leqslant c_1 < c_2 \leqslant 1$ exist, such that for all $u, v \in \tilde{V}$ it holds that*

$$\Delta(\mathbf{h}_u, \mathbf{h}_v) \begin{cases} \leqslant c_1 m & \text{if } i_u = i_v \quad \text{(Case (i))}, \\ \geqslant c_2 m & \text{otherwise} \quad \text{(Case (ii))}, \end{cases}$$

where $i_u = 1$ if $u \in V_1$ and $i_u = 2$ if $u \in V_2$.

3 First Moment Analysis

In this section, we analyze the expected behaviour of Algorithm Averaging(1/2) on an almost-regular graph G (see Definition 2.1). The evolution of the resulting process can be formally described by the recursion $\mathbf{x}^{(t+1)} = W_t \cdot \mathbf{x}^{(t)}$, where $W_t = (W_t(i,j))$ is the random matrix that defines the updates of the values at round t, i.e.,

$$W_t(i,j) = \begin{cases} 0 & \text{if } i \neq j \text{ and } \{i,j\} \text{ is not sampled (at round } t), \\ 1/2 & \begin{array}{l}\text{if } i = j \text{ and an edge with endpoint } i \text{ is sampled} \\ \text{or } i \neq j \text{ and edge } \{i,j\} \text{ is sampled,}\end{array} \\ 1 & \text{if } i = j \text{ and } i \text{ is not an endpoint of sampled edge.} \end{cases} \tag{1}$$

and the initial random vector $\mathbf{x}^{(0)}$ is uniformly distributed in $\{-1, 1\}^n$.[11]

Notice that random matrices $\{W_t : t \geqslant 0\}$ are independent and identically distributed and simple calculus shows that their expectation can be expressed as (see Observation A.2 in the Appendix):

$$\overline{W} := \mathbb{E}[W_t] = I - \frac{1}{2m} L\,, \tag{2}$$

[10] This class is more general than the one introduced in [BCN+17], since there is no regularity constraint on the outer node degree, i.e., on the number of edges a node can have towards the other community.

[11] Notice that, since each node chooses value ± 1 with probability 1/2 the first time it is active, by using the principle of deferred decisions we can assume there exists an "initial" random vector $\mathbf{x}^{(0)}$ uniformly distributed in $\{-1, +1\}^n$.

where $L = D - A$ is the Laplacian matrix of G. Matrix $\overline{W}$ is thus symmetric and doubly-stochastic. We denote its eigenvalues as $\bar{\lambda}_1, \ldots, \bar{\lambda}_n$, with $1 = \bar{\lambda}_1 \geqslant \bar{\lambda}_2 \geqslant \cdots \bar{\lambda}_n \geqslant -1$.

We next provide a first moment analysis for (n, d, β)-almost regular graphs that exhibit a clustered structure. Our analysis proves the following results.

Theorem 3.1. *Let $G = (V, E)$ be an (n, d, β)-almost regular graph $G = (V, E)$ with a balanced partition $V = (V_1, V_2)$ and such that: (i) The cut $E(V_1, V_2)$ is sparse, i.e., $m_{1,2} = |E(V_1, V_2)| = o(m)$; (ii) The gap $\lambda_3 - \lambda_2 = \Omega(1)$.*[12] *If nodes of G execute Protocol* AVERAGING *then, with constant probability w.r.t. the initial random vector $\mathbf{x}^{(0)} \in \{-1, 1\}^n$, after $\Theta(n \log n)$ rounds the following holds for all but $o(n)$ nodes: (i) The expected value of a node u increases or decreases depending on the community it belongs to, i.e., $\mathbf{sgn}\left(\mathbb{E}\left[\mathbf{x}_u^{(t-1)} \mid \mathbf{x}^{(0)}\right] - \mathbb{E}\left[\mathbf{x}_u^{(t)} \mid \mathbf{x}^{(0)}\right]\right) = \mathbf{sgn}\left(\boldsymbol{\chi}_u\right)$; (ii) Over a time window of length $\Omega(n \log n)$ the sign of the expected value of a node u reflects the community u belongs to, i.e., $\mathbf{sgn}\left(\mathbb{E}\left[\mathbf{x}_u^{(t)} \mid \mathbf{x}^{(0)}\right]\right) = \mathbf{sgn}\left(\alpha_2 \boldsymbol{\chi}_u\right)$, for some $\alpha_2 = \alpha_2(\mathbf{x}^{(0)})$.*

Proof of Theorem 3.1: An outline

The proof makes a black-box use of technical results that are rigorously given in Appendix B.1. We believe, these results are interesting in their own right, since they shed light on the evolution of the dynamics and its algebraic structure.

The hypotheses of Theorem 3.1 involve the eigenvalues of the normalized Laplacian matrix $\mathcal{L}$ of the graph, while the expected evolution of the process is governed by matrix $\overline{W}$ and its eigenvalues $1 = \bar{\lambda}_1 \geqslant \cdots \bar{\lambda}_n \geqslant -1$ (see Lemma B.1). However, (n, d, β)-almost regularity implies that the spectra of these two matrices are related. In particular, it is easy to see that, under the hypotheses of Theorem 3.1, we have (see Observation A.5 in the Appendix)

$$\frac{d}{2m}(1 - 2\gamma)\left(\lambda_3 - \lambda_2\right) \leqslant \bar{\lambda}_2 - \bar{\lambda}_3 \leqslant \frac{d}{2m}(1 + 2\gamma)\left(\lambda_3 - \lambda_2\right) \tag{3}$$

In Lemma B.1, we decompose $\mathbb{E}\left[\mathbf{x}^{(t)}\right]$ into its components along the first two eigenvectors of $\overline{W}$ and into the corresponding orthogonal component $\mathbf{e}^{(t)}$ (note that $\overline{W}$ admits an orthonormal eigenvector basis since it is symmetric). We further decompose the component along the second eigenvector of $\overline{W}$ into its component parallel to the partition indicator vector ($\alpha_2 \bar{\lambda}_2^t \boldsymbol{\chi}$) and into the corresponding orthogonal one ($\alpha_2 \sqrt{n}\, \bar{\lambda}_2^t \mathbf{f}_\perp$). As a consequence, we can rewrite $\mathbb{E}\left[\mathbf{x}^{(t)}\right]$ as

$$\mathbb{E}\left[\mathbf{x}_u^{(t)}\right] = \alpha_1 + \alpha_2 \bar{\lambda}_2^t \left[\boldsymbol{\chi}_u + \sqrt{n}\, \mathbf{f}_{\perp,u}\right] + \mathbf{e}_u^{(t)}, \tag{4}$$

where $\left\|\mathbf{e}^{(t)}\right\| \leqslant \bar{\lambda}_3^t \sqrt{n}$. Hence, if $\alpha_2 \neq 0$ and $\bar{\lambda}_3 < \bar{\lambda}_2$, the term $\mathbf{e}_u^{(t)}$ becomes negligible w.r.t. the other two from some round t onward. Moreover, for any node u with $\mathbf{f}_{\perp,u} < 1/\sqrt{n}$, $\mathbf{sgn}([\boldsymbol{\chi}_u + \sqrt{n}\, \mathbf{f}_{\perp,u}]) = \mathbf{sgn}(\boldsymbol{\chi}_u)$, i.e., $\mathbf{x}_u^{(t)}$ *identifies* the community V_h node u belongs to. Accordingly, we say a node $u \in [n]$ is *ε-bad* if it does not satisfy the above property (see Definition B.1).[13] Next, we derive an upper bound on the number of ε-bad nodes. To this purpose, we first prove an upper bound on the square norm of $\mathbf{f}_\perp$, as a function of the gap $\bar{\lambda}_2 - \bar{\lambda}_3$ and the ratio between the size of the cut $m_{1,2}$ and the total number of edges in the graph (Lemma B.2). This easily implies an upper bound on the number of ε-bad nodes (Corollary B.1) as a function of the gap $\bar{\lambda}_2 - \bar{\lambda}_3$. From (3) and the hypothesis $\lambda_3 - \lambda_2 = \Omega(1)$ the upper bound in Corollary B.1 turns out to be $\mathcal{O}(m_{1,2}/d)$, which in turn is $o(n)$ under the hypothesis $m_{1,2} = o(m)$.

These results and (4) imply the following conclusions for all ε-good nodes u:

[12] In practice, this means that each of the subgraphs induced by community V_i ($i = 1, 2$) is an expander.

[13] Consistently, a node is *ε-good* otherwise.

(i) For all $t = \Omega(n \log n)$, the evolution of $\mathbb{E}\left[\mathbf{x}_u^{(t)}\right]$ along two consecutive rounds identifies the block u belongs to (Lemma B.3), namely:

$$\mathbf{sgn}\left(\mathbb{E}\left[\mathbf{x}_u^{(t-1)}\right] - \mathbb{E}\left[\mathbf{x}_u^{(t)}\right]\right) = \mathbf{sgn}\left(\boldsymbol{\chi}_u\right)$$

(ii) If $|\alpha_2|$ is sufficiently larger than $|\alpha_1|$ and the second and third largest eigenvalues of $\overline{W}$ satisfy appropriate conditions, for all t falling in a suitable time window, the sign of $\mathbb{E}\left[\mathbf{x}_u^{(t)}\right]$ identifies the community node u belongs to (Lemma B.4), namely:

$$\mathbf{sgn}\left(\mathbb{E}\left[\mathbf{x}_u^{(t)}\right]\right) = \mathbf{sgn}(\alpha_2 \boldsymbol{\chi}_u)$$

Moreover, Lemma B.5 implies that the initial random vector $\mathbf{x}$ satisfies the hypotheses of Lemma B.3 w.h.p. (i.e., $\alpha_2(\mathbf{x}) \neq 0$ w.h.p.) and those of Lemma B.4 with constant probability (i.e., $|\alpha_2(\mathbf{x})| \geqslant 2|\alpha_1(\mathbf{x})|/(1-\varepsilon)$ with constant probability).

As a result, we can claim the following for any non-bad node u: if we consider the r.v. $\mathbf{h}_u^{\text{jump},(t)} = \mathbf{sgn}\left(\mathbb{E}\left[\mathbf{x}_u^{(t-1)} - \mathbf{x}_u^{(t)} \mid \mathbf{x}^{(0)}\right]\right)$, Lemma B.3 implies $\mathbf{h}_u^{\text{jump},(t)} = \mathbf{sgn}(\alpha_2 \boldsymbol{\chi}_u)$ w.h.p., for every t such that

$$t \geqslant 3 \log\left(\frac{n}{1-\varepsilon}\right) / \log(\bar{\lambda}_2/\bar{\lambda}_3). \tag{5}$$

Likewise, if we consider the r.v. $\mathbf{h}_u^{\text{sign},(t)} = \mathbf{sgn}\left(\mathbb{E}\left[\mathbf{x}_u^{(t)}\right]\right)$, Lemma B.4 implies

$$\mathbb{P}\left[\mathbf{h}^{\text{jump},(t)} = \mathbf{sgn}(\alpha_2 \boldsymbol{\chi}_u)\right] = \Omega(1),$$

for all t such that

$$\frac{1}{\log(1/\bar{\lambda}_3)} \log(n/|\alpha_1|) \leqslant t \leqslant \frac{1}{\log(1/\bar{\lambda}_2)} \log\left(\frac{|\alpha_2|(1-\varepsilon)}{2|\alpha_1|}\right). \tag{6}$$

Finally, note that under the hypothesis $\lambda_3 - \lambda_2 = \Omega(1)$, the lower bounds on t in (5) and (6) are both $\mathcal{O}(n \log n)$.

4 Regular Graphs with a Sparse Cut

We next provide a second moment analysis of the Averaging(δ) with $\delta = 1/2$ on the class of (n,d,b)-*clustered regular graphs* (see Definition 2.2) when the cut between the two communities is relatively sparse, i.e., for $\lambda_2 = 2b/d = o(\lambda_3/\log n)$. This analysis is consistent with the "expected" clustering behaviour of the dynamics explored in the previous section and highlights clustering properties that emerge well before global mixing time, as we show in Section 4.2. In particular, the main analysis results are discussed in Section 4.1, while in Section 4.2, we describe how the above analysis in concentration can be exploited to get an opportunistic protocol for provably-good community-sensitive labeling.

4.1 Second moment analysis for sparse cuts

Restriction to (n,d,b)-*clustered regular graphs* simplifies the analysis of the Averaging dynamics. When G is regular, $\overline{W}$ defined in (2) can be written as

$$\overline{W} = \left(1 - \frac{1}{n}\right) I + \frac{1}{n} P = I - \frac{1}{n} \mathcal{L}$$

This obviously implies that $\overline{W}$ and $\mathcal{L}$ share the same eigenvectors, while every eigenvalue λ_i of $\mathcal{L}$ corresponds to an eigenvalue $\bar{\lambda}_i = 1 - \lambda_i/n$ of $\overline{W}$. For (n, d, b)-*clustered regular graphs*, these facts and our preliminary remarks in Section 2 further imply $\bar{\lambda}_2 = 1 - \lambda_2/n = 1 - 2b/dn$ whenever $\lambda_3 > \frac{2b}{d}$ while, very importantly, the partition indicator vector $\boldsymbol{\chi}$ turns out to be the eigenvector of $\overline{W}$ corresponding to $\bar{\lambda}_2$ (see (2)). As a consequence, the orthogonal component $\mathbf{f}_\perp$ in (4) is $\mathbf{0}$ in this case.

On the other hand, even in this restricted setting, our second moment analysis requires new, non-standard concentration results for the product of random matrices that apply to far more general settings and may be of independent interest.

For the sake of readability, we here denote by $\mathbf{y}^{(t)} = Q_2\mathbf{x}^{(t)}$ the component of the state vector in the eigenspace of the second eigenvalue of $\overline{W}$, while $\mathbf{z}^{(t)} = Q_{3\cdots n}\mathbf{x}^{(t)}$ denotes $\mathbf{x}^{(t)}$'s projection onto the subspace orthogonal to $\mathbf{1}$ and $\boldsymbol{\chi}$. If we also set $\mathbf{x}_\| = Q_1\mathbf{x}^{(0)}$, we can write:

$$\mathbf{x}^{(t)} = \mathbf{x}_\| + \mathbf{y}^{(t)} + \mathbf{z}^{(t)}. \tag{7}$$

Notice that, by taking expectations in the equation above, we get (4) with $\mathbb{E}\left[\mathbf{y}^{(t)}\right] = \alpha_2\bar{\lambda}_2^t\boldsymbol{\chi}$ and $\mathbb{E}\left[\mathbf{z}^{(t)}\right] = \mathbf{e}^{(t)}$.

Our analysis of the process induced by Averaging(1/2) provides the following bound.

Theorem 4.1 (Second moment analysis)**.** *Let G be an (n, d, b)-clustered regular graph with $\lambda_2 = \frac{2b}{d} = o(\lambda_3/\log n)$. Then, for every $\frac{3n}{\lambda_3}\log n \leqslant t \leqslant \frac{n}{4\lambda_2}$ it holds that*

$$\mathbb{E}\left[\left\|\mathbf{y}^{(t)} + \mathbf{z}^{(t)} - \mathbf{y}^{(0)}\right\|^2\right] \leqslant \frac{3\lambda_2 t}{n}.$$

We prove Theorem 4.1 by bounding and tracking the lengths of the projections of $\mathbf{x}^{(t)}$ onto the eigenspace of λ_2 and onto the space orthogonal to $\mathbf{1}$ and $\boldsymbol{\chi}$, i.e. $\|\mathbf{y}^{(t)}\|^2$ and $\|\mathbf{z}^{(t)}\|^2$. Due to lack of space, the proof is deferred to Appendix C.1.

We here want just to remark that the only part using the regularity of the graph is the derivation of the upper bound on $\mathbb{E}\left[\|\mathbf{y}^{(t+1)}\|^2\right]$ (see Lemma C.2), in particular its second addend. This term arises from an expression involving the Laplacian of G, which is far from simple in general, but that very nicely simplifies in the regular case. We suspect that increasingly weaker bounds should be achievable as the graph deviates from regularity.

Theorem 4.1 gives an upper bound on the squared norm of the difference of the state vector at step t with the state vector at step 0. Corollary 4.1 below shows how such a *global* bound can be used to derive *pointwise* bounds on the values of the nodes.

Definition 4.1. *A node v is* ε-good *at time t if*

$$(\mathbf{x}_v^{(t)} - (\mathbf{x}_{\|,v} + \mathbf{y}_v^{(0)}))^2 \leqslant \frac{\varepsilon^2}{n}\|\mathbf{y}^{(0)}\|^2,$$

it is ε-bad *otherwise. We define by B_t the set of nodes that are ε-bad at time t: $B_t = \{u :$ u is ε-bad at time $t\}$.*

Observe first that, by definition of ε-bad node and some counting argument, we can prove both the next inequality and the corollary below (see Appendix C.2 for their proofs)

$$|B_t| \leqslant \frac{n}{\varepsilon^2\|\mathbf{y}^{(0)}\|^2}\|\mathbf{y}^{(t)} + \mathbf{z}^{(t)} - \mathbf{y}^{(0)}\|^2. \tag{8}$$

Corollary 4.1. *Assume $3\frac{n}{\lambda_3}\log n \leqslant t \leqslant 3c\frac{n}{\lambda_3}\log n$ for any absolute constant $c \geqslant 1$ and $\lambda_2/\lambda_3 \leqslant \varepsilon^4/(4c\log n)$:*

$$\mathbb{P}\left[|B_t| > \varepsilon n \,|\, \mathbf{x}^{(0)} = \mathbf{x}\right] \leqslant \varepsilon. \tag{9}$$

The next lemma gives a bound on the number of nodes that are good over a relatively large time-window. This is the key-property that we will use to analyse the asynchronous protocol Sign-Labeling (see the next subsection and Lemma C.7).

Lemma 4.2 (Non-ephemeral good nodes). *Let $\varepsilon > 0$ be an arbitrarily small value, let G be an (n, d, b)-clustered regular graph with $\frac{\lambda_2}{\lambda_3} \leqslant \frac{\lambda_3 \varepsilon^4}{c \log^2 n}$, for a large enough costant c. If we execute* Averaging(1/2) *on G, it holds that*

$$\mathbb{P}\left[|B_t| \leqslant 3\varepsilon \cdot n\,,\, \forall t \,:\, 6\frac{n}{\lambda_3}\log n \leqslant t \leqslant 12\frac{n}{\lambda_3}\log n\right] \geqslant 1 - \varepsilon\,.$$

4.1.1 Proof of Lemma 4.2: An overview

The main idea of the proof is to first show that with probability strictly larger than $1-\varepsilon$, the number of ε-good nodes is at least $n \cdot (1 - \varepsilon/\log n)$ in every round $t \in [t_1, 2t_1]$. Theorem 4.1 already ensures this to be true in any given time step within a suitable window, but simply taking a union bound will not work, since we have $n \log n$ time steps and only a $1-\varepsilon$ probability of observing the desired outcome in each of them. We will instead argue about the possible magnitude of the change in $\|\mathbf{y}^{(t)} + \mathbf{z}^{(t)} - \mathbf{y}^{(0)}\|^2$ over time, assuming this quantity is small at time $6\frac{n}{\lambda_3}\log n$. We will then show that our argument implies that, with probability $1-\varepsilon$, at least $n - \varepsilon n$ nodes remain ε-good over the entire window $[6\frac{n}{\lambda_3}\log n, 12\frac{n}{\lambda_3}\log n]$.

The full proof of Lemma 4.2 is given in Appendix C.3.

4.2 The Sign-Labeling protocol

Leveraging the results of Subsection 4.1, we next propose a simple, lightweight opportunistic protocol that provides community-sensitive labeling for graphs that exhibit a relatively sparse cut.

The algorithm, denoted as Sign-Labeling, adds a simple *labeling rule* to the Averaging(1/2) process: Each node keeps track of the number of times it is activated. Upon its T-th activation, for a suitable $T = \Theta(\log n)$, the node uses the sign of its current value as a binary label. The above local strategy is applied to ℓ independent runs of Averaging(1/2), so that every node is eventually assigned a binary signature of length ℓ.

Sign-Labeling(T,ℓ) (for a node u that is one of the two endpoints of an active edge) **Component selection:** Jointly[14] with the other endpoint choose a component $j \in [\ell]$ u.a.r. **Initialization and update:** Run one step of Averaging (1/2) for component j. **Labeling:** If this is the T-th activation of component j: set $\mathbf{h}_u^{sign}(j) = \mathbf{sgn}(\mathbf{x}_u(j))$.

Algorithm 2: Sign-Labeling algorithm for a node u of an active edge.

Algorithm Sign-Labeling achieves community-sensitive labeling (see Definition 2.4), as stated in the following theorem and corollary.

Theorem 4.3 (Community-sensitive labeling). *Let $\varepsilon > 0$ be an arbitrarily small value, let G be an (n, d, b)-clustered regular graph with $\frac{\lambda_2}{\lambda_3} \leqslant \frac{\lambda_3 \varepsilon^4}{c \log^2 n}$, for a large enough constant c. Then, proto-*

col SIGN-LABELING (T, ℓ) *with* $T = (8/\lambda_3) \log n$ *and* $\ell = 10\varepsilon^{-1} \log n$ *performs a* γ*-community-sensitive labeling of* G *according to Definition 2.4 with* $c_1 = 4\varepsilon$*,* $c_2 = 1/6$ *and* $\gamma = 6\varepsilon$*, w.h.p. The convergence time is* $\mathcal{O}(n\ell \log n/\lambda_3)$ *and the work per node is* $\mathcal{O}(\ell \log n/\lambda_3)$*, w.h.p.*

Notice that, according to the hypothesis of Theorem 4.3, in order to set local parameters T and ℓ, nodes should know parameters ε and λ_3 (in addition to a polynomial upper bound on the number of the nodes). However, it easy to restate it in a slightly restricted form that does not require such assumptions on what nodes know about the underlying graph.

Corollary 4.2. *Protocol* SIGN-LABELING $(80 \log n, 600 \log n)$ *performs a* $(1/10)$*-community-sensitive labeling, according to Definition 2.4 with* $c_1 = 1/15$ *and* $c_2 = 1/6$*, of any* (n, d, b)*-clustered regular graph* G *with* $\lambda_3 \geqslant 1/10$ *and* $\lambda_2 \leqslant 1/(c \log^2 n)$ *for a large enough constant* c*.*

4.2.1 Proof of Theorem 4.3: An Overview

We here sketch the main arguments proving Theorem 4.3: Its full proof is deferred to Appendix C.4.

Lemma 4.2 essentially states that over a suitable time window of size $\Theta(n \log n)$, for all nodes u but a fraction $\mathcal{O}(\varepsilon/\log n)$, we have $\mathbf{sgn}(\mathbf{x}_u^{(t)}) = \mathbf{sgn}(\mathbf{x}_{\|,u} + \mathbf{y}_u^{(0)})))$. Recalling that $\mathbf{x}_\|$ and $\mathbf{y}^{(0)}$ respectively are $\mathbf{x}^{(0)}$'s projections along $\boldsymbol{\chi}/\sqrt{n}$ and $\mathbf{1}/\sqrt{n}$, this immediately implies that, with probability $1-\varepsilon$ and up to a fraction ε of the nodes, $\mathbf{sgn}(\mathbf{x}_u^{(t)}) = \mathbf{sgn}(\mathbf{x}_v^{(t)})$, whenever u and v belong to the same community and t falls within the aforementioned window. As to the latter condition, we prove that each node labels itself within the right window with probability at least $1 - 1/n$.[15] Moreover, $\mathbf{sgn}(\mathbf{x}_{\|,u} + \mathbf{y}_u^{(0)}))) = \mathbf{sgn}(\boldsymbol{\chi}_u)$, whenever $\mathbf{y}_u^{(0)}$ exceeds $\mathbf{x}_{\|,u}$ in modulus, which occurs with probability $1/2 - o(1)$ from the (independent) Rademacher initialization. As a consequence, if we run ℓ suitably independent copies of the process (see Algorithm 2), the following will happen for all but a fraction $\mathcal{O}(\varepsilon)$ of the nodes: the signatures of two nodes belonging to the same community will agree on $\ell - o(1)$ bits, whereas those of two nodes belonging to different communities will disagree on $\Omega(\ell)$ bits, i.e., our algorithm returns a community-sensitive labeling of the graph.

5 Regular Graphs with a Dense Cut

In this section, we extend our study to the lazy averaging algorithm AVERAGING(δ) where $\delta < 1/2$. Similar to the previous section, we assume that the underlying graph G is an (n, d, b)-clustered regular graph and $\lambda_3 > \lambda_2 = 2b/d$. However, this new analyses and the clustering protocol we derive from will work even for large (constant) λ_2, in contrast to those in Section 4 which only works for small $\lambda_2 \ll 1/\log^2 n$. The structure of this section is similar to the previous one. Indeed, in Subsection 5.1, we propose a second moment analysis of the AVERAGING(δ) for the above-mentioned regime of λ_2. Then, in Subsection 5.2, we exploit the analysis above to devise a protocol that guarantees a weak reconstruction for the underlying graph with arbitrarily-large constant probability and thus, by running independent "copies" of the protocol (so, similarly to the previous section), we easily obtain a community-sensitive labeling of the graph, with high probability.

[15] It may be worth noting that $\mathbf{sgn}(\mathbf{x}_u^{(t)}) = \mathbf{sgn}(\mathbf{x}_v^{(t)})$ for u and v belonging to the same community does not imply $\mathbf{sgn}(\mathbf{x}_u^{(t)}) \neq \mathbf{sgn}(\mathbf{x}_v^{(t)})$ when they don't.

5.1 Second moment analysis for large λ_2

Informally speaking, we show that, for an appropriate value of δ and any t such that $\Omega(n\log n) \leqslant t \leqslant \mathcal{O}(n^2)$, with large probability, the vector $\mathbf{y}^{(t)} + \mathbf{z}^{(t)}$ is almost parallel to χ, i.e., $\|\mathbf{z}^{(t)}\|$ is much smaller than $\|\mathbf{y}^{(t)}\|$. A more precise statement is given below as Theorem 5.1. Note that, for brevity, we write $\mathcal{E}$ here to denote the sequence $\{(u_t, v_t)\}_{t\in\mathbb{N}}$ of the edges chosen by the protocol.

Theorem 5.1. *For any sufficiently large $n \in \mathbb{N}$, any*[16] $\delta \in (0, 0.8(\lambda_3 - \lambda_2))$ *and any* $t \in \left[\Omega\left(\frac{n}{\delta(\lambda_3-\lambda_2)}\log(n/\delta)\right), \mathcal{O}\left(\frac{n^2}{\delta(\lambda_3-\lambda_2)}\left(\frac{d(\lambda_3-\lambda_2)}{\delta b}\right)^{2/3}\right)\right]$, *we have*

$$\mathop{\mathbb{P}}_{\mathbf{x}^{(0)},\mathcal{E}}\left[\|\mathbf{z}^{(t)}\|^2 \leqslant \sqrt{\frac{\delta b}{d(\lambda_3-\lambda_2)}}\|\mathbf{y}^{(t)}\|^2\right] \geqslant 1 - \mathcal{O}\left(\sqrt[3]{\frac{\delta b}{d(\lambda_3-\lambda_2)}} + \frac{1}{\sqrt{n}}\right).$$

Theorem 5.1 should be compared to Theorem 4.1: both assert that $\|\mathbf{y}^{(t)}\|$ is much larger than $\|\mathbf{z}^{(t)}\|$, but Theorem 5.1 works even when λ_2 is quite large whereas Theorem 4.1 only holds for $\lambda_2 \ll 1/\log n$.

While the parameter dependencies in Theorem 5.1 may look confusing at first, there are mainly two cases that are interesting here. First, for any error parameter ε, we can pick δ depending only on ε and $\lambda_3 - \lambda_2$ in such a way that Theorem 5.1 implies that, with probability $1-\varepsilon$, $\|\mathbf{z}^{(t)}\|^2$ is at most $\varepsilon\|\mathbf{y}^{(t)}\|^2$, as stated below.

Corollary 5.1. *For any constant $\varepsilon > 0$ and for any $\lambda_3 > \lambda_2$, there exists δ depending only on ε and $\lambda_3 - \lambda_2$ such that, for any sufficiently large n and for any $t \in [\Omega_{\varepsilon,\lambda_3-\lambda_2}(n\log n), \mathcal{O}(n^2)]$, we have*

$$\mathop{\mathbb{P}}_{\mathbf{x}^{(0)},\mathcal{E}}\left[\|\mathbf{z}^{(t)}\|^2 \leqslant \varepsilon\|\mathbf{y}^{(t)}\|^2\right] \geqslant 1-\varepsilon.$$

Another interesting case is when $\delta = 1/2$ (i.e., we consider the basic averaging protocol). Recalling that $\lambda_2 = 2b/d$, observe that λ_2 appears in both the bound on $\|\mathbf{z}^{(t)}\|^2$ and the error probability. Hence, we can derive a similar lemma as the one above, but with λ_2 depending on ε instead of δ:

Corollary 5.2. *Fix $\delta = 1/2$. For any constant $\varepsilon > 0$, any*[17] $\lambda_3 > 0.7$, *any sufficiently small λ_2 depending only on ε, any sufficiently large n and any $t \in [\Omega_\varepsilon(n\log n), \mathcal{O}(n^2)]$, we have*

$$\mathop{\mathbb{P}}_{\mathbf{x}^{(0)},\mathcal{E}}\left[\|\mathbf{z}^{(t)}\|^2 \leqslant \varepsilon\|\mathbf{y}^{(t)}\|^2\right] \geqslant 1-\varepsilon.$$

5.1.1 Proof of Theorem 5.1: An Overview

Due to space constraint, the full proof of Theorem 5.1 is deferred to Appendix D. We provide a brief summary of the ideas behind the proof here. Compared to the proof of Theorem 4.1, the main additional technical challenge in the new proof is to show that $\|\mathbf{y}^{(t)}\|$ is large with reasonably high probability. In Theorem 4.1, this is true because λ_2 is so small that $\mathbf{y}^{(t)}$ remains almost unchanged from $\mathbf{y}^{(0)}$. However, in the setting of large λ_2, this is not true anymore; for constant λ_2, even $\mathbb{E}[\mathbf{y}^{(t)}]$ shrinks by a constant factor from $\mathbf{y}^{(0)}$ when $t \geqslant \Omega_{\lambda_2}(n)$.

[16] Here 0.8 is arbitrary and can be changed to any constant less than 1. However, we pick an absolute constant here to avoid introducing another parameter to our theorem.

[17] 0.7 here can be replaced by any constant larger than 0.5.

As a result, we need to develop a more fine-grained understanding of how $\|\mathbf{y}^{(t)}\|$, $\|\mathbf{z}^{(t)}\|$ changes over time. Specifically, at the heart of our analysis lies the following lemma[18] which allows us to understand how $\|\mathbf{y}^{(t)}\|$, $\|\mathbf{z}^{(t)}\|$ behave, given $\|\mathbf{y}^{(t-1)}\|$, $\|\mathbf{z}^{(t-1)}\|$:

Lemma 5.2. *For any $t \in \mathbb{N}$,*

$$\mathbb{E}[\|\mathbf{y}^{(t)}\|^2] \leqslant \left(1 - \frac{4\delta\lambda_2}{n} + \frac{8\delta^2\lambda_2}{n^2}\right)\|\mathbf{y}^{(t-1)}\|^2 + \left(\frac{8\delta^2\lambda_2}{n^2}\right)\|\mathbf{z}^{(t-1)}\|^2$$

and

$$\mathbb{E}[\|\mathbf{z}^{(t)}\|^2] \leqslant \left(\frac{4\delta^2\lambda_2}{n}\right)\|\mathbf{y}^{(t-1)}\|^2 + \left(1 - \frac{4\delta(1-\delta)\lambda_3}{n}\right)\|\mathbf{z}^{(t-1)}\|^2$$

where the expectation is over the random edge selected at time t.

For simplicity of the overview, let us pretend that the cross terms were not there, i.e., that $\mathbb{E}[\|\mathbf{y}^{(t)}\|^2] \leqslant \left(1 - \frac{4\delta\lambda_2}{n} + \frac{8\delta^2\lambda_2}{n^2}\right)\|\mathbf{y}^{(t-1)}\|^2$ and $\mathbb{E}[\|\mathbf{z}^{(t)}\|^2] \leqslant \left(1 - \frac{4\delta(1-\delta)\lambda_3}{n}\right)\|\mathbf{z}^{(t-1)}\|^2$. These imply that

$$\mathbb{E}[\|\mathbf{y}^{(t)}\|^2] \leqslant \left(1 - \frac{4\delta\lambda_2}{n} + \frac{8\delta^2\lambda_2}{n^2}\right)^t \|\mathbf{y}^{(0)}\|^2 \tag{10}$$

and

$$\mathbb{E}[\|\mathbf{z}^{(t)}\|^2] \leqslant \left(1 - \frac{4\delta(1-\delta)\lambda_3}{n}\right)^t \|\mathbf{z}^{(0)}\|^2. \tag{11}$$

Now, by Markov's inequality, (11) implies that, with 0.99 probability, $\|\mathbf{z}^{(t)}\|$ is at most $\mathcal{O}\left(\left(1 - \frac{2\delta(1-\delta)\lambda_3}{n}\right)^t \|\mathbf{z}^{(0)}\|\right)$. However, it is not immediately clear how (10) can be used to lower bound $\|\mathbf{y}^{(t)}\|$. Fortunately for us, it is rather simple to see that, for a fixed $\mathbf{y}^{(0)}$, $\mathbb{E}[\mathbf{y}^{(t)}]$ can be computed exactly; in particular,

$$\mathbb{E}[\mathbf{y}^{(t)}] = \left(1 - \frac{2\delta\lambda_2}{n}\right)^t \mathbf{y}^{(0)}. \tag{12}$$

Let $a_y(t) \in \mathbb{R}$ be such that $\mathbf{y}^{(t)} = a_y(t) \cdot (\chi/\sqrt{n})$. (12) can equivalently be stated as $\mathbb{E}[a_y(t)] = (1 - 2\delta\lambda_2/n)^t a_y(0)$. This, together with (10), can be used to bound the variance of $a_y(t)$ as follows:

$$\begin{aligned}\mathrm{Var}(a_y(t)) &\leqslant \left(1 - \frac{4\delta\lambda_2}{n} + \frac{8\delta^2\lambda_2}{n^2}\right)^t a_y(0)^2 - (1 - 2\delta\lambda_2/n)^{2t} a_y(0)^2 \\ &= \mathcal{O}_{\delta,\lambda_2}(t/n^2)\left(\mathbb{E}[a_y(t)]\right)^2.\end{aligned}$$

Hence, when $t \ll n^2$, Chebyshev's inequality implies that $a_y(t)$ concentrates around $\mathbb{E}[a_y(t)]$ or, equivalently, $\|\mathbf{y}^{(t)}\|$ concentrates around $\left(1 - \frac{2\delta\lambda_2}{n}\right)^t \|\mathbf{y}_0\|$.

Finally, observe that, since $\lambda_2 < \lambda_3$, for sufficiently small δ, we have $2\delta\lambda_2 < 2\delta(1-\delta)\lambda_3$. Hence, when $t \gg n\log n$, $\left(1 - \frac{2\delta\lambda_2}{n}\right)^t$ is polynomially (say n^{10} times) larger than $\left(1 - \frac{2\delta(1-\delta)\lambda_3}{n}\right)^t$. It is also not hard to see that, for a random starting vector, $\|\mathbf{z}^{(0)}\| \ll n^{10}\|\mathbf{y}^{(0)}\|$

[18] Lemma 5.2 with its full statement and proof is given in Appendix D as Lemma D.2. Recall that $\lambda_2 = 2b/d$.

with high probability. This means that, for this range of t, we have $\left(1 - \frac{2\delta\lambda_2}{n}\right)^t \|\mathbf{y}^{(0)}\| \gg \left(1 - \frac{2\delta(1-\delta)\lambda_3}{n}\right)^t \|\mathbf{z}^{(0)}\|$ with high probability. Since $\|\mathbf{y}^{(t)}\|$ concentrates on the former quantity whereas $\|\mathbf{z}^{(t)}\|$ often does not exceed a constant factor of the latter, we can conclude that $\|\mathbf{y}^{(t)}\|$ is indeed often much larger than $\|\mathbf{z}^{(t)}\|$.

This wraps up our proof overview of Theorem 5.1.

5.2 The Jump-Labeling protocol

Relying on our insights from the previous section, we propose a lightweight protocol named Jump-Labeling, which makes use of the lazy version of the averaging process. Here $\delta \in [0, 1]$ and $\tau^{\mathrm{s}}, \tilde{\tau}^{\mathrm{s}}, \tau^{\mathrm{e}}, \tilde{\tau}^{\mathrm{e}} \in \mathbb{N}$ are parameters that will be chosen later. Intuitively, protocol Jump-Labeling exploits the expected monotonicity in the behaviour of $\mathbf{sgn}(\mathbf{x}_u^{(t)} - \mathbf{x}^{(t-1)})$ highlighted in Section 3. Though this property does not hold for a single realization of the averaging process in general, the results of Section 5 allow us to show that the sign of $\mathbf{x}^{(\tau_u^e)} - \mathbf{x}^{(\tau_u^s)}$ reflects u's community membership for most vertices with probability $1 - o(1)$ (i.e., the algorithm achieves weak reconstruction) when τ_u^s and τ_u^e are randomly chosen within a suitable interval. This is the intuition behind the main result of this section. Due to space constraints, the full proof of Theorem 5.3 below is deferred to Appendix E.

Theorem 5.3. *Let n be any sufficiently large even positive integer. For any $0 < \delta < 0.8(\lambda_3 - \lambda_2)$, there exist $\tau^s, \tilde{\tau}^s, \tau^e, \tilde{\tau}^e \in \mathbb{N}$ such that, with probability $1 - \mathcal{O}\left(\sqrt[8]{\frac{\delta b}{d(\lambda_3-\lambda_2)}} + \sqrt[4]{\frac{1}{\log n}}\right)$, after $\mathcal{O}\left(\frac{n}{\delta(\lambda_3-\lambda_2)} \log(n/\delta) + \frac{nd}{b\delta}\right)$ rounds of Jump-Labeling$(\delta, \tau^s, \tilde{\tau}^s, \tau^e, \tilde{\tau}^e)$, every node labels its cluster and this labelling is a $\left(\sqrt[8]{\frac{\delta b}{d(\lambda_3-\lambda_2)}} + \sqrt[4]{\frac{1}{\log n}}\right)$-weak reconstruction of G. The convergence time of this algorithm is $\Omega_\delta\left(n\left(\log n + \frac{d}{b}\right)\right)$.*

Jump-Labeling$(\delta, \tau^{\mathrm{s}}, \tilde{\tau}^{\mathrm{s}}, \tau^{\mathrm{e}}, \tilde{\tau}^{\mathrm{e}})$ (for a node u that is one of the two endpoints of an active edge)

Initialization: The first time it is activated, u chooses $\tau_u^{\mathrm{s}}, \tau_u^{\mathrm{e}} \in \mathbb{N}$ independently uniformly at random from $[\tau^{\mathrm{s}}, \tilde{\tau}^{\mathrm{s}}]$ and $[\tau^{\mathrm{e}}, \tilde{\tau}^{\mathrm{e}}]$ respectively. Moreover, let $\tau_u = 0$.

Update (and Averaging's initialization): Run one step of Averaging(δ).

Labeling: If $\tau_u = \tau_u^{\mathrm{s}}$, then set $x_u^{\mathrm{s}} = x_u$.
If $\tau_u = \tau_u^{\mathrm{e}}$, then label $\mathbf{h}_u^{jump} = \mathbf{sgn}(x_u^{\mathrm{s}} - x_u)$.

Algorithm 3: Jump-Labeling algorithm for a node u of an active edge. Here, τ_u is a local counter keeping track of the number of times u was an endpoint of an active edge, while x_u is u's current value.

Remark 1. *The nd/b dependency in the running time is necessary; imagine we start with a good state where $\mathbf{x}^{(0)} = \mathbf{z}^{(0)} = 0$. In this case, the values on one side of the partition are all $a_y(0)$ and the values on the other side are $-a_y(0)$. It is simple to see that, after $o(nd/b)$ steps of our protocol, $1 - o(1)$ fraction of the values remain the same. For these nodes, it is impossible them to determine which cluster they are in and, hence, no good reconstruction can be achieved.*

Similarly to our concentration result in Section 5, let us demonstrate the use of Theorem 5.3 to the two interesting cases. First, let us start with the case where $\lambda_3 - \lambda_2$ is constant. Again, in this case, for any error parameter $\varepsilon > 0$, we can pick $\delta = \delta(\varepsilon, \lambda_2 - \lambda_3)$ sufficiently small so that, with probability $1 - \varepsilon$, the protocol achieves ε-weak reconstruction, as stated below.

Corollary 5.3. *For any constant $\varepsilon > 0$ and for any λ_3, λ_2, there exists δ depending only on ε and $\lambda_3 - \lambda_2$ such that, for any sufficiently large n, there exists $\tau^s, \tilde{\tau}^s, \tau^e, \tilde{\tau}^e \in \mathbb{N}$ such that, with probability $1 - \varepsilon$, after $\mathcal{O}_{\varepsilon,\lambda_3-\lambda_2}\left(n \log n + \frac{n}{\lambda_2}\right)$ rounds of* Jump-Labeling*($\delta, \tau^s, \tilde{\tau}^s, \tau^e, \tilde{\tau}^e$), every node labels its cluster and this labelling is a ε-weak reconstruction of G.*

As in Section 5, we can consider the (non-lazy) averaging protocol and view λ_2 instead as a parameter. On this front, we arrive at the following reconstruction guarantee.

Corollary 5.4. *Fix $\delta = 1/2$. For any constant $\varepsilon > 0$, any $\lambda_3 > 0.7$, any sufficiently small λ_2 depending only on ε, any sufficiently large n, there exists $\tau^s, \tilde{\tau}^s, \tau^e, \tilde{\tau}^e \in \mathbb{N}$ such that, with probability $1 - \varepsilon$, after $\mathcal{O}_{\varepsilon}\left(n \log n + \frac{n}{\lambda_2}\right)$ rounds of* Jump-Labeling*($\delta, \tau^s, \tilde{\tau}^s, \tau^e, \tilde{\tau}^e$), the nodes' labelling is a ε-weak reconstruction of G.*

While the weak reconstruction in the above claims is guaranteed only with arbitrarily-large constant probability, we can boost this success probability considering the same approach we used in Subsection 4.2 to get community-sensitive binary strings of size $\ell = \Theta(\log n)$ from the sign-based protocol.

Indeed, we first run $\ell = \Theta_{\varepsilon}(\log n)$ copies of Jump-Labeling where, similarly to Algorithm 2, "running ℓ copies" of Jump-Labeling means that each node keeps ℓ copies of the states of Jump-Labeling and, when an edge $\{u, v\}$ is activated, u and v jointly sample a random $j \in [\ell]$ and run the j-th copy of Jump-Labeling.

In the previous section, we have seen that Lemma 4.2 and the repetition approach above allowed us to get a good community-sensitive labeling, w.h.p. (not a good weak-reconstruction). Interestingly enough, the somewhat stronger concentration results given in this section allow us to "add" a simple *majority* rule on the top of the ℓ components and get a "good" single-bit label, as described below.

When all ℓ components of a node u have been set, node u sets $\mathbf{h}_u^{jump} = Majority_{j\in[\ell]}(\mathbf{h}_u^{jump}(i))$ where $\mathbf{h}_u^{jump}(j)$ is the binary label of u from the j-th copy of the protocol.

Observe that the weak reconstruction guarantee of Jump-Labeling shown earlier implies that the expected number of mislabelings of each copy is at most $2\varepsilon n$, i.e., $\mathbb{E}[\{u \in V \mid |\mathbf{h}_u^{jump}(i) \neq \chi_u|\}] \leqslant 2\varepsilon n$. Now, since the number of mislabelings of each copy is independent, the total number of mislabelings is at most $\mathcal{O}(\varepsilon n \ell)$, w.h.p. However, if the eventual label of u is incorrect, it must contributes to mislabeling across at least $\ell/2$ copies. As a result, there are at most $\mathcal{O}(\varepsilon n)$ mislabelings in the new protocol, w.h.p.

The above approach in fact works for any weak reconstruction protocol (not just Jump-Labeling) and, in our case, it easily gives the following result.

Corollary 5.5. *For any constant $\varepsilon > 0$ and $\lambda_3 > \lambda_2$, there is a protocol that yields an ε-weak reconstruction of G , w.h.p. The convergence time is $\Theta_{\varepsilon,\lambda_3-\lambda_2}\left(n\left(\log^2 n + \frac{\log n}{\lambda_2}\right)\right)$ rounds, while the work per node is $\mathcal{O}_{\varepsilon,\lambda_3-\lambda_2}\left(\log^2 n + \frac{\log n}{\lambda_2}\right)$.*

We finally remark that, for the dense-cut case we focus on in this section (i.e. $\lambda_2 = 2b/d = \Theta(1)$), the fraction of outliers turns out to be a constant we can made arbitrarily small. If we relax the condition to $\lambda_2 = o(1)$, then this fraction can be made $o(1)$, accordingly. This issue will be clarified in the full version of the paper.

5.2.1 Proof of Theorem 5.3: An Overview

We now give an informal overview of our proof, which builds on the concentration results from Section 5. Since our discussion here will involve both local times and global times, let us define the following notation to facilitate the discussion: for each vertex $u \in V$, let $T_u : \mathbb{N} \to \mathbb{N}$ be a function that maps the local time of u to the global time, i.e., $T_u(\tau) \triangleq \min\{t \in \mathbb{N} \mid |\{i \leqslant t \mid u \in \{u_i, v_i\}\}| \geqslant \tau\}$ where $(\{u_i, v_i\})_{i\in\mathbb{N}}$ is the sequence of active edges.

Recall from the previous section that we let $a_y(t) \in \mathbb{R}$ be such that $\mathbf{y}^{(t)} = a_y(t)\cdot(\chi/\sqrt{n})$. Let us also assume without loss of generality that $a_y(0) \geqslant 0$. Observe first that our concentration result implies the following: for any t such that $\Omega(n \log n) \leqslant t \leqslant \mathcal{O}(n^2)$, with large probability, $\chi_u(\mathbf{x}_u^{(t)} - \mathbf{x}_{||,u})$ is roughly $\mathbb{E}_{\mathcal{E}}\, a_y(t)/n$ for most vertices $u \in V$; let us call these vertices *good for time* t. Imagine for a moment that we change the protocol in such a way that each u has access to the global time t and u assigns $\mathbf{h}_u^{jump} = \mathbf{sgn}(\mathbf{x}_u^{(t^{\mathrm{e}})} - \mathbf{x}_u^{(t^{\mathrm{s}})})$ for some $t^{\mathrm{s}}, t^{\mathrm{e}} \in [\Omega(n \log n), \mathcal{O}(n^2)]$ that do not depend on u. If $t^{\mathrm{e}} - t^{\mathrm{s}}$ is large enough, then $\mathbb{E}_{\mathcal{E}}\, a_y(t^{\mathrm{s}}) \gg \mathbb{E}_{\mathcal{E}}\, a_y(t^{\mathrm{e}})$. This means that, if a vertex $u \in V$ is good at both times t^{s} and t^{e}, then we have that $\chi_u(\mathbf{x}_u^{(t^{\mathrm{s}})} - \mathbf{x}_{||,u}) \approx \mathbb{E}_{\mathcal{E}}\, a_y(t^{\mathrm{s}})/n \gg \mathbb{E}_{\mathcal{E}}\, a_y(t^{\mathrm{e}})/n \approx \chi_u(\mathbf{x}_u^{(t^{\mathrm{e}})} - \mathbf{x}_{||,u})$. Note that when $\chi_u \cdot \mathbf{x}_u^{(t^{\mathrm{s}})} > \chi_u \cdot \mathbf{x}_u^{(t^{\mathrm{e}})}$, we have $\mathbf{h}_u^{jump} = \chi_u$. From this and from almost all vertices are good at both times t^{s} and t^{e}, $\mathbf{h}^{jump}$ is indeed a good weak reconstruction for the graph!

The problem of the modified protocol above is of course that, in our settings, each vertex does not know the global time t. Perhaps the simplest approach to imitate the above algorithm in this regime is to fix $\tau^{\mathrm{s}}, \tau^{\mathrm{e}} \in [\Omega(\log n), \mathcal{O}(n)]$ and, for each $u \in V$, proceed as in Jump-Labeling except with $\tau_u^{\mathrm{s}} = \tau^{\mathrm{s}}$ and $\tau_u^{\mathrm{e}} = \tau^{\mathrm{e}}$. In other words, u assigns $\mathbf{h}_u^{jump} = \mathbf{sgn}(\mathbf{x}_u^{(T_u(\tau^{\mathrm{s}}))} - \mathbf{x}_u^{(T_u(\tau^{\mathrm{e}}))})$. The problem about this approach is that, while we know that $\mathbb{E}_{\mathcal{E}}\, T_u(\tau^{\mathrm{s}}) = 0.5n\tau^{\mathrm{s}}$ and $\mathbb{E}_{\mathcal{E}}\, T_u(\tau^{\mathrm{e}}) = 0.5n\tau^{\mathrm{e}}$, the actual values of $T_u(\tau^{\mathrm{s}})$ and $T_u(\tau^{\mathrm{e}})$ differ quite a bit from their means, i.e., on average they will be $\Omega(n\sqrt{\log n})$ of away their mean. Since our concentration result only says that, at each time t, we expect 99% of the vertices to be good, it is unclear how this can rule out the following extreme case: for many $u \in V$, $T_u(\tau^{\mathrm{s}})$ or $T_u(\tau^{\mathrm{e}})$ is a time step at which u is bad. This case results in $\mathbf{h}^{jump}$ not being a good weak reconstruction of V.

The above issue motivates us to arrive at our eventual algorithm, in which τ_u^{s} and τ_u^{e} are not fixed to be the same for every u, but instead each u pick these values randomly from specified intervals $[\tau^{\mathrm{s}}, \tilde{\tau}^{\mathrm{s}}]$ and $[\tau^{\mathrm{e}}, \tilde{\tau}^{\mathrm{e}}]$. To demonstrate why this overcomes the above problem, let us focus on the interval $[\tau^{\mathrm{s}}, \tilde{\tau}^{\mathrm{s}}]$. While $T_u(\tau^{\mathrm{s}})$ and $T_u(\tilde{\tau}^{\mathrm{s}})$ can still differ from their means, the interval $[T_u(\tau^{\mathrm{s}}), T_u(\tilde{\tau}^{\mathrm{s}})]$ still, with large probability, overlaps with most of $[0.5n\tau^{\mathrm{s}}, 0.5n\tilde{\tau}^{\mathrm{s}}]$ if $\tilde{\tau}^{\mathrm{s}} - \tau^{\mathrm{s}}$ is sufficiently large. Now, if $T_u(\tau+1) - T_u(\tau)$ are the same for all $\tau \in [\tau^{\mathrm{s}}, \tilde{\tau}^{\mathrm{s}}]$, then the distribution of $\mathbf{x}_u^{(T_u(\tau^{\mathrm{s}}))}$ is very close to $\mathbf{x}_u^{(t_u^{\mathrm{s}})}$ if we pick t_u^{s} randomly from $[0.5n\tau^{\mathrm{s}}, 0.5n\tilde{\tau}^{\mathrm{s}}]$. From the usual global time step argument, it is easy to see that the latter distribution results in most u being good at time t_u^{s}. Of course, $T_u(\tau+1) - T_u(\tau)$ will not be the same for all $\tau \in [\tau^{\mathrm{s}}, \tilde{\tau}^{\mathrm{s}}]$, but we will be able to argue that, for almost all such τ, $T_u(\tau+1) - T_u(\tau)$ is not too small, which is sufficient for our purpose.

References

[AAER07] Dana Angluin, James Aspnes, David Eisenstat, and Eric Ruppert. The computational power of population protocols. *Distributed Computing*, 20(4):279–304, 2007.

[ABH14] 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.

[BCN+17] Luca Becchetti, Andrea Clementi, Emanuele Natale, Francesco Pasquale, and Luca Trevisan. Find your place: Simple distributed algorithms for community detection. In *Proc. of the 28th Ann. ACM-SIAM Symp. on Discrete Algorithms (SODA'17)*, pages 940–959. SIAM, 2017.

[BDG+15] 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 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 371–390. ACM, 2015.

[BGPS06] Stephen Boyd, Arpita Ghosh, Balaji Prabhakar, and Devavrat Shah. Randomized gossip algorithms. *IEEE/ACM Transactions on Networking*, 14:2508–2530, 2006.

[Bor15] Charles Bordenave. A new proof of Friedman's second eigenvalue Theorem and its extension to random lifts. *arXiv preprint arXiv:1502.04482*, 2015.

[CO10] A. Coja-Oghlan. Graph partitioning via adaptive spectral techniques. *Combinatorics, Probability and Computing*, 19(02):227–284, 2010.

[CPV93] Andrea Crisanti, Giovanni Paladin, and Angelo Vulpiani. *Products of Random Matrices*, volume 104 of *Springer Series in Solid-State Sciences*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1993. DOI: 10.1007/978-3-642-84942-8.

[DF89] M. E. Dyer and A. M. Frieze. The solution of some random NP-hard problems in polynomial expected time. *Journal of Algorithms*, 10(4):451–489, 1989.

[DKMZ11] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84(6):066106, 2011.

[DP09] D. P. Dubhashi and A. Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, 2009.

[HLL83] P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

[Ips15] J. R. Ipsen. Products of Independent Gaussian Random Matrices. *arXiv:1510.06128 [math-ph]*, October 2015. arXiv: 1510.06128.

[JS98] M. Jerrum and G. B. Sorkin. The metropolis algorithm for graph bisection. *Discrete Applied Mathematics*, 82(1), 1998.

[KM04] D. Kempe and F. McSherry. A decentralized algorithm for spectral analysis. In *Proc. of the ACM Symposium on Theory of Computing (STOC)*, pages 561–568. ACM, 2004.

[Mas14] L. Massoulie. Community Detection Thresholds and the Weak Ramanujan Property. In *Proc. of the ACM Symposium on Theory of Computing (STOC)*, pages 694–703. ACM, 2014.

[McS01] F. McSherry. Spectral partitioning of random graphs. In *Proc. of the IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 529–537, 2001.

[MNS13] E. Mossel, J. Neeman, and A. Sly. A proof of the block model threshold conjecture. *arXiv preprint arXiv:1311.4115*, 2013.

[MNS14] E. Mossel, J. Neeman, and A. Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162(3-4):431–461, 2014.

[MNS16] E. Mossel, J. Neeman, and A. Sly. Belief propagation, robust reconstruction and optimal recovery of block models. *The Annals of Applied Probability*, 26(4):2211–2256, 2016.

[SZ16] H. Sun and L. Zanetti. Distributed graph clustering by load balancing. *CoRR*, abs/1607.04984, 2016.

[SZ17] H. Sun and L. Zanetti. Distributed graph clustering and sparsification. *CoRR*, abs/1711.01262, 2017.

[WWA12] M.J. Williams, R.M. Whitaker, and S.M. Allen. Decentralised detection of periodic encounter communities in opportunistic networks. *Ad Hoc Networks*, 10(8):1544–1556, 2012.

Appendix

A Tools from linear algebra

A.1 Projections on the main eigenspaces

Lemma A.1 (Projection on the first two eigenvectors)**.** *For all $\varepsilon \in (0,1)$, for a random $\mathbf{x} \in \{-1,1\}^n$, with probability at least $1-\mathcal{O}(\varepsilon)$ we have,*

$$\mathbb{P}\left[|\mathbf{x}\cdot\mathbf{1} \pm \mathbf{x}\cdot\boldsymbol{\chi}| \geqslant \varepsilon\cdot\sqrt{n}\right] \geqslant 1-\mathcal{O}(\varepsilon) \quad \textit{and}$$
$$\mathbb{P}\left[|\mathbf{x}\cdot\mathbf{1}| \leqslant |\mathbf{x}\cdot\boldsymbol{\chi}| - \varepsilon\cdot\sqrt{n} \mid |\mathbf{x}\cdot\mathbf{1} \pm \mathbf{x}\cdot\boldsymbol{\chi}| \geqslant \varepsilon\cdot\sqrt{n}\right] = \frac{1}{2}.$$

Proof. Note that $\mathbf{x}\cdot(\mathbf{1}+\boldsymbol{\chi}) = 2\mathbf{x}\cdot\mathbf{1}_{V_1}$ and $\mathbf{x}\cdot(\mathbf{1}-\boldsymbol{\chi}) = 2\mathbf{x}\cdot\mathbf{1}_{V_2}$. Using properties of the binomial distribution, it is easy to see that

$$\mathbb{P}\left(\text{“}|\mathbf{x}\cdot\mathbf{1}+\mathbf{x}\cdot\boldsymbol{\chi}| \geqslant \varepsilon\sqrt{n}\text{”} \wedge \text{“}|\mathbf{x}\cdot\mathbf{1}-\mathbf{x}\cdot\boldsymbol{\chi}| \geqslant \varepsilon\sqrt{n}\text{”}\right) \geqslant 1 - \frac{4}{\sqrt{2\pi}}\varepsilon.$$

The above event implies $||\mathbf{x}\cdot\mathbf{1}| - |\mathbf{x}\cdot\boldsymbol{\chi}|| \geqslant \varepsilon\sqrt{n}$. Since $\mathbf{x}\cdot\mathbf{1}$ and $\mathbf{x}\cdot\boldsymbol{\chi}$ are independent sums of Rademacher random variables, they have the same chances of being positive or negative, thus with probability at least $\frac{1}{2}$ we will have $|\mathbf{x}\cdot\mathbf{1}| \leqslant |\mathbf{x}\cdot\boldsymbol{\chi}|$. □

A.2 Properties of the spectrum of the main matrices

We consider here Algorithm Averaging(δ) assuming $\delta = 1/2$ and recall the main notations:

- A is the adjacency matrix of the clustered graph $G((V_1,V_2);E)$, with $|V_h| = n/2$, $m = |E|$ and $m_{1,2} = |E(V_1,V_2)|$ is the number of edges in the cut (V_1,V_2);
- D is the diagonal matrix with the degrees of nodes;
- $L = D - A$ is the Laplacian matrix;
- $\mathcal{L} = D^{-1/2}LD^{-1/2}$is the normalized Laplacian;
- $P = D^{-1}A$ is the transition matrix;
- For each node $i = 1,\ldots,n$, we name $d_i = a_i + b_i$ the degree of node i, where a_i is the number of neighbors in its own block and b_i is the number of neighbors in the other block

The next facts are often used in our analysis.

Observation A.1. *Let $W = (W(i,j)) \sim \mathcal{W}$ be the random matrix of one step of the averaging process, then*

$$w_{i,j} = \begin{cases} 0 & \textit{if } i \neq j \textit{ and } \{i,j\} \textit{ not sampled} \\ 1/2 & \textit{if } i = j \textit{ and some edge incident on } i \textit{ sampled or } i \neq j \textit{ and edge } \{i,j\} \textit{ sampled} \\ 1 & \textit{if } i = j \textit{ and } i \textit{ not incident to a sampled edge.} \end{cases}$$

Observation A.2. *The expectation of W is*

$$\overline{W} := \mathbb{E}[W] = I - \frac{1}{2m}L\,.$$

Proof. 1. If $i \neq j$ then $\mathbb{E}[W(i,j)] = \frac{1}{2m}$

2. If $i = j$ then $\mathbb{E}[W(i,j)] = 1\left(1 - \frac{d_i}{m}\right) + \frac{1}{2}\frac{d_i}{m} = 1 - \frac{d_i}{2m}$

□

As for the spectrum of the main matrices above, defined by the averaging process, we have the following useful properties we can derive from standard spectral algebra.

Observation A.3. *1. Il λ is an egenvalue of L then $1 - \lambda/(2m)$ is an eigenvalue of $\overline{W}$.*

2. Vector $\mathbf{1}$ is an eigenvector of $\overline{W}$.

3. If the underlying graph G is (n,d,b)-regular then $\boldsymbol{\chi}$ is an eigenvector of $\overline{W}$.

Observation A.4. *Consider a graph G with adjacency matrix A and diagonal degree matrix D.*

1. Let $1 = \lambda_1^P \geqslant \lambda_2^P \geqslant \cdots \geqslant \lambda_n^P$ be the eigenvalues of the transition matrix $P = D^{-1}A$ and let $0 = \lambda_1 \leqslant \lambda_1 \leqslant \cdots \leqslant \lambda_1$ be the eigenvalues of the normalized Laplacian $\mathcal{L} = I - D^{-1/2}AD^{-1/2}$. For every $i = 1,\ldots,n$ it holds that

$$\lambda_i = 1 - \lambda_i^P\,.$$

2. Let $1 = \bar{\lambda}_1 \geqslant \bar{\lambda}_2 \geqslant \cdots \geqslant \bar{\lambda}_n$ be the eigenvalues of $\overline{W} = I - L/(2m)$ and let $0 = \lambda_1^L \leqslant \lambda_2^L \leqslant \cdots \leqslant \lambda_n^L$ be the eigenvalues of the Laplacian matrix $L = D - A$. For every $i = 1,\ldots,n$ it holds that

$$\lambda_i^L = 2m(1 - \bar{\lambda}_i)\,.$$

3. For every $i = 1,\ldots,n$, the eigenvalues of L and $\mathcal{L}$ satisfy

$$d_{min}\lambda_i \leqslant \lambda_i^L \leqslant d_{max}\lambda_i\,,$$

where d_{min} and d_{max} are the minimum and the maximum degree of the nodes, respectively.

As a consequence of the above relationships among the eigenvalues of matrices L, $\mathcal{L}$, and $\overline{W}$, we easily get the following useful bounds.

Observation A.5. *Let G be an (n,d,γ)-clustered graph and let $\bar{\lambda}_i$ and λ_i, for $i = 1,\ldots,n$, be the eigenvalues of $\overline{W}$ and $\mathcal{L}$, respectively, in non-decreasing order. It holds that*

$$\frac{d}{2m}(1-2\gamma)\left(\lambda_3 - \lambda_2\right) \leqslant \bar{\lambda}_2 - \bar{\lambda}_3 \leqslant \frac{d}{2m}(1+2\gamma)\left(\lambda_3 - \lambda_2\right)\,.$$

Proof. We have:

$$\bar{\lambda}_2 - \bar{\lambda}_3 = \frac{\lambda_3^L - \lambda_2^L}{2m}.$$

To derive the lower bound, we write:

$$\frac{\lambda_3^L - \lambda_2^L}{2m} \geqslant \frac{\lambda_3 d_{\min} - \lambda_2 d_{\max}}{2m} \geqslant \frac{\lambda_3 d(1-\gamma) - \lambda_2 d(1+\gamma)}{2m} = \frac{(\lambda_3 - \lambda_2)d - \gamma d(\lambda_3 + \lambda_2)}{2m} \tag{13}$$

$$\geqslant \frac{(1-2\gamma)(\lambda_3 - \lambda_2)}{2m}. \tag{14}$$

Here, the first inequality is a direct consequence of Observation A.4, the second follows from the definition of (n,d,γ)-clustered graph, while the last inequality follows since $\lambda_3 + \lambda_2 \leqslant 2(\lambda_3 - \lambda_2)$, whenever $\lambda_3 \geqslant 3\lambda_2$.[19] The upper bound is derived in the same way, again using $\lambda_3 \geqslant 3\lambda_2$. □

[19] Note that the latter condition holds, since the hypotheses of Theorem 3.1 state that $\lambda_3 - \lambda_2 = \Omega(1)$, while the conditions on the cut implies that the graph's conductance is $o(1)$. The condition $\lambda_3 \geqslant 3\lambda_2$ is thus a consequence of Cheeger's inequality.

Observation A.6. *Since the random edges sequentially selected by the process are mutually independent, the expected state of the process at time t can be written as :*

$$\mathbb{E}\left[\mathbf{x}^{(t)}\right] \;=\; \mathbb{E}[W_t \cdot \ldots \cdot W_1 \cdot \mathbf{x}] = \left(\mathbb{E}[W]\right)^t \mathbf{x} \;=\; \overline{W}^t \cdot \mathbf{x}\,. \tag{15}$$

B Proofs for Section 3

B.1 Proof of Theorem 3.1: Technical lemmas

The next lemma decomposes the state of the system at time t, explicitely identifying components parallel to $\mathbf{1}$ and $\boldsymbol{\chi}$ respectively.

Lemma B.1 (Main decomposition)**.** *Let $\mathbf{x} \in \{-1,1\}^n$ be an arbitrary initial vector of values and, for $h = 1,2$, let $\mu_h = \mu_h(\mathbf{x}) = (2/n)\sum_{i\in V_h}\mathbf{x}(i)$ be the average of the initial values in block h. The expected vector of values at round t conditional on the initial vector being $\mathbf{x}$ can be written as*

$$\mathbb{E}\left[\mathbf{x}^{(t)} \,|\, \mathbf{x}^{(0)} = \mathbf{x}\right] = \alpha_1\mathbf{1} + \alpha_2\bar{\lambda}_2^t\boldsymbol{\chi} + \alpha_2\sqrt{n}\,\bar{\lambda}_2^t\mathbf{f}_\perp + \mathbf{e}^{(t)}\,,$$

where

$$\alpha_1 = \frac{\mu_1+\mu_2}{2}\,, \qquad\qquad \alpha_2 = \frac{1}{\|\mathbf{f}-\mathbf{f}_\perp\|^2}\left(\frac{\mu_1-\mu_2}{2} - \frac{\langle\mathbf{x},\mathbf{f}_\perp\rangle}{\sqrt{n}}\right)$$

and, moreover,

$$\left\|\mathbf{e}^{(t)}\right\| \leqslant \bar{\lambda}_3^t\sqrt{n}\,.$$

Proof. Let $1 = \bar{\lambda}_1 > \bar{\lambda}_2 \geqslant \cdots \geqslant \bar{\lambda}_n$ be the eigenvalues of $\overline{W}$ and let $\mathbf{w}_1 = \mathbf{1}/\sqrt{n}, \mathbf{w}_2, \ldots, \mathbf{w}_n$ be a basis of orthonormal eigenvectors of $\overline{W}$, so that we can write

$$\mathbb{E}\left[\mathbf{x}^{(t)} \,|\, \mathbf{x}^{(0)} = \mathbf{x}\right] = \sum_{i=1}^{n}\bar{\lambda}_i^t\langle\mathbf{x},\mathbf{w}_i\rangle\mathbf{w}_i\,.$$

Since $\bar{\lambda}_1 = 1$ and $\mathbf{w}_1 = \mathbf{1}/\sqrt{n}$, we have that $\langle\mathbf{x},\mathbf{w}_1\rangle\mathbf{w}_1 = (1/n)\langle\mathbf{x},\mathbf{1}\rangle\mathbf{1}$. Hence, $\alpha_1 = (1/n)\sum_{i\in V}\mathbf{x}(i) = (\mu_1+\mu_2)/2$.
Since $\mathbf{w}_2 = (\mathbf{f}-\mathbf{f}_\perp)/\|\mathbf{f}-\mathbf{f}_\perp\|$ we have that

$$\begin{aligned}\langle\mathbf{x},\mathbf{w}_2\rangle\mathbf{w}_2 &= \left\langle\mathbf{x}, \frac{\mathbf{f}-\mathbf{f}_\perp}{\|\mathbf{f}-\mathbf{f}_\perp\|}\right\rangle\frac{\mathbf{f}-\mathbf{f}_\perp}{\|\mathbf{f}-\mathbf{f}_\perp\|} = \frac{\langle\mathbf{x},\mathbf{f}-\mathbf{f}_\perp\rangle\mathbf{f} - \langle\mathbf{x},\mathbf{f}-\mathbf{f}_\perp\rangle\mathbf{f}_\perp}{\|\mathbf{f}-\mathbf{f}_\perp\|^2}\\ &= \frac{\langle\mathbf{x},\mathbf{f}-\mathbf{f}_\perp\rangle}{\|\mathbf{f}-\mathbf{f}_\perp\|^2\sqrt{n}}\boldsymbol{\chi} - \frac{\langle\mathbf{x},\mathbf{f}-\mathbf{f}_\perp\rangle}{\|\mathbf{f}-\mathbf{f}_\perp\|^2}\mathbf{f}_\perp\end{aligned}$$

Hence,

$$\begin{aligned}\alpha_2 &= \frac{\langle\mathbf{x},\mathbf{f}-\mathbf{f}_\perp\rangle}{\|\mathbf{f}-\mathbf{f}_\perp\|^2\sqrt{n}} = \frac{1}{\|\mathbf{f}-\mathbf{f}_\perp\|^2}\left(\frac{\langle\mathbf{x},\mathbf{f}\rangle}{\sqrt{n}} - \frac{\langle\mathbf{x},\mathbf{f}_\perp\rangle}{\sqrt{n}}\right)\\ &= \frac{1}{\|\mathbf{f}-\mathbf{f}_\perp\|^2}\left(\frac{\langle\mathbf{x},\boldsymbol{\chi}\rangle}{n} - \frac{\langle\mathbf{x},\mathbf{f}_\perp\rangle}{\sqrt{n}}\right) = \frac{1}{\|\mathbf{f}-\mathbf{f}_\perp\|^2}\left(\frac{\mu_1-\mu_2}{2} - \frac{\langle\mathbf{x},\mathbf{f}_\perp\rangle}{\sqrt{n}}\right)\end{aligned}$$

Finally, the bound on $\left\|\mathbf{e}^{(t)}\right\|$ easily follows since, by definition, $\mathbf{e}\perp\mathbf{w}_1,\mathbf{w}_2$ and $\bar{\lambda}_3 \geqslant \max\{\bar{\lambda}_i \,|\, i \geqslant 4\}$. □

Lemma B.2. *Recall that we name $m_{1,2} = |E(V_1,V_2)|$ the size of the cut. It holds that*

$$\|\mathbf{f}_\perp\|^2 \leqslant \frac{2}{\bar{\lambda}_2-\bar{\lambda}_3}\cdot\frac{m_{1,2}}{nm}\,.$$

Proof. Observe that since $\mathbf{f}$ is orthogonal to $\mathbf{1}$, we can write $\mathbf{f}^\intercal \overline{W} \mathbf{f}$ as

$$\begin{aligned}\mathbf{f}^\intercal \overline{W} \mathbf{f} &= \left(\mathbf{f}_{\|} + \mathbf{f}_{\perp}\right)^\intercal \overline{W} \left(\mathbf{f}_{\|} + \mathbf{f}_{\perp}\right) = \mathbf{f}_{\|}^\intercal \overline{W} \mathbf{f}_{\|} + \mathbf{f}_{\perp}^\intercal \overline{W} \mathbf{f}_{\perp} + 2\mathbf{f}_{\|}^\intercal \overline{W} \mathbf{f}_{\perp} \\ &= \mathbf{f}_{\|}^\intercal \overline{W} \mathbf{f}_{\|} + \mathbf{f}_{\perp}^\intercal \overline{W} \mathbf{f}_{\perp} = \bar{\lambda}_2 \|\mathbf{f}_{\|}\|^2 + \mathbf{f}_{\perp}^\intercal \overline{W} \mathbf{f}_{\perp} \end{aligned} \tag{16}$$

where we used the fact that $\overline{W}$ is symmetric, thus $\mathbf{f}_{\perp}^\intercal \overline{W} \mathbf{f}_{\|} = \mathbf{f}_{\|}^\intercal \overline{W} \mathbf{f}_{\perp} = 0$, and the fact that $\mathbf{f}_{\|}$ is in the second eigenspace of $\overline{W}$, thus $\mathbf{f}_{\|}^\intercal \overline{W} \mathbf{f}_{\|} = \bar{\lambda}_2 \|\mathbf{f}_{\|}\|^2$. Moreover, since $\mathbf{f}_{\perp}$ is orthogonal to the first two eigenspaces of $\overline{W}$, the third eigenvalue of $\overline{W}$ is $\bar{\lambda}_3 = \sup_{\mathbf{x} \perp \mathbf{1}, \mathbf{w}_2} \frac{\mathbf{x}^\intercal \overline{W} \mathbf{x}}{\|\mathbf{x}\|^2} \geqslant \frac{\mathbf{f}_{\perp}^\intercal \overline{W} \mathbf{f}_{\perp}}{\|\mathbf{f}_{\perp}\|^2}$ from (16) it follows that

$$\mathbf{f}^\intercal \overline{W} \mathbf{f} \leqslant \bar{\lambda}_2 \|\mathbf{f}_{\|}\|^2 + \bar{\lambda}_3 \|\mathbf{f}_{\perp}\|^2 \tag{17}$$

Observe that, since $\overline{W} = I - L/(2m) = I - \frac{D-A}{2m}$ we can also write $\mathbf{f}^\intercal \overline{W} \mathbf{f}$ as a function of $m_{1,2}$, indeed

$$\mathbf{f}^\intercal \overline{W} \mathbf{f} = 1 - \frac{1}{2m} \left(\mathbf{f}^\intercal D \mathbf{f} - \mathbf{f}^\intercal A \mathbf{f}\right) = 1 - \frac{1}{2mn} \left(\boldsymbol{\chi}^\intercal D \boldsymbol{\chi} - \boldsymbol{\chi}^\intercal A \boldsymbol{\chi}\right) = 1 - 2\frac{m_{1,2}}{mn} \tag{18}$$

where we used the fact that $\boldsymbol{\chi}^\intercal D \boldsymbol{\chi} = 2m$, the fact that $\boldsymbol{\chi}^\intercal A \boldsymbol{\chi} = \sum_i a_i - \sum_i b_i = 2m - 4m_{1,2}$.

From (17) and (18) and the fact that $1 = \|\mathbf{f}\|^2 = \|\mathbf{f}_{\|}\|^2 + \|\mathbf{f}_{\perp}\|^2$ we have

$$1 - 2\frac{m_{1,2}}{mn} \leqslant \bar{\lambda}_2 \|\mathbf{f}_{\|}\|^2 + \bar{\lambda}_3 \|\mathbf{f}_{\perp}\|^2 = \bar{\lambda}_2 - (\bar{\lambda}_2 - \bar{\lambda}_3)\|\mathbf{f}_{\perp}\|^2$$

and thus

$$\|\mathbf{f}_{\perp}\|^2 \leqslant \frac{2\frac{m_{1,2}}{mn} - (1 - \bar{\lambda}_2)}{\bar{\lambda}_2 - \bar{\lambda}_3} \leqslant \frac{2}{\bar{\lambda}_2 - \bar{\lambda}_3} \cdot \frac{m_{1,2}}{nm} .$$

□

Definition B.1 (Bad nodes)**.** *We say a node $u \in [n]$ is ε-*bad *if $|\mathbf{f}_{\perp,u}| \geqslant \varepsilon/\sqrt{n}$, for some $\varepsilon > 0$ and we call B_ε the set of ε-bad nodes.*

Notice that the property of being a bad node only depends on the graph and on the protocol, not on the "execution" of the protocol.

From the Lemma B.2, an upper bound on the number of ε-bad nodes easily follows.

Corollary B.1 (Number of bad nodes)**.** *The number $|B_\varepsilon|$ of ε-bad nodes is upper bounded by*

$$|B_\varepsilon| \leqslant \frac{2m_{1,2}}{\varepsilon^2(\bar{\lambda}_2 - \bar{\lambda}_3)m} .$$

Proof. Assume B_ε vertices satisfy $(\mathbf{f}_{\perp})_i > \frac{\varepsilon}{\sqrt{n}}$. Lemma B.2 implies:

$$\frac{\varepsilon^2}{n} B_\varepsilon \leqslant \frac{2}{\bar{\lambda}_2 - \bar{\lambda}_3} \cdot \frac{m_{1,2}}{nm},$$

from which the thesis follows. □

Lemma B.3 (Monotonicity property)**.** *Let G be a connected graph, let $\bar{\lambda}_i$ for $i = 1, \ldots, n$ be the eigenvalues of matrix $\overline{W}$, let ε be such that $0 < \varepsilon < 1$, and let $\mathbf{x} \in \{-1, 1\}^n$ be an arbitrary initial vector such that $|\alpha_2| = |\alpha_2(\mathbf{x})| > 0$. Then, for every node $u \notin B_\varepsilon$ and for any round t such that $t \geqslant 3 \log\left(\frac{n}{1-\varepsilon}\right) / \log(\bar{\lambda}_2/\bar{\lambda}_3)$, it holds that*

$$\mathbf{sgn}\left(\mathbb{E}\left[\mathbf{x}_u^{(t-1)} \mid \mathbf{x}^{(0)} = \mathbf{x}\right] - \mathbb{E}\left[\mathbf{x}_u^{(t)} \mid \mathbf{x}^{(0)} = \mathbf{x}\right]\right) = \mathbf{sgn}\left(\alpha_2 \boldsymbol{\chi}_u\right) . \tag{19}$$

Proof. Thanks to Lemma B.1, the expected difference of the value of a node in two consecutive rounds is

$$\mathbb{E}\left[\mathbf{x}_u^{(t-1)} \,|\, \mathbf{x}^{(0)} = \mathbf{x}\right] - \mathbb{E}\left[\mathbf{x}_u^{(t)} \,|\, \mathbf{x}^{(0)} = \mathbf{x}\right] = \alpha_2 \bar{\lambda}_2^{(t-1)}(1-\bar{\lambda}_2)[\boldsymbol{\chi}_u - \sqrt{n}\,\mathbf{f}_{\perp,u}] + \mathbf{e}^{(t-1)} - \mathbf{e}^{(t)}$$

From the above equation we get that, as soon as

$$|\mathbf{e}_u^{(t-1)} - \mathbf{e}_u^{(t)}| < |\alpha_2|\,\bar{\lambda}_2^t(1-\bar{\lambda}_2)|\boldsymbol{\chi}_u \pm \varepsilon|,$$

the sign of the expected difference between two consecutive values of a non ε-bad node u indicates the community node u belongs to. Moreover, since $|\boldsymbol{\chi}_u \pm \varepsilon| \geqslant 1-\varepsilon$ and $|\mathbf{e}_u^{(t-1)} - \mathbf{e}_u^{(t)}| \leqslant 2\,\bar{\lambda}_3^t$, the above sign property turns out to be true for every round t such that

$$t \geqslant \log\left(\frac{2}{|\alpha_2|(1-\bar{\lambda}_2)(1-\varepsilon)}\right) / \log(\bar{\lambda}_2/\bar{\lambda}_3) \tag{20}$$

The thesis thus follows from the fact that, if $|\alpha_2| > 0$ then it is at least $1/n$ and if the graph is connected then $1-\bar{\lambda}_2 \geqslant 1/n$. Hence, any

$$t \geqslant 3\log\left(\frac{n}{1-\varepsilon}\right) / \log(\bar{\lambda}_2/\bar{\lambda}_3)$$

satisfies (20). □

Lemma B.4 (Sign property)**.** *Let G be a connected graph, let $\bar{\lambda}_i$ for $i = 1,\ldots,n$ be the eigenvalues of matrix $\overline{W}$, let ε be such that $0 < \varepsilon < 1$, and let $\mathbf{x} \in \{-1,1\}^n$ be an arbitrary initial vector such that $\alpha_1 = \alpha_1(\mathbf{x})$ and $\alpha_2 = \alpha_2(\mathbf{x})$ satisfy $|\alpha_2| > 2|\alpha_1|/(1-\varepsilon)$. Then, for every node $u \notin B_\varepsilon$ and for any round t such that*

$$\frac{1}{\log(1/\bar{\lambda}_3)}\log(n/|\alpha_1|) \leqslant t \leqslant \frac{1}{\log(1/\bar{\lambda}_2)}\log\left(\frac{|\alpha_2|(1-\varepsilon)}{2|\alpha_1|}\right)$$

it holds that $\mathbf{sgn}\left(\mathbb{E}\left[\mathbf{x}_u^{(t)} \,|\, \mathbf{x}^{(0)} = \mathbf{x}\right]\right) = \mathbf{sgn}\left(\alpha_2\boldsymbol{\chi}_u\right)$.

Proof. From (4) it follows that, if $|\alpha_1 + \mathbf{e}_u^{(t)}| < |\alpha_2|\,\bar{\lambda}_2^t\,|\boldsymbol{\chi}_u \pm \varepsilon|$, then the sign of the expected value of a non ε-bad node u indicates the block node u belongs to. Notice that, for

$$t \geqslant \frac{1}{\log(1/\bar{\lambda}_3)}\log(n/|\alpha_1|) \tag{21}$$

we have that

$$|\alpha_1 + \mathbf{e}_u^{(t)}| \leqslant |\alpha_1| + |\mathbf{e}_u^{(t)}| \leqslant |\alpha_1| + n\bar{\lambda}_3^{(t)} \leqslant 2|\alpha_1|$$

And for

$$t \leqslant \frac{1}{\log(1/\bar{\lambda}_2)}\log\left(\frac{|\alpha_2|(1-\varepsilon)}{2|\alpha_1|}\right) \tag{22}$$

we have that

$$2|\alpha_1| \leqslant |\alpha_2|\,\bar{\lambda}_2^t\,(1-\varepsilon) \leqslant |\alpha_2|\,\bar{\lambda}_2^t\,|\boldsymbol{\chi}_u \pm \varepsilon|$$

Hence, if the time-window defined by (21) and (22) is non-empty then for all t in it

$$\frac{1}{\log(1/\bar{\lambda}_3)}\log(n/|\alpha_1|) \leqslant t \leqslant \frac{1}{\log(1/\bar{\lambda}_2)}\log\left(\frac{|\alpha_2|(1-\varepsilon)}{2|\alpha_1|}\right)$$

the sign of the expected value of a non-bad node u equals the sign of $\alpha_2\boldsymbol{\chi}_u$. □

Lemma B.5 (Projection of the initial random vector.). *Let* $\mathbf{x}$ *be chosen uniformly at random in* $\{-1,1\}^n$. *Then, two absolute constants* $\beta_1, \beta_2 > 0$ *exist, such that:*

- *with probability at least* β_1, *it holds that* $\alpha_2 = \alpha_2(\mathbf{x}) > 0$,
- *with probability at least* β_2, *it holds that* $|\alpha_2| > 2|\alpha_1|/(1-\varepsilon)$.

Sketch of the Proof. Both α_1 and α_2 are linear combinations of a sum of n independent Rademacher random variables. Then, the two claims can be derived easily from Lemma A.1. □

C Proofs for Section 4.1

C.1 Proof of Theorem 4.1

From the fact that random matrix $W \sim \mathcal{W}$ defined by one step of our averaging process is symmetric and idempotent ($W^\intercal W = W$) we get the following upper bound on the expected squared norm of $\mathbf{y}+\mathbf{z}$ at the next step as a function of their squared norm at the current step. For readability sake, in the following proofs of this section we use $\mathbf{y}'$ and $\mathbf{z}'$ for random variables $\mathbf{y}^{(t+1)}$ and $\mathbf{z}^{(t+1)}$ conditional on the state at round t being $\mathbf{x}^{(t)} = \mathbf{x} = \mathbf{x}_\parallel + \mathbf{y} + \mathbf{z}$.

Lemma C.1. *Let* $\mathbf{x} = \mathbf{x}_\parallel + \mathbf{y} + \mathbf{z} \in [-1,1]^n$ *be an arbitrary vector of states. After one step of Algorithm 1 it holds that*

$$\mathbb{E}\left[\|\mathbf{y}^{(t+1)} + \mathbf{z}^{(t+1)}\|^2 \,|\, \mathbf{x}^{(t)} = \mathbf{x}\right] \leqslant \left(1 - \frac{\lambda_2}{n}\right)\|\mathbf{y}\|^2 + \left(1 - \frac{\lambda_3}{n}\right)\|\mathbf{z}\|^2.$$

Proof. Since random matrix W is symmetric and idempotent, it holds that

$$\begin{aligned}
\mathbb{E}\left[\|\mathbf{y}' + \mathbf{z}'\|^2\right] &= \mathbb{E}\left[\|W(\mathbf{y}+\mathbf{z})\|^2\right] \\
&= (\mathbf{y}+\mathbf{z})^T\left(I - \frac{1}{n}L\right)(\mathbf{y}+\mathbf{z}) \\
&= \|\mathbf{y}\|^2 + \|\mathbf{z}\|^2 - \frac{1}{n}\mathbf{y}^T L\mathbf{y} - \frac{1}{n}\mathbf{z}^T L\mathbf{z} \\
&\leqslant \left(1 - \frac{\lambda_2}{n}\right)\|\mathbf{y}\|^2 + \left(1 - \frac{\lambda_3}{n}\right)\|\mathbf{z}\|^2.
\end{aligned}$$

□

The squared norm of $\mathbf{y}$ at the next step can be lower bounded as a function of the squared norms of $\mathbf{y}$ and $\mathbf{z}$ at the current time step as follows. If the underlying graph is (n,d,b)-regular, we can get an upper bound as well.

Lemma C.2. *Let* $\mathbf{x} = \mathbf{x}_\parallel + \mathbf{y} + \mathbf{z} \in [-1,1]^n$ *be an arbitrary vector of states. After one step of Algorithm 1 it holds that*

$$\mathbb{E}\left[\|\mathbf{y}^{(t+1)}\|^2 \,|\, \mathbf{x}^{(t)} = \mathbf{x}\right] \geqslant \left(1 - \frac{2\lambda_2}{n}\right)\|\mathbf{y}\|^2.$$

Moreover, if the underlying graph G *is an* (n,d,b)*-clustered regular graph with* $\lambda_2 = 2b/d = o(\lambda_3/\log n)$ *we also have that*

$$\mathbb{E}\left[\|\mathbf{y}^{(t+1)}\|^2 \,|\, \mathbf{x}^{(t)} = \mathbf{x}\right] \leqslant \left(1 - \frac{2\lambda_2}{n}\right)\|\mathbf{y}\|^2 + 2\frac{\lambda_2}{n^2}\left(\|\mathbf{y}\|^2 + \|\mathbf{z}\|^2\right).$$

Proof. Let $\{u,v\} \in E$ be the random edge sampled at step t, call $W_{u,v} \sim \mathcal{W}$ be the corresponding random matrix with $L_{u,v}$ be such that $W = I - \frac{1}{2}L_{u,v}$. As for the lower bound we have

$$\begin{aligned}
\mathbb{E}\left[\|\mathbf{y}'\|^2\right] &= \mathbb{E}\left[\|Q_2 W_{u,v}(\mathbf{y}+\mathbf{z})\|^2\right] \\
&= \mathbb{E}\left[\|Q_2\left(I - \frac{1}{2}L_{u,v}\right)(\mathbf{y}+\mathbf{z})\|^2\right] \\
&= \mathbb{E}\left[\|\mathbf{y} - \frac{1}{2}Q_2 L_{u,v}(\mathbf{y}+\mathbf{z})\|^2\right] \\
&= \mathbb{E}\left[\|\mathbf{y}\|^2 - \mathbf{y}^T Q_2 L_{u,v}(\mathbf{y}+\mathbf{z}) + \|\frac{1}{2}Q_2 L_{u,v}(\mathbf{y}+\mathbf{z})\|^2\right] \\
&\geqslant \|\mathbf{y}\|^2 - \mathbf{y}^T Q_2 \mathbb{E}[L_{u,v}](\mathbf{y}+\mathbf{z}) \\
&= \|\mathbf{y}\|^2 - \frac{1}{n}\mathbf{y}^T Q_2 L \mathbf{y} \\
&= \|\mathbf{y}\|^2\left(1 - \frac{2\lambda_2}{n}\right),
\end{aligned}$$

where the last equality follows since Q_2 is the projector along the direction of $\mathbf{v}_2 = \boldsymbol{\chi}/\|\boldsymbol{\chi}\|$, which in turn is L's second eigenvector.
As for the upper bound, it holds that

$$\begin{aligned}
\mathbb{E}\left[\|\mathbf{y}'\|^2\right] &= \mathbb{E}\left[\|Q_2 W_{u,v}(\mathbf{y}+\mathbf{z})\|^2\right] \\
&= \mathbb{E}\left[\|Q_2\left(I - \frac{1}{2}L_{u,v}\right)(\mathbf{y}+\mathbf{z})\|^2\right] \\
&= \mathbb{E}\left[\|\mathbf{y} - \frac{1}{2}Q_2 L_{u,v}(\mathbf{y}+\mathbf{z})\|^2\right] \\
&= \mathbb{E}\left[\|\mathbf{y}\|^2 - \mathbf{y}^T Q_2 L_{u,v}(\mathbf{y}+\mathbf{z}) + \|\frac{1}{2}Q_2 L_{u,v}(\mathbf{y}+\mathbf{z})\|^2\right] \\
&= \left(1 - \frac{2\lambda_2}{n}\right)\|\mathbf{y}\|^2 + \mathbb{E}\left[\left(\frac{1}{2}\mathbf{v}_2^T L_{u,v}(\mathbf{y}+\mathbf{z})\right)^2\right] \\
&= \left(1 - \frac{2\lambda_2}{n}\right)\|\mathbf{y}\|^2 + \frac{1}{4}\|\mathbf{y}\|^2\mathbb{E}\left[\left(\mathbf{v}_2^T L_{u,v}\mathbf{v}_2\right)^2\right] \\
&\quad + \frac{1}{2}\mathbb{E}\left[\mathbf{v}_2^T L_{u,v}\mathbf{y}\mathbf{v}_2^T L_{u,v}\mathbf{z}\right] + \frac{1}{4}\mathbb{E}\left[\left(\mathbf{v}_2^T L_{u,v}\mathbf{z}\right)^2\right] \\
&= \left(1 - \frac{2\lambda_2}{n}\right)\|\mathbf{y}\|^2 + \frac{1}{4}\|\mathbf{y}\|^2\mathbb{E}\left[\left(\mathbf{v}_2(u) - \mathbf{v}_2(v)\right)^4\right] \\
&+ \frac{1}{2}\|\mathbf{y}\|\mathbb{E}\left[\left(\mathbf{v}_2(u) - \mathbf{v}_2(v)\right)^3\left(\mathbf{z}(u) - \mathbf{z}(v)\right)\right] \\
&\quad + \frac{1}{4}\mathbb{E}\left[\left(\mathbf{v}_2(u) - \mathbf{v}_2(v)\right)^2\left(\mathbf{z}(u) - \mathbf{z}(v)\right)^2\right]. \qquad (23)
\end{aligned}$$

Next, note that we have

$$\mathbb{E}\left[\left(\mathbf{v}_2(u) - \mathbf{v}_2(v)\right)^4\right] = \frac{1}{nd}\sum_{(u,v)\in E(V_1,V_2)}\frac{4}{n^2} = \frac{b}{dn^2} = \frac{\lambda_2}{2n^2}, \qquad (24)$$

where we used that $\lambda_2 = 2b/d$ and the fact that $\mathbf{v}_2(u) = 1/\sqrt{n}$ if u belongs to first community and $\mathbf{v}_2(u) = -1/\sqrt{n}$ when v belongs to the second community. We further get that

$$\mathbb{E}\left[\left(\mathbf{v}_2(u) - \mathbf{v}_2(v)\right)^3\left(\mathbf{z}(u) - \mathbf{z}(v)\right)\right] = \frac{1}{nd}\cdot\frac{8}{n\sqrt{n}}\sum_{(u,v)\in E(V_1,V_2)}\left(\mathbf{z}(u) - \mathbf{z}(v)\right)$$

$$= \frac{1}{nd} \cdot \frac{8}{n\sqrt{n}} \left(\sum_{u \in V_1} b\mathbf{z}(u) - \sum_{v \in V_2} b\mathbf{z}(v) \right) = 0, \tag{25}$$

where it is understood that if (u, v) belongs to the cut, then $u \in V_1$ and $v \in V_2$ and where, to derive the last equality, we recall that $\mathbf{z} \perp \mathbf{span}\{\mathbf{1}, \boldsymbol{\chi}\}$.
Finally, we get that

$$\mathbb{E}\left[(\mathbf{v}_2(u) - \mathbf{v}_2(v))^2 (\mathbf{z}(u) - \mathbf{z}(v))^2\right] = \frac{8}{dn^2} \sum_{(u,v) \in E(V_1, V_2)} (\mathbf{z}(u) - \mathbf{z}(v))^2 \tag{26}$$

$$= \frac{8}{dn^2} \|\mathbf{z}\|^2 \sum_{(u,v) \in E(V_1, V_2)} \frac{(\mathbf{z}(u) - \mathbf{z}(v))^2}{\|\mathbf{z}\|^2}$$

$$\leqslant \frac{16b}{dn^2} \|\mathbf{z}\|^2 = \frac{8\lambda_2}{n^2} \|\mathbf{z}\|^2, \tag{27}$$

where the last inequality follows by observing that $\sum_{(u,v) \in E(V_1, V_2)} \frac{(\mathbf{z}(u) - \mathbf{z}(v))^2}{\|\mathbf{z}\|^2}$ is the Rayleigh quotient of the unnormalized Laplacian of a bipartite b-regular graph and the largest possible eigenvalue is $2b$. The thesis follows by using (24), (25), and (26) in (23). □

Lemma C.3. *Let $\mathbf{x} = \mathbf{x}_{\parallel} + \mathbf{y} + \mathbf{z} \in [-1, 1]^n$ be an arbitrary vector of states. After one step of Algorithm 1 it holds that*

$$\mathbb{E}\left[\|\mathbf{z}^{(t+1)}\| \mid \mathbf{x}^{(t)} = \mathbf{x}\right]^2 \leqslant \frac{\lambda_2}{n} \|\mathbf{y}\|^2 + \left(1 - \frac{\lambda_3}{n}\right) \|\mathbf{z}\|^2.$$

Proof. From Pythagoras' Theorem and Lemmas C.1 and C.2, we get

$$\begin{aligned} \mathbb{E}\left[\|\mathbf{z}'\|^2\right] &= \mathbb{E}\left[\|\mathbf{y}' + \mathbf{z}'\|^2\right] - \mathbb{E}\left[\|\mathbf{y}'\|^2\right] \\ &\leqslant \left(1 - \frac{\lambda_2}{n}\right) \|\mathbf{y}\|^2 + \left(1 - \frac{\lambda_3}{n}\right) \|\mathbf{z}\|^2 - \left(1 - \frac{2\lambda_2}{n}\right) \|\mathbf{y}\|^2 \\ &= \frac{\lambda_2}{n} \|\mathbf{y}\|^2 + \left(1 - \frac{\lambda_3}{n}\right) \|\mathbf{z}\|^2. \end{aligned}$$

□

Finally, by unrolling the double recursion, we get that the expected squared norm of $\mathbf{z}$ and $\mathbf{y}$ at round t satisfy the following inequality.

Lemma C.4. *Let G be an (n, d, b)-clustered regular graph with $\lambda_2 = \frac{2b}{d} = o(\lambda_3 / \log n)$. For every starting state $\mathbf{x}^{(0)} \in \{-1, +1\}^n$ and for every $t \in \mathbb{N}$ it holds that*

$$\mathbb{E}\left[\|\mathbf{z}^{(t)}\|^2\right] \leqslant \frac{\lambda_2}{\lambda_3} \left(1 - \frac{\lambda_2}{n}\right)^{-t} \mathbb{E}\left[\|\mathbf{y}^{(t)}\|^2\right] + \left(1 - \frac{\lambda_3}{n}\right)^t \mathbb{E}\left[\|\mathbf{z}^{(0)}\|^2\right].$$

Proof. We first prove the following inequality

$$\mathbb{E}\left[\|\mathbf{z}^{(t)}\|^2\right] \leqslant \frac{\lambda_2}{\lambda_3} \max_i \mathbb{E}\left[\|\mathbf{y}^{(i)}\|^2\right] + \left(1 - \frac{\lambda_3}{n}\right)^t \mathbb{E}\left[\|\mathbf{z}^{(0)}\|^2\right]. \tag{28}$$

Indeed, from Lemma C.3 we get

$$\mathbb{E}\left[\|\mathbf{z}^{(t)}\|^2\right] \leqslant \frac{\lambda_2}{n} \mathbb{E}\left[\|\mathbf{y}^{(t-1)}\|^2\right] + \left(1 - \frac{\lambda_3}{n}\right) \mathbb{E}\left[\|\mathbf{z}^{(t-1)}\|^2\right]$$

$$\leqslant \frac{\lambda_2}{n} \sum_{i \leqslant t-1} \mathbb{E}\left[\|\mathbf{y}^{(i)}\|^2\right] \left(1 - \frac{\lambda_3}{n}\right)^{t-1-i} + \left(1 - \frac{\lambda_3}{n}\right)^t \mathbb{E}\left[\|\mathbf{z}^{(0)}\|^2\right]$$

$$\leqslant \frac{\lambda_2}{n} \max_{i \leqslant t-1} \mathbb{E}\left[\|\mathbf{y}^{(i)}\|^2\right] \sum_i \left(1 - \frac{\lambda_3}{n}\right)^{t-1-i} + \left(1 - \frac{\lambda_3}{n}\right)^t \mathbb{E}\left[\|\mathbf{z}^{(0)}\|^2\right]$$

$$= \frac{\lambda_2}{n} \max_{i \leqslant t-1} \mathbb{E}\left[\|\mathbf{y}^{(i)}\|^2\right] \frac{1 - \left(1 - \frac{\lambda_3}{n}\right)^t}{\frac{\lambda_3}{n}} + \left(1 - \frac{\lambda_3}{n}\right)^t \mathbb{E}\left[\|\mathbf{z}^{(0)}\|^2\right]$$

$$\leqslant \frac{\lambda_2}{\lambda_3} \max_{i \leqslant t-1} \mathbb{E}\left[\|\mathbf{y}^{(i)}\|^2\right] + \left(1 - \frac{\lambda_3}{n}\right)^t \mathbb{E}\left[\|\mathbf{z}^{(0)}\|^2\right].$$

Next we observe from Lemma C.2, for each i we have

$$\mathbb{E}\left[\|\mathbf{y}^{(t)}\|^2\right] \geqslant \left(1 - \frac{2\lambda_2}{n}\right)^i \mathbb{E}\left[\|\mathbf{y}^{(t-i)}\|^2\right]$$

which means that

$$\mathbb{E}\left[\|\mathbf{y}^{(t-i)}\|^2\right] \leqslant \left(1 - \frac{2\lambda_2}{n}\right)^{-i} \mathbb{E}\left[\|\mathbf{y}^{(t)}\|^2\right] \leqslant \left(1 - \frac{2\lambda_2}{n}\right)^{-t} \mathbb{E}\left[\|\mathbf{y}^{(t)}\|^2\right].$$

Hence,

$$\max_{i \leqslant t-1} \mathbb{E}\left[\|\mathbf{y}^{(i)}\|^2\right] \leqslant \left(1 - \frac{\lambda_2}{n}\right)^{-t} \mathbb{E}\left[\|\mathbf{y}^{(t)}\|^2\right]. \tag{29}$$

The thesis follows by using (29) in (28). □

Wrapping up: Proof of Theorem 4.1

Since $\mathbf{y}^{(t)} - \mathbf{y}^{(0)}$ and $\mathbf{z}^{(t)}$ are orthogonal we can write

$$\mathbb{E}\left[\left\|\mathbf{y}^{(t)} + \mathbf{z}^{(t)} - \mathbf{y}^{(0)}\right\|^2\right] = \mathbb{E}\left[\left\|\mathbf{y}^{(t)} - \mathbf{y}^{(0)}\right\|^2\right] + \mathbb{E}\left[\left\|\mathbf{z}^{(t)}\right\|^2\right]. \tag{30}$$

The proof proceeds by bounding the two terms above separately. As for the first term, observe that

$$\begin{aligned}\mathbb{E}\left[\|\mathbf{y}^{(t)} - \mathbf{y}^{(0)}\|^2 \,|\mathbf{y}^{(0)}\right] &= \mathbb{E}\left[\|\mathbf{y}^{(t)}\|^2\right] + \|\mathbf{y}^{(0)}\|^2 - 2\left\langle \mathbb{E}\left[\mathbf{y}^{(t)}\right], \mathbf{y}^{(0)}\right\rangle \\ &\leqslant \mathbb{E}\left[\|\mathbf{y}^{(t)}\|^2\right] + \|\mathbf{y}^{(0)}\|^2 - 2\left(1 - \frac{\lambda_2}{n}\right)^t \|\mathbf{y}^{(0)}\|^2, \end{aligned} \tag{31}$$

where the last inequality follows from the lower bound in Lemma C.2. By taking the expectation of both sides of the previous inequality with respect to the random choice of the initial state, we immediately have that (31) holds if $\|\mathbf{y}^{(0)}\|^2$ is replaced by $\mathbb{E}\left[\|\mathbf{y}^{(0)}\|^2\right]$. Moreover, the upper bound in Lemma C.2 yields

$$\mathbb{E}\left[\left\|\mathbf{y}'\right\|^2\right] \leqslant \left(1 - \frac{2\lambda_2}{n}\right) \mathbb{E}[\|\mathbf{y}\|^2] + 2\frac{\lambda_2}{n^2} \left(\mathbb{E}[\|\mathbf{y}\|^2] + \mathbb{E}[\|\mathbf{z}\|^2]\right).$$

Recall the upper bound on $\mathbb{E}\left[\|\mathbf{z}^{(t-1)}\|^2\right]$ given by Lemma C.4, namely

$$\mathbb{E}\left[\|\mathbf{z}^{(t-1)}\|^2\right] \leqslant \frac{\lambda_2}{\lambda_3}\left(1 - \frac{\lambda_2}{n}\right)^{-(t-1)} \mathbb{E}\left[\|\mathbf{y}^{(t-1)}\|^2\right] + \left(1 - \frac{\lambda_3}{n}\right)^{t-1} \mathbb{E}\left[\|\mathbf{z}^{(0)}\|^2\right]. \tag{32}$$

We thus get, for $t = \mathcal{O}(n/\lambda_2)$,

$$\mathbb{E}\left[\|\mathbf{y}^{(t)}\|^2\right] \leqslant f(n)\mathbb{E}\left[\|\mathbf{y}^{(t-1)}\|^2\right] + \frac{2\lambda_2}{n^2}\left(1-\frac{\lambda_3}{n}\right)^{t-1}\mathbb{E}\left[\|\mathbf{z}^{(0)}\|^2\right], \tag{33}$$

where

$$f(n) = \left(1-\frac{2\lambda_2}{n}\right) + \frac{2\lambda_2}{n^2} + \frac{2\lambda_2^2}{\lambda_3 n^2}\left(1-\frac{\lambda_2}{n}\right)^{-(t-1)} \leqslant \left(1-\frac{\lambda_2}{n}\right) \leqslant 1, \tag{34}$$

for the values of t under consideration. We can unfold the recursion above to obtain

$$\begin{aligned}\mathbb{E}\left[\|\mathbf{y}^{(t)}\|^2\right] &\leqslant \left(1-\frac{\lambda_2}{n}\right)^t\mathbb{E}\left[\|\mathbf{y}^{(0)}\|^2\right] + \frac{2\lambda_2}{n^2}\mathbb{E}\left[\|\mathbf{z}^{(0)}\|^2\right]\sum_{j=0}^{t-1}\left(1-\frac{\lambda_3}{n}\right)^j \\ &\leqslant \left(1-\frac{\lambda_2}{n}\right)^t\mathbb{E}\left[\|\mathbf{y}^{(0)}\|^2\right] + \frac{2\lambda_2}{n\lambda_3}\mathbb{E}\left[\|\mathbf{z}^{(0)}\|^2\right]. \end{aligned}\tag{35}$$

As for the second term in (30), from (32) we have that

$$\begin{aligned}\mathbb{E}\left[\|\mathbf{z}^{(t)}\|^2\right] &\leqslant \frac{\lambda_2}{\lambda_3}\left(1-\frac{\lambda_2}{n}\right)^{-t}\mathbb{E}\left[\|\mathbf{y}^{(t)}\|^2\right] + \left(1-\frac{\lambda_3}{n}\right)^t\mathbb{E}\left[\|\mathbf{z}^{(0)}\|^2\right] \\ &\leqslant \frac{\lambda_2}{\lambda_3}\mathbb{E}\left[\|\mathbf{y}^{(0)}\|^2\right] + \frac{2}{n}\left(1-\frac{\lambda_2}{n}\right)^{-t}\left(\frac{\lambda_2}{\lambda_3}\right)^2\mathbb{E}\left[\|\mathbf{z}^{(0)}\|^2\right] + \left(1-\frac{\lambda_3}{n}\right)^t\mathbb{E}\left[\|\mathbf{z}^{(0)}\|^2\right] \\ &\leqslant \frac{\lambda_2}{\lambda_3}\mathbb{E}\left[\|\mathbf{y}^{(0)}\|^2\right] + \left(\frac{4}{n}\left(\frac{\lambda_2}{\lambda_3}\right)^2 + \frac{1}{n^3}\right)\mathbb{E}\left[\|\mathbf{z}^{(0)}\|^2\right], \end{aligned}\tag{36}$$

where in the last inequality we used the fact that $(1-\lambda_2/n)^{-t} \leqslant 2$ for $t \leqslant n/(4\lambda_2)$ and the fact that $(1-\lambda_3/n)^t \leqslant 1/n^3$ for $t \geqslant (3n/\lambda_3)\log n$.

Using (35) in (31) and then (31) and (36) in (30) we get

$$\begin{aligned}&\mathbb{E}\left[\left\|\mathbf{y}^{(t)} + \mathbf{z}^{(t)} - \mathbf{y}^{(0)}\right\|^2\right] \\ &\leqslant \left[1-\left(1-\frac{\lambda_2}{n}\right)^t + \frac{\lambda_2}{\lambda_3}\right]\mathbb{E}\left[\|\mathbf{y}^{(0)}\|^2\right] + \left[\frac{2\lambda_2}{n\lambda_3} + \frac{4}{n}\left(\frac{\lambda_2}{\lambda_3}\right)^2 + \frac{1}{n^3}\right]\mathbb{E}\left[\|\mathbf{z}^{(0)}\|^2\right] \\ &\leqslant 2\frac{\lambda_2 t}{n}\mathbb{E}\left[\|\mathbf{y}^{(0)}\|^2\right] + \frac{1}{n^3}\mathbb{E}\left[\|\mathbf{z}^{(0)}\|^2\right].\end{aligned}$$

The thesis then follows from the fact that $\mathbb{E}\left[\|\mathbf{y}^{(0)}\|^2\right] = 1$ and $\mathbb{E}\left[\|\mathbf{z}^{(0)}\|^2\right] \leqslant 1/n$. □

C.2 Proofs of Corollary 4.1 and of Equation (8)

- As for Corollary 4.1, we first note that t meets the conditions of Theorem 4.1. Moreover, $t \leqslant c\frac{n}{\lambda_3}\log n$, so we immediately have:

$$\mathbb{E}[B_t] \leqslant \frac{n}{\varepsilon^2\|\mathbf{y}^{(0)}\|^2}\mathbb{E}\left[\|\mathbf{y}^{(t)} + \mathbf{z}^{(t)} - \mathbf{y}^{(0)}\|^2\right] \leqslant \frac{4\lambda_2 t}{\varepsilon^2} \leqslant 4c\frac{\lambda_2}{\lambda_3}n\log n,$$

which is at most $\varepsilon^2 n$, whenever $\frac{\lambda_2}{\lambda_3} \leqslant \frac{\varepsilon^4}{4c\log n}$. Hence, the second claim follows directly from Markov's inequality.

- As for Equation (8), note that the definition of ε-bad node implies:

$$|B_t| \leqslant \frac{n}{\varepsilon^2\|\mathbf{y}^{(0)}\|^2}\|\mathbf{x}^{(t)} - \mathbf{x}_\parallel - \mathbf{y}^{(0)}\|^2 \stackrel{(a)}{=} \frac{n}{\varepsilon^2\|\mathbf{y}^{(0)}\|^2}\|\mathbf{y}^{(t)} + \mathbf{z}^{(t)} - \mathbf{y}^{(0)}\|^2,$$

where in (a) we used (7). This easily implies (8).

C.3 Proof of Lemma 4.2

In order to prove Lemma 4.2, we need some preliminary results.

Claim 1. *Let $W_1, \dots, W_t$ be denote a sequence of matrices describing t steps of the* Averaging *protocol that includes c cross edges and $t-c$ internal edges. Then*

$$\|W_t \cdots W_1 \boldsymbol{\chi} - \boldsymbol{\chi}\|^2 \leqslant 4c.$$

Furthermore, if $W_1, \dots, W_t$ are chosen randomly according to the Averaging *protocol we have,*

$$\mathbb{E}\left[\|W_t \cdots W_1 \boldsymbol{\chi} - \boldsymbol{\chi}\|^2\right] \leqslant 4\frac{tb}{d} \text{ and}$$
$$\mathbb{P}\left[\|W_t \cdots W_1 \boldsymbol{\chi} - \boldsymbol{\chi}\|^2 \geqslant 8t\frac{b}{d}\right] \leqslant e^{-\Omega(bt/d)}.$$

Proof. We have

$$\begin{aligned}\|W_t \cdots W_1 \boldsymbol{\chi} - \boldsymbol{\chi}\|^2 &= \|W_t \cdots W_1 \boldsymbol{\chi}\|^2 - 2\boldsymbol{\chi}^T W_t \cdots W_1 \boldsymbol{\chi} + \|\boldsymbol{\chi}\|^2 \\ &\leqslant 2n - 2\boldsymbol{\chi}^T W_t \cdots W_1 \boldsymbol{\chi}.\end{aligned}$$

To complete the proof, observe that for every vector $\mathbf{w}$ such that $\|\mathbf{w}\|_\infty \leqslant 1$, we have $\boldsymbol{\chi}^T W_{u,v} \mathbf{w} = \boldsymbol{\chi}^T \mathbf{w}$ if (u,v) is an internal edge and

$$\begin{aligned}\boldsymbol{\chi}^T W_{u,v} \mathbf{w} &= \boldsymbol{\chi}^T \left(\mathbf{w} + \frac{1}{2}(w_u - w_v)\mathbf{1}_v + \frac{1}{2}(w_v - w_u)\mathbf{1}_u\right) \\ &\geqslant \boldsymbol{\chi}^T \mathbf{w} - 2.\end{aligned}$$

By induction we thus get

$$\boldsymbol{\chi}^T W_t \cdots W_1 \boldsymbol{\chi} \geqslant n - 2c,$$

which implies the first part of the lemma. The furthermore part follows by noting that the average of c is bt/d, and that Chernoff bounds imply that c is concentrated around its average. □

Claim 2. *Let $t_1 = 6\frac{n}{\lambda_3}\log n$ and assume $\frac{\lambda_2}{\lambda_3} \leqslant \frac{\varepsilon^4}{96 \log n}$. Then*

$$\mathbb{P}\left[\forall t \in \{t_1, \dots, 2t_1\} : |B_t| \leqslant \frac{48\lambda_2 t}{\varepsilon^3}\right] \geqslant 1 - \varepsilon - \frac{\log n}{n^2}. \tag{37}$$

Proof. Recall that $\mathbf{y}^{(t)} + \mathbf{z}^{(t)} = W_t \dots W_{t_1}(\mathbf{y}^{(t_1)} + \mathbf{z}^{(t_1)})$. Then, we deterministically have:

$$\begin{aligned}&\|\mathbf{y}^{(t)} + \mathbf{z}^{(t)} - \mathbf{y}^{(0)}\|^2 = \|W_t \dots W_{t_1}(\mathbf{y}^{(t_1)} + \mathbf{z}^{(t_1)}) - \mathbf{y}^{(0)}\|^2 \\ &\leqslant 3\|W_t \dots W_{t_1}\mathbf{y}^{(t_1)} - \mathbf{y}^{(t_1)}\|^2 + 3\|\mathbf{y}^{(t_1)} - \mathbf{y}^{(0)}\|^2 + 3\|W_t \dots W_{t_1}\mathbf{z}^{(t_1)}\|^2 \\ &\leqslant 3\|W_t \dots W_{t_1}\mathbf{y}^{(t_1)} - \mathbf{y}^{(t_1)}\|^2 + 3\|\mathbf{y}^{(t_1)} - \mathbf{y}^{(0)}\|^2 + 3\|\mathbf{z}^{(t_1)}\|^2 \\ &= 3\|W_t \dots W_{t_1}\mathbf{y}^{(t_1)} - \mathbf{y}^{(t_1)}\|^2 + 3\|\mathbf{y}^{(t_1)} + \mathbf{z}^{(t_1)} - \mathbf{y}^{(0)}\|^2,\end{aligned} \tag{38}$$

where the first inequality follows from obvious manipulations, the second follows since $W_t \dots W_{t_1}$ has norm one, while the last equality follows from Pythagoras' Theorem.

The proof of Claim 2 next proceeds in the following steps:

1. Upper bound to $\|\mathbf{y}^{(t_1)} + \mathbf{z}^{(t_1)} - \mathbf{y}^{(0)}\|^2$.

Theorem 4.1 and Markov's inequality immediately imply:

$$\mathbb{P}\left[\|\mathbf{y}^{(t_1)} + \mathbf{z}^{(t_1)} - \mathbf{y}^{(0)}\|^2 \leqslant \frac{4\lambda_2 t_1}{\varepsilon n}\|\mathbf{y}^{(0)}\|^2\right] \geqslant 1 - \varepsilon. \tag{39}$$

Moreover, since $\|\mathbf{y}^{(t_1)} + \mathbf{z}^{(t_1)} - \mathbf{y}^{(0)}\|^2 = \|\mathbf{y}^{(t_1)} - \mathbf{y}^{(0)}\|^2 + \|\mathbf{z}^{(t_1)}\|^2$, we have the following useful derivations:

$$\begin{aligned}
&\|\mathbf{y}^{(t_1)} + \mathbf{z}^{(t_1)} - \mathbf{y}^{(0)}\|^2 \leqslant \frac{4\lambda_2 t_1}{\varepsilon n}\|\mathbf{y}^{(0)}\|^2 \Longrightarrow \|\mathbf{y}^{(t_1)} - \mathbf{y}^{(0)}\|^2 \leqslant \frac{4\lambda_2 t_1}{\varepsilon n}\|\mathbf{y}^{(0)}\|^2 \\
&\Longrightarrow \|\mathbf{y}^{(t_1)}\|^2 \leqslant \left(1 + \frac{4\lambda_2 t_1}{\varepsilon n}\right)\|y^{(0)}\|^2
\end{aligned} \tag{40}$$

2. Upper bound to $\|W_t \dots W_{t_1}\mathbf{y}^{(t_1)} - \mathbf{y}^{(t_1)}\|^2$, for $t \in \{t_1, \dots, 2t_1\}$.
We have:

$$\|W_t \dots W_{t_1}\mathbf{y}^{(t_1)} - \mathbf{y}^{(t_1)}\|^2 = \|W_t \dots W_1\boldsymbol{\chi} - \boldsymbol{\chi}\|^2 \frac{\|\mathbf{y}^{(t_1)}\|^2}{n}$$

Next, for any steps $\tau_1, \tau_2 \geqslant t_1$, denote by $C_{[\tau_1,\tau_2]}$ the number of cross edges sampled over the steps $\tau_1, \dots, \tau_2$. Claim 1 then implies:

$$\|W_t \dots W_{t_1}\mathbf{y}^{(t_1)} - \mathbf{y}^{(t_1)}\|^2 \leqslant 6C_{[t_1,t]}\frac{\|\mathbf{y}^{(t_1)}\|^2}{n} \tag{41}$$

The deterministic upper bound stated by (41) immediately implies:

$$\mathbb{P}\left[\|W_t \dots W_{t_1}\mathbf{y}^{(t_1)} - \mathbf{y}^{(t_1)}\|^2 > \frac{12\lambda_2 t}{n}\|\mathbf{y}^{(t_1)}\|^2\right] \leqslant \mathbb{P}\left[C_{[1,t]} > 2\lambda_2 t\right].$$

On the other hand, we immediately have $\mathbb{E}\left[C_{[1,t]}\right] \leqslant \frac{2b}{d}t = \lambda_2 t$. Application of Chernoff bound then implies:

$$\mathbb{P}\left[C_{[1,t]} > 2\lambda_2 t\right] \leqslant e^{-\frac{\lambda_2 t}{2}} \leqslant e^{-3\log n},$$

since $t \geqslant t_1 = 6\frac{n}{\lambda_3}\log n$ and $\lambda_2 \geqslant 1/n$, given the the graph is regular. As a result:

$$\mathbb{P}\left[\exists t \in \{t_1, \dots, 2t_1\} : \|W_t \dots W_{t_1}\mathbf{y}^{(t_1)} - \mathbf{y}^{(t_1)}\|^2 > \frac{12\lambda_2 t}{n}\|\mathbf{y}^{(t_1)}\|^2\right] < \frac{\log n}{n^2}. \tag{42}$$

Recalling (38) and using (39) and (42) we finally obtain:

$$\mathbb{P}\left[\exists t \in \{t_1, \dots, 2t_1\} : \|\mathbf{y}^{(t)} + \mathbf{z}^{(t)} - \mathbf{y}^{(0)}\|^2 > \frac{48\lambda_2 t}{\varepsilon n}\|\mathbf{y}^{(0)}\|^2\right] < \varepsilon + \frac{\log n}{n^2}. \tag{43}$$

In particular, from (39), (40) and (42), we know that with probability $1 - \varepsilon - \frac{\log n}{n^2}$:

$$\begin{aligned}
&\|\mathbf{y}^{(t)} + \mathbf{z}^{(t)} - \mathbf{y}^{(0)}\|^2 \leqslant 3\|W_t \dots W_{t_1}\mathbf{y}^{(t_1)} - \mathbf{y}^{(t_1)}\| + 3\|\mathbf{y}^{(t_1)} + \mathbf{z}^{(t_1)} - \mathbf{y}^{(0)}\|^2 \\
&\leqslant \frac{12\lambda_2 t_1}{\varepsilon n}\|\mathbf{y}^{(0)}\|^2 + \frac{36\lambda_2 t}{n}\|\mathbf{y}^{(t_1)}\|^2 \leqslant \frac{12\lambda_2 t}{\varepsilon n}\|\mathbf{y}^{(0)}\|^2 + \frac{36\lambda_2 t}{n}\left(1 + \frac{4\lambda_2 t_1}{\varepsilon n}\right)\|\mathbf{y}^{(0)}\|^2 \leqslant \frac{48\lambda_2 t}{\varepsilon n}\|\mathbf{y}^{(0)}\|^2.
\end{aligned}$$

The last inequality in the derivations above follows by noting that $\frac{36\lambda_2 t}{n}\left(1 + \frac{4\lambda_2 t_1}{\varepsilon n}\right) \leqslant \frac{36\lambda_2 t}{\varepsilon n}$ since $t \leqslant 2t_1 \leqslant 4n/\lambda_2$, which in turn follows if $\frac{\lambda_2}{\lambda_3} \leqslant \frac{\varepsilon^4}{96\log n}$ as we assume.

Upper bound on $|B_t|$. (8) and (37) immediately imply that, with probability at least $1 - \varepsilon - \log n/n^2$

$$\forall t \in \{t_1, \dots, 2t_1\} : |B_t| \leqslant \frac{48\lambda_2 t}{\varepsilon^3}.$$

This concludes the proof of Claim 2 . □

Wrapping up: Proof of Lemma 4.2

Let $t_1 = 6\frac{n}{\lambda_3}\log n$ and define $A_{t_1+1} := V - B_{t_1+1}$ the complement of B_{t_1+1} and, for each $t \in [t_1+1, 2t_1]$, let A_t denote the set of nodes in A_{t_1+1} that have not been averaged along a cross edge or with a node in B_t. Inductively, if $e_t = (u_t, v_t)$ is the edge chosen at time t then

$$A_t = \begin{cases} A_{t-1} & \text{if } e_{t-1} \text{ is not a cross edge and } e_{t-1} \cap B_{t-1} = \emptyset, \\ A_{t-1} \setminus \{u_t, v_t\} & \text{otherwise.} \end{cases} \tag{44}$$

We say that *we sampled a good edge at time t* in the first case, otherwise we say that *we sampled a bad edge*. The proof proceeds in two steps:

Deterministic lower bound on the number of goods nodes. We note that $|A_{2t_1}|$ is a lower bound on the number of nodes that are ε-good at every step $t \in [t_1, 2t_1]$. Indeed, every node v in A_{2t_1} was ε-good at all times between t_1 and $2t_1$, because it is a node whose value was good at time t_1+1, and then was averaged only with nodes u on the same side of the partition (that is, such that $\mathbf{sgn}(\mathbf{y}_u) = \mathbf{sgn}(\mathbf{y}_v)$) and at times in which u was good as well. Moreover, $|A_t|$ decreases by at most 2 compared to $|A_{t-1}|$ and this can only happen if we sample a bad edge. As a consequence, $|A_{2t_1}| \geqslant |A_{t_1}| - Z$, with Z the number of bad edges sampled in the interval $[t_1, 2t_1]$.

Lower bound on $|A_{2t_1}|$. Next, we condition on the event $\mathcal{B} = (\forall t \in \{t_1, \dots, 2t_1\} : |B_t| \leqslant \frac{\lambda_3 \varepsilon n}{12\log n})$. From Claim 2, this event holds with probability at least $1-\varepsilon$, whenever $\frac{\lambda_2}{\lambda_3} \leqslant \frac{\lambda_3 \varepsilon^4}{6192 \log^2 n}$. Now, observe that, for a given time step t, conditioning on the event $\mathcal{B}$ can only increase the probability of sampling a good edge. As a consequence, recalling that, without conditioning we sample edges uniformly at random, for any integer $x > 0$ we have:

$$\mathbb{P}\left[Z > x \,|\, \mathcal{B}\right] \leqslant \mathbb{P}\left[\hat{Z} > x\right],$$

where $\hat{Z}$ is the sum of independent Bernoulli variables with parameter $\frac{2b}{d} + \frac{\lambda_3\varepsilon}{12\log n}$. Here, the first term is the probability of sampling a cross edge, while the second is worst-case upper bound on the probability of sampling an edge with an endpoint in A_t and the other in B_t, provided that $|B_t| \leqslant \frac{\lambda_3 \varepsilon n}{12\log n}$. As a consequence, recalling that $\lambda_2 = 2b/d$:

$$\mathbb{E}\left[\hat{Z}\right] = \lambda_2 t_1 + \frac{\lambda_3\varepsilon}{12\log n} t_1 = 6\frac{n}{\lambda_3}\log n\left(\lambda_2 + \frac{\lambda_3\varepsilon}{12\log n}\right) = 6\frac{\lambda_2}{\lambda_3} n\log n + \frac{\varepsilon n}{2} \leqslant \varepsilon n,$$

where the last inequality obviously follows given our assumptions on λ_2/λ_3. At this point, a simple application of Chernoff bound allows to conclude that $\hat{Z}$ (and thus Z, when conditioned to $\mathcal{B}$), is at most $2\varepsilon n$ with probability $1 - e^{-\frac{\varepsilon n}{2}}$. So, conditioned to the event $\mathcal{B}$, $A_{2t_1} \geqslant n - \frac{\lambda_3 n}{12\log n} - 2\varepsilon n$ with probability $1 - e^{-\frac{\varepsilon n}{2}}$. This concludes the proof. □

C.4 Tools for the analysis of Algorithm 2

Proof of Theorem 4.3

In the next two subsections, we provide Lemmas C.5 and C.7, respectively. The two lemmas together easily imply Theorem 4.3.

Lucky nodes and community-sensitive labeling

We next consider the behavior of Algorithm SIGN-LABELING$(T, 1)$ on a graph $G = (V, E)$ where V has a sparse cut (V_1, V_2) that we wish to discover. For every node u, let $p_u^{sign}(T)$ be the probability that node u sets $\mathbf{h}_u^{sign}$ differently from the sign of the average of the initial values of the nodes in its own community

$$p_u^{sign}(T) = \begin{cases} \mathbb{P}\left[\mathbf{h}_u^{sign} \neq \mathbf{sgn}\left(\frac{2}{n}\sum_{v\in V_1} \mathbf{x}_v\right)\right] & \text{if } u \in V_1, \\ \mathbb{P}\left[\mathbf{h}_u^{sign} \neq \mathbf{sgn}\left(\frac{2}{n}\sum_{v\in V_2} \mathbf{x}_v\right)\right] & \text{if } u \in V_2, \end{cases}$$

where the randomness is over the initial choice of $\mathbf{x} \sim \{-1, 1\}^n$ and over the execution of Algorithm SIGN-LABELING$(T, 1)$. Notice that, for a given graph G and partition of the nodes V_1, V_2, $p_u^{sign}(T)$ depends on the node u, on the protocol (*sign*), and on the number of activations T after which the node sets $\mathbf{h}_u^{\text{sign}}$ (we will omit parameter T from p_u^{sign} when clear from context).

To understand the point of this definition, consider the extreme case in which the cut (V_1, V_2) is empty, while V_1 and V_2 induce connected graphs. In this case, the averaging process AVERAGING will converge to a global state in which all nodes in V_1 have a local state close to the average $\frac{2}{n}\sum_{v\in V_1} \mathbf{x}_v$ and all nodes in V_2 have a state close to $\frac{2}{n}\sum_{v\in V_2} \mathbf{x}_v$. We call τ_u the (random) step in which the node u sets its $\mathbf{h}_u^{\text{sign}}$ to the sign of the state of u, i.e., the global round when u achieves T activations. If T is chosen so that τ_u is large enough, we would expect $\mathbf{h}_u^{\text{sign}}$ to agree with the sign of $\frac{2}{n}\sum_{v\in V_1} \mathbf{x}_v$ if $u \in V_1$ and with the sign of $\frac{2}{n}\sum_{v\in V_2} \mathbf{x}_v$ if $u \in V_2$, with p_u^{sign} small for all u. It seems reasonable that a possibly weakened version of the considerations above should apply to graphs exhibiting a sparse, rather than empty, cut, provided the subgraphs induced by V_1 and V_2 are good expanders. To quantitavely capture this intuition, we introduce the notion of a *(un)lucky* node.

Definition C.1 (Unlucky nodes)**.** *We say that node u is ε-*unlucky *if p_u^{sign} is larger than ε. We thus define the set of ε-unlucky nodes as follows*

$$U^{\varepsilon,sign}_{G,(V_1,V_2)} = \left\{u \,|\, p_u^{sign} \geqslant \varepsilon\right\} .$$

We write $U^{\varepsilon,sign}$ in place of $U^{\varepsilon,sign}_{G,(V_1,V_2)}$, when the underlying graph and partition of the nodes are clear from the context.

Lemma C.5. *Let $G = (V, E)$ be a graph, V_1, V_2 be a partition of V and fix $\varepsilon \in (0, \frac{1}{12}]$. Then,* SIGN-LABELING $(T, 10\varepsilon^{-1} \log n)$ *performs a community-sensitive labeling of G according to Definition 2.4, with $c_1 = 4\varepsilon$, $c_2 = 1/6$ and $\gamma = |U^{\varepsilon,T}|/n$.*

Proof. The proof of the theorem relies on the mutual independence among the components of any label $\mathbf{h}(\cdot)$ and some standard arguments. We remark that the independence crucially depends on the fact that SIGN-LABELING (T,m) updates one component per interaction: the evolution of $\mathbf{x}(j_1)$ and $\mathbf{x}(j_2)$ depends solely on the respective initial vector values (which are independent), and on the sequence of sampled edges which update component j_1 and j_2 (which are independent conditional on their number).

Call $\ell = \frac{10 \log n}{\varepsilon}$ and denote $\mathbf{h}_{V_i} := (h_{V_i}(1), \ldots, h_{V_i}(\ell))$ where $h_{V_i}(j) := \mathbf{sgn}\left(\sum_{v\in V_i} x_v\right)$.

We first claim that w.h.p. for every vertex $u \in V_1 \setminus U^\varepsilon$, $\Delta(\mathbf{h}_u, \mathbf{h}_{V_1}) \leqslant 2\varepsilon\ell$. Observe that by definition of U^ε,

$$\mathbb{E}[\Delta(\mathbf{h}_u, \mathbf{h}_{V_1})] = p(u)\ell \leqslant \varepsilon\ell.$$

Since the ℓ components are mutually independent, the Chernoff bound [DP09] implies that $\Delta(\mathbf{h}_u, \mathbf{h}_{V_1}) \leqslant 2\varepsilon\ell$, w.h.p. A union bound over vertices in $V_1 \setminus U^\varepsilon$ implies the claim.

Henceforth, let us assume that $\Delta(\mathbf{h}_u, \mathbf{h}_{V_1}) \leqslant 2\varepsilon\ell$ for each $u \in V_1 \setminus U^\varepsilon$ and a similar claim for all vertices $v \in V_2 \setminus U^\varepsilon$.

As for Case (i), w.l.o.g. let us consider $u, v \in V_1 \setminus U$. By triangle inequality, we get the desired claim.

$$\Delta(\mathbf{h}_u, \mathbf{h}_v) \leqslant \Delta(\mathbf{h}_u, \mathbf{h}_{V_1}) + \Delta(\mathbf{h}_{V_1}, \mathbf{h}_v) \leqslant 4\varepsilon\ell.$$

As for Case (ii), since the initial values of $x_u(j)$ $(u \in V_1 \cup V_2)$ are chosen independently and uniformly at random in $\{-1, 1\}$, simple symmetry arguments show that the probability of the event "$\mathbf{sgn}(\sum_{u \in V_1} x_u) = \mathbf{sgn}(\sum_{u \in V_2} x_u)$" is 1/2. Hence, $\mathbb{E}[\Delta(\mathbf{h}_{V_1}, \mathbf{h}_{V_2})] = \ell/2$ and from Chernoff bounds we get that

$$\Delta(\mathbf{h}_{V_1}, \mathbf{h}_{V_2}) \geqslant \frac{\ell}{3}, \tag{45}$$

with all but a probability exponentially small in ℓ. Henceforth, let us condition on the event that $\Delta(\mathbf{h}_{V_1}, \mathbf{h}_{V_2}) \leqslant \frac{\ell}{3}$

Consider $u \in V_1$ and $v \in V_2$. By triangle inequality, we have that

$$\Delta(\mathbf{h}_u, \mathbf{h}_v) \geqslant \Delta(\mathbf{h}_{V_1}, \mathbf{h}_{V_2}) - \Delta(\mathbf{h}_u, \mathbf{h}_{V_1}) - \Delta(\mathbf{h}_{V_2}, \mathbf{h}_v) \tag{46}$$

$$\geqslant \frac{\ell}{3} - 2\varepsilon\ell - 2\varepsilon\ell \geqslant \frac{\ell}{6}, \tag{47}$$

concluding the proof.

□

C.4.1 A bound on the number of unlucky nodes

In Lemma C.7 we give an upper bound on the number of ε-unlucky nodes. This is the second key step toward proving Theorem 4.3.

We first prove the following technical lemma on the range in which τ_v falls w.h.p.

Lemma C.6. *If $T > 72 \log n$ and $t_1 = 3Tn/4$ then*

$$\mathbb{P}\left[\{\tau_v \mid v \in V\} \subseteq [t_1, 2t_1]\right] \geqslant 1 - \frac{1}{n}.$$

Proof. For each node v, let $X_v^{(i)} = \mathbf{1}_{[v \text{ is activated at round } i]}$. Fix a node v. By applying the Chernoff bound on the i.i.d. random variables $\{X_v^{(i)}\}_{i \geqslant 0}$, we have

$$\mathbb{P}\left[\sum_{i=1}^{3Tn/4} X_v^{(i)} \geqslant T\right] \leqslant e^{-\frac{T}{36}} \quad \text{and} \quad \mathbb{P}\left[\sum_{i=1}^{3Tn/2} X_v^{(i)} \leqslant T\right] \leqslant e^{-\frac{T}{12}}.$$

The claim follows by applying a union bound over the nodes. □

Lemma C.7 (Number of unlucky nodes)**.** *Let $\varepsilon > 0$ be an arbitrarily small value and let G be an (n, d, b)-clustered regular graph with $\frac{\lambda_2}{\lambda_3} \leqslant \frac{\lambda_3 \varepsilon^4}{c \log^2 n}$, for a large enough costant c. If $T = \frac{8}{\lambda_3} \log n$ then the number of $\sqrt{\varepsilon}$-unlucky nodes is*

$$\left|U^{\sqrt{\varepsilon}, sign}\right| \leqslant 6\sqrt{\varepsilon}\, n.$$

Proof. Let L^{sign} be the set of nodes that freeze their sign $\mathbf{h}_v^{sign}$ according to the sign of $\mathbf{x}_{\|,v}+\mathbf{y}_v^{(0)}$,

$$L^{sign} = \left\{ v \in V_1 \cup V_2 \, : \, \mathbf{sgn}\left(\mathbf{x}_v^{\tau_v}\right) = \mathbf{sgn}\left(\mathbf{x}_{\|,v} + \mathbf{y}_v^{(0)}\right) \right\}.$$

We first observe that, given any $\varepsilon > 0$, if we have a lower bound on the expected size of L^{sign}, namely $\mathbb{E}\left[|L^{sign}|\right] \geqslant n - \varepsilon n$, then we have an upper bound on the number of $\sqrt{\varepsilon}$-unlucky nodes, namely $|U^{\sqrt{\varepsilon},sign}| \leqslant 6\sqrt{\varepsilon}n$. Indeed,

$$\begin{aligned}
\mathbb{E}\left[|L^{sign}|\right] &= \sum_{u \in U^{\sqrt{\varepsilon},sign}} \mathbb{P}\left[u \in L^{sign}\right] + \sum_{u \notin U^{\sqrt{\varepsilon},sign}} \mathbb{P}\left[u \in L^{sign}\right] \\
&\leqslant (1 - \sqrt{\varepsilon}) \left|U^{\sqrt{\varepsilon},sign}\right| + n - \left|U^{\sqrt{\varepsilon},sign}\right| \\
&= n - \sqrt{\varepsilon} \left|U^{\sqrt{\varepsilon},sign}\right|. \qquad (48)
\end{aligned}$$

We now give a lower bound on the expected size of L^{sign}. Let Γ be the event

$$\Gamma = \text{“}|\mathbf{y}_v^{(0)} + \mathbf{x}_{\|,v}| \geqslant \frac{\varepsilon}{\sqrt{n}}\text{”}.$$

Recall that τ_v denotes the time at which node v freezes its value of $\mathbf{h}_v^{sign}$. Notice that the value $|\mathbf{y}_v^{(0)} + \mathbf{x}_{\|,v}|$ does not depend on the node v, only on the initial assignment. Hence, for any node $u \in V_1 \cup V_2$ we have that

$$\begin{aligned}
\mathbb{P}\left[u \in L^{sign}\right] &\geqslant \mathbb{P}\left[\Gamma \wedge \{u \text{ is } \varepsilon\text{-good at round } \tau_u\}\right] \\
&\geqslant \mathbb{P}\left[\Gamma \wedge \{u \text{ is } \varepsilon\text{-good at all rounds } t \in [t_1, 2t_1]\} \wedge \{\tau_u \in [t_1, 2t_1]\}\right] \\
&\geqslant 1 - \mathbb{P}\left[\overline{\Gamma}\right] - \mathbb{P}\left[\{u \text{ is } \textit{not } \varepsilon\text{-good at } \textit{some} \text{ round } t \in [t_1, 2t_1]\}\right] - \mathbb{P}\left[\tau_u \notin [t_1, 2t_1]\right] \\
&= \mathbb{P}\left[u \text{ is } \varepsilon\text{-good at all rounds } t \in [t_1, 2t_1]\right] - \mathbb{P}\left[\overline{\Gamma}\right] - \mathbb{P}\left[\tau_u \notin [t_1, 2t_1]\right]. \qquad (49)
\end{aligned}$$

From Lemmas A.1 and C.6 it follows that $\mathbb{P}\left[\overline{\Gamma}\right] \leqslant \mathcal{O}(\varepsilon)$ and that $\mathbb{P}\left[\tau_u \notin [t_1, 2t_1]\right] \leqslant 1/n$. Hence, from (49) the expected size of L^{sign} is

$$\begin{aligned}
\mathbb{E}\left[|L^{sign}|\right] &= \sum_u \mathbb{P}\left[u \in L^{sign}\right] \\
&\geqslant \mathbb{E}[|\{u \, : \, u \text{ is } \varepsilon\text{-good at all rounds } t \in [t_1, 2t_1]\}|] - \frac{4}{\sqrt{2\pi}}\varepsilon n - 1. \qquad (50)
\end{aligned}$$

Finally, from Lemma 4.2 we have

$$\mathbb{E}[|\{u \, : \, u \text{ is } \varepsilon\text{-good at all rounds } t \in [t_1, 2t_1]\}|] \geqslant (1 - 3\varepsilon)(1 - \varepsilon)n.$$

Thus from (50) and the previous inequality we get that

$$\mathbb{E}\left[|L^{sign}|\right] \geqslant (1 - 3\varepsilon)(1 - \varepsilon)n - \frac{4}{\sqrt{2\pi}}\varepsilon n - 1 > (1 - 6\varepsilon)n$$

and the thesis follows from (48). □

D Omitted Proofs from Section 5

The main goal of this subsection is to prove Theorem 5.1. For the sake of convenience, we rewrite the state of the averaging process as

$$\mathbf{x}^{(t)} = a_{||} \cdot \left(\mathbf{1}/\sqrt{n}\right) + a_y(t) \cdot \left(\chi/\sqrt{n}\right) + \mathbf{z}^{(t)}$$

where $a_{||}, a_y(t) \in \mathbb{R}$ and $\mathbf{z}^{(t)}$ is orthogonal to both χ and $\mathbf{1}$. Recall also that $\mathbf{y}^{(t)} = a_y(t) \cdot (\chi/\sqrt{n})$ and that $a_{||}$ remains unchanged throughout the algorithm. Suppose now we fix a starting vector $\mathbf{x}^{(0)}$, then we can exactly compute the expectation of $a_y(t)$ as stated more formally below.

Observation D.1. *For all $t \in \mathbb{Z}_{\geqslant 0}$, we have $\mathbb{E}_{\mathcal{E}}[a_y(t)] = \left(1 - \frac{2\delta\lambda_2}{n}\right)^t a_y(0)$.*

Proof of Observation D.1. To prove the above statement, it is enough to show that $\mathbb{E}_{(u_t,v_t)}[a_y(t)] = \left(1 - \frac{2\delta\lambda_2}{n}\right) a_y(t-1)$ for every $t \in \mathbb{N}$. Indeed, $\mathbb{E}_{(u_t,v_t)}[a_y(t)]$ can be rewritten as follows.

$$\mathop{\mathbb{E}}_{W \sim \mathcal{W}}\left[\frac{\chi^T W \mathbf{x}^{(t-1)}}{\sqrt{n}}\right] = \frac{\chi^T \bar{W} \mathbf{x}^{(t-1)}}{\sqrt{n}} = \frac{\left(1 - \frac{2\delta\lambda_2}{n}\right)\chi^T \mathbf{x}^{(t-1)}}{\sqrt{n}} = \left(1 - \frac{2\delta\lambda_2}{n}\right) a_y(t-1),$$

concluding the proof. $\square$

Let $\mu(t) \triangleq \mathbb{E}[a_y(t)] = \left(1 - \frac{4\delta b}{dn}\right)^t a_y(0)$ be the expectation of $a_y(t)$. We will show that, if we start with $\mathbf{x}^{(0)}$ such that $\|\mathbf{z}^{(0)}\|^2$ is not too much larger than $n\|\mathbf{y}^{(0)}\|^2$, then $a_y(t)$ concentrates around $\mu(t)$, as long as $t \leqslant O_{b,d,\delta}(n^2)$. Moreover, we will also show that, for $t \geqslant \Omega_{b,d,\delta}(n \log n)$, $\|\mathbf{z}^{(t)}\|$ becomes small compared to $\mu(t)$. This is stated more precisely below.

Theorem D.1. *Let β be any real number such that $1 \leqslant \beta \leqslant \frac{d}{\varepsilon b}$. For any initial vector $\mathbf{x}^{(0)}$ that satisfies satisfies $\|\mathbf{z}^{(0)}\|^2 \leqslant n\beta\|\mathbf{y}(0)\|^2$ and for any $t \in \left[\frac{8n}{\delta(\lambda_3-\lambda_2)}\log\left(\frac{nd\beta}{\varepsilon b}\right), \frac{n^2\beta}{128\delta(\lambda_3-\lambda_2)}\right]$, we have*

$$\mathop{\mathbb{P}}_{\mathcal{E}}\left[0.5\mu(t) \leqslant a_y(t) \leqslant 1.5\mu(t)\right] \geqslant 1 - O(\varepsilon\beta b/d) \tag{51}$$

and

$$\mathop{\mathbb{P}}_{\mathcal{E}}\left[\|\mathbf{z}^{(t)}\| \leqslant \left(0.5\sqrt[4]{\varepsilon b/d}\right)\mu(t)\right] \geqslant 1 - O(\sqrt{\varepsilon b/d}). \tag{52}$$

We defer the proof of Theorem D.1 to Subsection D.1. For now, let us turn our attention back to the proof of Theorem 5.1. To go from here to Theorem 5.1, we will also need to upper bound the probability that $\|\mathbf{z}^{(0)}\|^2 > n\beta\|\mathbf{y}^{(0)}\|^2$. More specifically, when $\mathbf{x}^{(0)}$ is a random ± 1 vector, we have the following bound.

Proposition 1. *For any $\beta > 0$, we have $\mathbb{P}_{\mathbf{x}^{(0)}\sim\{\pm1\}^n}\left[\|\mathbf{z}^{(0)}\|^2 > n\beta\|\mathbf{y}^{(0)}\|^2\right] \leqslant O(1/\sqrt{\beta} + 1/\sqrt{n})$.*

This proposition was also essentially proved in [BCN+17]; we repeat the proof from [BCN+17] below for completeness.

Proof of Proposition 1. First, note that $\|\mathbf{z}(0)\|^2 \leqslant \|\mathbf{x}(0)\|^2 = n$. Hence, it suffices to upper bound the probability that $\|\mathbf{y}(0)\|^2$ is less than $1/\beta$. Since $\|\mathbf{y}(0)\| = \left|\frac{\chi^T \mathbf{x}(0)}{\sqrt{n}}\right|$, the probability that $\|\mathbf{y}(0)\|^2 < 1/\beta$ is exactly equal to the probability that a sum of n i.i.d. Rademacher random variables lie in $[-\sqrt{n/\beta}, \sqrt{n/\beta}]$. The latter probability is exactly equal to

$$\frac{1}{2^n} \sum_{i=\left(n-\lfloor\sqrt{n/\beta}\rfloor\right)/2}^{\left(n+\lfloor\sqrt{n/\beta}\rfloor\right)/2} \binom{n}{i} \leqslant \left(\sqrt{n/\beta}+1\right) \frac{\binom{n}{n/2}}{2^n} \leqslant O(1/\sqrt{\beta}+1/\sqrt{n}),$$

where the second inequality comes from a well-known fact that $\binom{n}{n/2} = O(2^n/\sqrt{n})$. □

By combining Theorem D.1 and Proposition 1, we immediately get Theorem 5.1.

Proof of Theorem 5.1. choosing $\beta = \left(\frac{d}{\varepsilon b}\right)^{2/3}$, we can upper bound $\mathbb{P}_{\mathbf{x}^{(0)},\mathcal{E}}\left[\|\mathbf{z}^{(t)}\|^2 \leqslant \sqrt{\varepsilon b/d}\|\mathbf{y}^{(t)}\|^2\right]$ by

$$\mathbb{P}_{\mathcal{E}}\left[\|\mathbf{z}^{(t)}\|^2 \leqslant \sqrt{\varepsilon b/d}\|\mathbf{y}^{(t)}\|^2 \;\middle|\; \|\mathbf{z}^{(0)}\|^2 \leqslant n\beta\|\mathbf{y}^{(0)}\|^2\right] + \mathbb{P}_{\mathbf{x}^{(0)}\sim\{\pm 1\}^n}\left[\|\mathbf{z}^{(0)}\|^2 > n\beta\|\mathbf{y}^{(0)}\|^2\right].$$

Then, from Theorem D.1, the first term is at most $O(\varepsilon\beta b/d) + O(\sqrt{\varepsilon b/d}) = O(\sqrt[3]{\varepsilon b/d})$. Moreover, from Proposition 1, the second term is also at most $O(1/\sqrt{\beta}+1/\sqrt{n}) = O(\sqrt[3]{\varepsilon b/d} + 1/\sqrt{n})$. □ □

D.1 Proof of Theorem D.1

D.1.1 Evolution of State in One Time Step

The first step in proving Theorem D.1 is to understand what happens in a single update. Specifically, we would like to understand how $\|\mathbf{y}^{(t)}\|$ and $\|\mathbf{z}^{(t)}\|$ behave, given $\|\mathbf{y}^{(t-1)}\|$ and $\|\mathbf{z}^{(t-1)}\|$. To this end, we prove the following lemma, which gives bounds on expectations of $\|\mathbf{y}^{(t)}\|^2$ and $\|\mathbf{z}^{(t)}\|^2$ based on $\|\mathbf{y}^{(t-1)}\|^2$ and $\|\mathbf{z}^{(t-1)}\|^2$.

Lemma D.2. *Let G be as above. Let $\mathbf{y}$ be a vector parallel to χ and $\mathbf{z}$ be a vector orthogonal to $\mathbf{y}$ and to $\mathbf{1}$. Let P_χ be the projection matrix on χ, that is, $P_\chi = \frac{1}{n}\chi\chi^T$ and let $P_\perp$ be the the projection matrix on the space orthogonal χ, that is, $P_\perp = I - \frac{1}{n}\chi\chi^T$. Moreover, let $\mathbf{y}' = P_\chi W(y+z)$ and $\mathbf{z}' = P_\perp W(y+z)$ where W is randomly selected according to $\mathcal{W}$. Then,*

$$\mathbb{E}_W[\|\mathbf{y}'\|^2] \leqslant \left(1 - \frac{8\delta b}{dn} + \frac{16\delta^2 b}{dn^2}\right)\|\mathbf{y}\|^2 + \left(\frac{16\delta^2 b}{dn^2}\right)\|\mathbf{z}\|^2$$

and

$$\mathbb{E}_W[\|\mathbf{z}'\|^2] \leqslant \left(\frac{8\delta^2 b}{dn}\right)\|\mathbf{y}\|^2 + \left(1 - \frac{4\delta(1-\delta)\lambda_3}{n}\right)\|\mathbf{z}\|^2.$$

We note here that Lemma D.2 is simply a restatement of Lemma 5.2.

Proof. Note that $\mathbf{y}'$ and $\mathbf{z}'$ are orthogonal. We will estimate the expectation of $\|\mathbf{y}'+\mathbf{z}'\|^2$ and of $\|\mathbf{y}'\|^2$, and the we will use Pythagoras's theorem to deduce a bound on $\|\mathbf{z}'\|^2$.

To estimate the expected norm squared of $\mathbf{y}'+\mathbf{z}'$ we see that

$$\begin{aligned}\mathbb{E}_W[\|\mathbf{y}'+\mathbf{z}'\|^2] &= \mathbb{E}_W[\|W(\mathbf{y}+\mathbf{z})\|^2] \\ &= \mathbb{E}_W[(\mathbf{y}+\mathbf{z})^T W^T W(\mathbf{y}+\mathbf{z})]\end{aligned}$$

$$
\begin{aligned}
&= \mathop{\mathbb{E}}_{W}[(\mathbf{y}+\mathbf{z})^T W^2 (\mathbf{y}+\mathbf{z})] \\
&= \mathop{\mathbb{E}}_{W}[(\mathbf{y}+\mathbf{z})^T ((2\delta-1)I + 2(1-\delta)W)(\mathbf{y}+\mathbf{z})] \\
&= (2\delta-1)\left(\|\mathbf{y}\|^2 + \|\mathbf{z}\|^2\right) + 2(1-\delta)(\mathbf{y}+\mathbf{z})^T \bar{W}(\mathbf{y}+\mathbf{z}) \\
&= (2\delta-1)\left(\|\mathbf{y}\|^2 + \|\mathbf{z}\|^2\right) + 2(1-\delta)\left(\mathbf{y}^T\bar{W}\mathbf{y} + 2\mathbf{z}^T\bar{W}\mathbf{y} + \mathbf{z}^T\bar{W}\mathbf{z}\right) \\
&= (2\delta-1)\left(\|\mathbf{y}\|^2 + \|\mathbf{z}\|^2\right) + 2(1-\delta)\left(\left(1-\frac{4\delta b}{dn}\right)\|\mathbf{y}\|^2 + \mathbf{z}^T\bar{W}\mathbf{z}\right),
\end{aligned}
$$

where $\bar{W} := \mathbb{E}[W]$, and in (a) we used that $\bar{W}\mathbf{y} = \left(1 - \frac{4\delta b}{dn}\right)\mathbf{y}$.

Moreover, note that $\mathbf{z}$ can be written as a linear combination of eigenvectors of $\bar{W}$ whose eigenvalues are at most $1 - \frac{2\delta\lambda_3}{n}$; This implies that $\mathbf{z}^T\bar{W}\mathbf{z} \leqslant \left(1 - \frac{2\delta\lambda_3}{n}\right)\|\mathbf{z}\|^2$. Plugging this inequality into the above equality, we have

$$
\mathop{\mathbb{E}}_{W}[\|\mathbf{y}' + \mathbf{z}'\|^2] \leqslant \left(1 - \frac{8\delta(1-\delta)b}{dn}\right)\|\mathbf{y}\|^2 + \left(1 - \frac{4\delta(1-\delta)\lambda_3}{n}\right)\|\mathbf{z}\|^2. \tag{53}
$$

Now we estimate the expected squared norm of $\mathbf{y}'$. Observe that

$$
\begin{aligned}
\mathop{\mathbb{E}}_{W}[\|\mathbf{y}'\|^2] &= \mathop{\mathbb{E}}_{W}[\|P_\chi W(\mathbf{y}+\mathbf{z})\|^2] \\
&= \mathop{\mathbb{E}}_{W}\ [(\mathbf{y}+\mathbf{z})^T W^T P_\chi^T P_\chi W (\mathbf{y}+\mathbf{z})] \\
&= \mathop{\mathbb{E}}_{W}\ [(\mathbf{y}+\mathbf{z})^T W^T P_\chi W (\mathbf{y}+\mathbf{z})].
\end{aligned}
$$

Let us consider two cases: whether the edge (u,v) defining W is an *internal* edge, that is, an edge whose endpoints are on the same side of the partition, or it is a *cross* edge having endpoints on different sides of the partition.

1. If (u,v) is an internal edge, which happens with probability $1 - \frac{b}{d}$, then $W\chi = \chi$, and so $WP_\chi W = \frac{1}{n}W\chi\chi^T W = \frac{1}{n}\chi\chi^T = P_\chi$. This implies that $(\mathbf{y}+\mathbf{z})^T W^T P_\chi W(\mathbf{y}+\mathbf{z}) = \|\mathbf{y}\|^2$.

2. If (u,v) is a cross edge such that $u \in V_1$ and $v \in V_2$, recall that $W = I - \delta\mathbf{e}_{u,v}\mathbf{e}_{u,v}^T$. To bound $\|P_\chi W(\mathbf{y}+\mathbf{z})\|^2$, first observe that

$$
P_\chi W\chi = P_\chi(\chi - \delta\mathbf{e}_{u,v}(\mathbf{e}_{u,v}^T\chi)) = \frac{\chi}{n}(\chi^T\chi - \delta\|\mathbf{e}_{u,v}^T\chi\|^2) = \chi - \frac{4\delta}{n}\chi.
$$

   Hence, we have

$$
P_\chi W\mathbf{y} = \left(1 - \frac{4\delta}{n}\right)\mathbf{y}. \tag{54}
$$

   Now, let us consider $P_\chi W\mathbf{z}$. Observe that

$$
P_\chi(W\mathbf{z}) = P_\chi(\mathbf{z} - \delta\mathbf{e}_{u,v}(\mathbf{e}_{u,v}^T\mathbf{z})) = P_\chi(\mathbf{z} - \delta(z_u - z_v)\mathbf{e}_{u,v}) = -\frac{2\delta(z_u - z_v)}{n}\chi. \tag{55}
$$

   By combining (54) and (55), we have

$$
\|P_\chi W(\mathbf{y}+\mathbf{z})\|^2 = \left(1 - \frac{4\delta}{n}\right)^2\|\mathbf{y}\|^2 - 2\left(\frac{2\delta(z_u - z_v)}{\sqrt{n}}\right)\left(1 - \frac{4\delta}{n}\right)\|\mathbf{y}\| + \frac{4\delta^2(z_u - z_v)^2}{n}.
$$

Now, if we take the expectation over cross edges (u, v), the second term becomes zero, because both z_u and z_v average to zero for a random cross edge (the margninal of u is uniform over V_1 and the marginal of v is uniform over V_2). Moreover, we have

$$\mathop{\mathbb{E}}_{(u,v)\text{ crossedge}}[(z_u - z_v)^2] \leqslant \mathop{\mathbb{E}}_{(u,v)\text{crossedge}}[2(z_u^2 + z_v^2)] = \mathop{\mathbb{E}}_{u \in V}[4z_u^2] = \frac{4}{n}\|\mathbf{z}\|^2.$$

where the first equality follows from the fact that each vertex has b cross edges.

Thus, in this case, we have

$$\left(1 - \frac{4\delta}{n}\right)^2 \|\mathbf{y}\|^2 \leqslant \|P_\chi W(\mathbf{y} + \mathbf{z})\|^2 \leqslant \left(1 - \frac{4\delta}{n}\right)^2 \|\mathbf{y}\|^2 + \frac{16\delta^2}{n^2}\|\mathbf{z}\|^2.$$

Putting the two cases together, we arrive at the following inequality.

$$\left(1 - \frac{8\delta b}{dn} + \frac{16\delta^2 b}{dn^2}\right)\|\mathbf{y}\|^2 \leqslant \mathop{\mathbb{E}}_{W}[\|\mathbf{y}'\|^2] \leqslant \left(1 - \frac{8\delta b}{dn} + \frac{16\delta^2 b}{dn^2}\right)\|\mathbf{y}\|^2 + \left(\frac{16\delta^2 b}{dn^2}\right)\|\mathbf{z}\|^2. \quad (56)$$

Finally, note that the upper bound in (56) is already the desired upper bound for $\mathbb{E}_W[\|\mathbf{y}'\|^2]$ and that the lower bound in (56) together with (53) implies the desired upper bound on $\mathbb{E}_W[\|\mathbf{z}'\|^2]$. □

D.1.2 From Evolution of State to Bounds on $\mathbb{E}_\mathcal{E}[\|\mathbf{y}^{(t)}\|^2]$ and $\mathbb{E}_\mathcal{E}[\|\mathbf{z}^{(t)}\|^2]$

We next turn the bounds from Lemma D.2 to bounds on $\mathbb{E}_\mathcal{E}[\|\mathbf{y}^{(t)}\|^2]$ and $\mathbb{E}_\mathcal{E}[\|\mathbf{z}^{(t)}\|^2]$ based only on $\|\mathbf{y}(0)\|^2, \|\mathbf{z}(0)\|^2, \delta$ and the parameters of our graph. This bound will indeed be enough for us to prove certain concentrations of $\|\mathbf{y}^{(t)}\|$ and $\|\mathbf{z}^{(t)}\|$, which are at the heart of the analysis. Before we state the bounds on $\mathbb{E}_\mathcal{E}[\|\mathbf{y}^{(t)}\|^2]$ and $\mathbb{E}_\mathcal{E}[\|\mathbf{z}^{(t)}\|^2]$, let us define the following shorthands for some expressions that will appear regularly throughout the rest of the section.

- Let $\xi \triangleq \left(1 - \frac{4\delta b}{dn}\right)^2, \xi_1 \triangleq \left(1 - \frac{8\delta b}{dn} + \frac{336\delta^2 b}{dn^2}\right), \xi_2 \triangleq \left(1 - \frac{4\delta(1-\delta)\lambda_3}{n}\right)$. Note that $\mu(t) = \xi^{t/2} a_y(0)$.
- Let $\kappa \triangleq 1 + (40\varepsilon b/d)\beta$. Recall that β is a parameter in Theorem D.1 which satisfies $\beta \geqslant \frac{\|\mathbf{z}(0)\|^2}{n\|\mathbf{y}(0)\|^2}$.

We can now state our bounds on $\mathbb{E}_\mathcal{E}[\|\mathbf{y}^{(t)}\|^2]$ and $\mathbb{E}_\mathcal{E}[\|\mathbf{z}^{(t)}\|^2]$:

Lemma D.3. *For any $t \in \mathbb{Z}_{\geqslant 0}$, we have*

$$\mathop{\mathbb{E}}_{\mathcal{E}}[\|\mathbf{y}^{(t)}\|^2] \leqslant (\kappa\xi_1^t)\|\mathbf{y}(0)\|^2 \qquad \textit{and} \qquad \mathop{\mathbb{E}}_{\mathcal{E}}[\|\mathbf{z}^{(t)}\|^2] \leqslant \left((20\varepsilon b/d)\kappa\xi_1^t + \beta n \xi_2^t\right)\|\mathbf{y}(0)\|^2.$$

We defer the proof of Lemma D.3, which is essentially solving the recurrence relation from Lemma D.2, to Subsection D.1.5. Let us now proceed to use this lemma to derive concentrations of $\|\mathbf{y}^{(t)}\|, \|\mathbf{z}^{(t)}\|$.

D.1.3 Concentrations of $\|\mathbf{y}^{(t)}\|$ and $\|\mathbf{z}^{(t)}\|$

A direction application of Markov's inequality to the bound on $\mathbb{E}_\mathcal{E}[\|\mathbf{z}^{(t)}\|^2]$ from Lemma D.3 gives us the desired concentration for $\|\mathbf{z}^{(t)}\|$:

Lemma D.4. *For every* $t \in \mathbb{Z}_{\geqslant 0}$, $\mathbb{P}_{\mathcal{E}}\left[\|\mathbf{z}^{(t)}\|^2 \geqslant 0.25\sqrt{\varepsilon b/d}(\mu(t))^2\right] \leqslant 80\sqrt{\varepsilon b/d}\kappa(\xi_1/\xi)^t + \frac{4\beta n}{\sqrt{\varepsilon b/d}}(\xi_2/\xi)^t$.

For $\|\mathbf{y}^{(t)}\|$, since we know that $\|\mathbf{y}^{(t)}\|^2$ is simply $a_y(t)^2$ and we also know $\mu(t) = \mathbb{E}_{\mathcal{E}}[a_y(t)]$, we can apply Cherbychev's inequality on $a_y(t)$, which results in the following lemma.

Lemma D.5. *For every* $t \in \mathbb{Z}_{\geqslant 0}$, $\mathbb{P}_{\mathcal{E}}[a_y(t) \notin (0.5\mu(t), 1.5\mu(t))] \leqslant 4\left(\kappa(\xi_1/\xi)^t - 1\right)$.

Proof. Recall from Observation D.1 that $\mathbb{E}_{\mathcal{E}}[a_y(t)] = \mu(t) = \xi^{t/2}a_y(0)$. Moreover, from Lemma D.3, we have $\mathbb{E}_{\mathcal{E}}[a_y(t)^2] = \mathbb{E}_{\mathcal{E}}[\|\mathbf{y}^{(t)}\|^2] \leqslant (\kappa\xi_1^t)(a_{y^{(0)}})^2$. Hence, from Chebyshev's inequality, we have

$$\mathbb{P}_{\mathcal{E}}[a_y(t) \notin (0.5\mu(t), 1.5\mu(t))] \leqslant \frac{\mathbb{E}_{\mathcal{E}}[a_y(t)^2] - (\mu(t))^2}{(0.5\mu(t))^2} = 4\left(\kappa(\xi_1/\xi)^t - 1\right)$$

as desired. ∎

D.1.4 Putting things together

Finally, we will now prove Theorem D.1 by plugging in the appropriate value for variables in Lemma D.4 and Lemma D.5. To this end, let us first state a couple of inequalities that will be useful.

Lemma D.6. *If* $\beta \leqslant \frac{d}{\varepsilon b}$, *then, for any* $t \leqslant \frac{n^2\beta}{1344\delta(\lambda_3-\lambda_2)}$, *we have* $\kappa(\xi_1/\xi)^t \leqslant 1 + 81(\varepsilon b\beta/d)$.

Lemma D.7. *If* $\beta \leqslant \frac{d}{\varepsilon b}$, *then, for any* $t \geqslant \frac{8n}{\delta(\lambda_3-\lambda_2)}\log\left(\frac{nd\beta}{\varepsilon b}\right)$, *we have* $\frac{4\beta n}{\sqrt{\varepsilon b/d}}(\xi_2/\xi)^t \leqslant 4\sqrt{\varepsilon b/d}$.

We defer the proofs of both lemmas, which are basically calculations, to Appendix D.1.6. Let us now proceed to prove Theorem D.1.

Proof of Theorem D.1. Let t be any positive integer such that $\frac{8n}{\delta(\lambda_3-\lambda_2)}\log\left(\frac{nd}{\varepsilon b}\right) \leqslant t \leqslant \frac{n^2\beta}{1344\delta(\lambda_3-\lambda_2)}$. From Lemma D.5 and Lemma D.6, we have $\mathbb{P}_{\mathcal{E}}[a_y(t) \notin (0.5\mu(t), 1.5\mu(t))] \leqslant O(\varepsilon b\beta/d)$.

Moreover, from Lemma D.4, Lemma D.6 and Lemma D.7, we have

$$\mathbb{P}_{\mathcal{E}}\left[\|\mathbf{z}^{(t)}\| \geqslant 0.5\sqrt[4]{\varepsilon b/d}\mu(t)\right] = \mathbb{P}_{\mathcal{E}}\left[\|\mathbf{z}^{(t)}\|^2 \geqslant 0.25\sqrt{\varepsilon b/d}(\mu(t))^2\right] \leqslant 80\sqrt{\varepsilon b/d}\kappa(\xi_1/\xi)^t + \frac{4\beta n}{\sqrt{\varepsilon b/d}}(\xi_2/\xi)^t$$
$$\text{(From Lemma D.6 and Lemma D.7)} \leqslant O(\sqrt{\varepsilon b/d}),$$

which concludes the proof of Theorem D.1. ∎

D.1.5 Proof of Lemma D.3

The goal of this subsection is to prove Lemma D.3, which is essentially just solving the recurrence relation from Lemma D.2. Before we proceed to the proof, we state a fact and an observation regarding eigenvalues and eigenvectors of certain 2×2 natruces, which will be useful in our proof.

Fact D.1. *Let* $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \in \mathbb{R}^{2\times 2}$ *be a* 2×2 *real-valued matrix such that* $a_{11} \neq 0$ *and* $(a_{11} - a_{22})^2 \neq 4a_{12}a_{21}$. *Then, its eigenvalues are*

$$\alpha_1(A) \triangleq \frac{1}{2}\left(a_{11} + a_{22} + \sqrt{(a_{11} - a_{22})^2 + 4a_{12}a_{21}}\right) \text{ and}$$

$$\alpha_2(A) \triangleq \frac{1}{2}\left(a_{11} + a_{22} - \sqrt{(a_{11}-a_{22})^2 + 4a_{12}a_{21}}\right),$$

and its eigenvectors are

$$\begin{bmatrix} 1 \\ \frac{\alpha_1(A)-a_{11}}{a_{12}} \end{bmatrix} \text{ and } \begin{bmatrix} 1 \\ \frac{\alpha_2(A)-a_{11}}{a_{12}} \end{bmatrix}.$$

Observation D.2. *Let $A, a_{11}, a_{12}, a_{21}, a_{22}, \alpha_1(A), \alpha_2(A)$ be as in Fact D.1. Suppose further that $a_{11}, a_{12}, a_{21}, a_{22} \geqslant 0, a_{11} > a_{22}$ and that $(a_{11}-a_{22})^2 > 4a_{12}a_{21}$. Then,*

$$a_{11} + \frac{a_{12}a_{21}}{a_{11}-a_{22}} \geqslant \alpha_1(A) \geqslant a_{11} \geqslant a_{22} \geqslant \alpha_2(A) \geqslant a_{22} - \frac{a_{12}a_{21}}{a_{11}-a_{22}}.$$

Proof. The inequalities come from an observation that $a_{11} - a_{22} \leqslant \sqrt{(a_{11}-a_{22})^2 + 4a_{12}a_{21}} \leqslant a_{11} - a_{22} + \frac{2a_{12}a_{21}}{a_{11}-a_{22}}$. □

We are now ready to prove Lemma D.3.

Proof of Lemma D.3. Let $\{y^{(t)}\}_{t\in\mathbb{Z}_{\geqslant 0}}$ and $\{z^{(t)}\}_{t\in\mathbb{Z}_{\geqslant 0}}$ be sequence of non-negative real numbers defined by $y^{(0)} = \|\mathbf{y}(0)\|^2, z^{(0)} = \|\mathbf{z}(0)\|^2$ and

$$\begin{bmatrix} y^{(t)} \\ z^{(t)} \end{bmatrix} = A \begin{bmatrix} y^{(t-1)} \\ z^{(t-1)} \end{bmatrix} \text{ where } A \triangleq \begin{bmatrix} 1 - \frac{8\delta b}{dn} + \frac{16\delta^2 b}{dn^2} & \frac{16\delta^2 b}{dn^2} \\ \frac{8\delta^2 b}{dn} & 1 - \frac{4\delta(1-\delta)\lambda_3}{n} \end{bmatrix}$$

for every $t \in \mathbb{N}$. Note that, from Lemma D.2 and from the initial values $y^{(0)}, z^{(0)}$, we have $y^{(t)} \geqslant \mathbb{E}_{\mathcal{E}}[\|\mathbf{y}(t)\|^2]$ and $z^{(t)} \geqslant \mathbb{E}_{\mathcal{E}}[\|\mathbf{z}(t)\|^2]$ for every $t \in \mathbb{Z}_{\geqslant 0}$. Hence, to prove the lemma, it suffices to prove that

$$y^{(t)} \leqslant \left(y^{(0)} + \left(\frac{40\varepsilon b}{dn}\right) z^{(0)}\right)\left(1 - \frac{8\delta b}{dn} + \frac{336\delta^2 b}{dn^2}\right)^t$$

and

$$z^{(t)} \leqslant \left((y^{(0)} + \left(\frac{40\varepsilon b}{dn}\right) z^{(0)}\right)\left(1 - \frac{8\delta b}{dn} + \frac{336\delta^2 b}{dn^2}\right)^t + z^{(0)}\left(1 - \frac{4\delta(1-\delta)\lambda_3}{n}\right)^t.$$

Let $a_{11}, a_{12}, a_{21}, a_{22}$ be the entries of A. Note that

$$\begin{aligned} a_{11} - a_{12} &= \left(1 - \frac{8\delta b}{dn} + \frac{16\delta^2 b}{dn^2}\right) - \left(1 - \frac{4\delta(1-\delta)\lambda_3}{n}\right) \\ &\geqslant \frac{4\delta(1-\delta)\lambda_3}{n} - \frac{8\delta b}{dn} \\ &= \frac{4\delta(1-\delta)\lambda_3}{n} - \frac{4\delta\lambda_2}{n} \\ &= \frac{4\delta}{n}(\lambda_3 - \lambda_2 - \delta\lambda_3). \end{aligned}$$

Observe here that, since the sum of eigenvalues of L is equal to $\mathrm{tr}(L) = n$ and $\lambda_1, \lambda_2 \geqslant 0$, we have $\lambda_3 \leqslant \frac{n}{n-2}$. Thus, from $\delta \leqslant 0.8(\lambda_3 - \lambda_2)$, we have that $\delta\lambda_3 \leqslant 0.9(\lambda_3 - \lambda_2)$ for any sufficiently large n. Plugging this into the above inequality gives

$$a_{11} - a_{12} \geqslant \frac{0.4\delta(\lambda_3 - \lambda_2)}{n} = \frac{0.4\delta^2}{\varepsilon n}. \tag{57}$$

The above inequality implies that $a_{11} - a_{22} > 0$. Moreover, for $n \geqslant 1000$, we have

$$(a_{11} - a_{22})^2 > \frac{0.16\delta^4}{\varepsilon^2 n^2} > \frac{128\delta^4 b^2}{d^2 n^3} = a_{12}a_{21}.$$

In other words, the conditions in Fact D.1 and Observation D.2 are satisfied. From Fact D.1, the eigenvalues of A are

$$\alpha_1 \triangleq \frac{1}{2}\left(a_{11} + a_{22} + \sqrt{(a_{11} - a_{22})^2 + 4a_{12}a_{21}}\right)$$

and

$$\alpha_2 \triangleq \frac{1}{2}\left(a_{11} + a_{22} - \sqrt{(a_{11} - a_{22})^2 - 4a_{12}a_{21}}\right).$$

Furthermore, the eigenvectors of A are

$$\mathbf{v}_1 \triangleq \begin{bmatrix} 1 \\ \frac{\alpha_1 - a_{11}}{a_{12}} \end{bmatrix} \text{ and } \mathbf{v}_2 \triangleq \begin{bmatrix} 1 \\ \frac{\alpha_2 - a_{11}}{a_{12}} \end{bmatrix}.$$

Let $\gamma_1 \triangleq \left(\frac{a_{11}-\alpha_2}{\alpha_1-\alpha_2}\right) y^{(0)} + \left(\frac{a_{12}}{\alpha_1-\alpha_2}\right) z^{(0)}$ and $\gamma_2 \triangleq \left(\frac{\alpha_1-a_{11}}{\alpha_1-\alpha_2}\right) y^{(0)} - \left(\frac{a_{12}}{\alpha_1-\alpha_2}\right) z^{(0)}$. It is easy to see that

$$\begin{bmatrix} y^{(0)} \\ z^{(0)} \end{bmatrix} = \gamma_1 \mathbf{v}_1 + \gamma_2 \mathbf{v}_2.$$

Since $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors of A with eigenvalues α_1 and α_2 respectively, we have

$$\begin{bmatrix} y^{(t)} \\ z^{(t)} \end{bmatrix} = \gamma_1 \alpha_1^t \mathbf{v}_1 + \gamma_2 \alpha_2^t \mathbf{v}_2$$

for every $t \in \mathbb{Z}$. In other words, we have

$$y^{(t)} = \gamma_1 \alpha_1^t + \gamma_2 \alpha_2^t, \qquad z^{(t)} = \left(\frac{\alpha_1 - a_{11}}{a_{12}}\right) \gamma_1 \alpha_1^t + \left(\frac{\alpha_2 - a_{11}}{a_{12}}\right) \gamma_2 \alpha_2^t.$$

Having derived the above formula, we will now bound $y^{(t)}, z^{(t)}$ by appropriately bounding the eigenvalues and coefficients. Before we do so, let us list a few inequalities that will be useful.

- From (57), we have the following three ineqalities.

$$\frac{a_{12}}{a_{11} - a_{22}} \leqslant \frac{40\varepsilon b}{dn}, \tag{58}$$

$$\frac{a_{21}}{a_{11} - a_{22}} \leqslant \frac{20\varepsilon b}{d}, \tag{59}$$

  and

$$\frac{a_{12}a_{21}}{a_{11} - a_{22}} \leqslant \frac{320\varepsilon\delta^2 b^2}{d^2 n^2} \leqslant \frac{320\delta^2 b}{dn^2} \tag{60}$$

  where the second inequality comes from $\varepsilon < 1$ and $b \leqslant d$.

- From (60) and from Observation D.2, we have

$$a_{11} + \frac{320\delta^2 b}{dn^2} \geqslant \alpha_1 \geqslant a_{11} \geqslant a_{22} \geqslant \alpha_2 \geqslant a_{22} - \frac{320\delta^2 b}{dn^2}. \tag{61}$$

  Note also that the right-most term above is non-negative for sufficiently large n.

Bounding $y^{(t)}$. With the above inequalities in place, it is now easy to bound $y^{(t)}$ as follows.

$$\begin{aligned}
y^{(t)} &= \gamma_1\alpha_1^t + \gamma_2\alpha_2^t \\
\left(\text{Since } \gamma_2 \leqslant \left(\frac{\alpha_1 - a_{11}}{\alpha_1 - \alpha_2}\right) y^{(0)} \text{ and } \alpha_2 \geqslant 0\right) &\leqslant \gamma_1\alpha_1^t + \left(\frac{\alpha_1 - a_{11}}{\alpha_1 - \alpha_2}\right) y^{(0)}\alpha_2^t \\
(\text{Since } \alpha_2 \leqslant \alpha_1) &\leqslant \gamma_1\alpha_1^t + \left(\frac{\alpha_1 - a_{11}}{\alpha_1 - \alpha_2}\right) y^{(0)}\alpha_1^t \\
&= \left(y^{(0)} + \left(\frac{a_{12}}{\alpha_1 - \alpha_2}\right) z^{(0)}\right)\alpha_1^t \\
(\text{From (61)}) &\leqslant \left(y^{(0)} + \left(\frac{a_{12}}{a_{11} - a_{22}}\right) z^{(0)}\right)\alpha_1^t \\
(\text{From (58)}) &\leqslant \left(y^{(0)} + \left(\frac{8\varepsilon b}{dn}\right) z^{(0)}\right)\alpha_1^t \\
(\text{From (61)}) &\leqslant \left(y^{(0)} + \left(\frac{40\varepsilon b}{dn}\right) z^{(0)}\right)\left(1 - \frac{8\delta b}{dn} + \frac{336\delta^2 b}{dn^2}\right)^t
\end{aligned}$$

as desired.

Bounding $z^{(t)}$. Recall that $z^{(t)} = \left(\frac{\alpha_1 - a_{11}}{a_{12}}\right)\gamma_1\alpha_1^t + \left(\frac{\alpha_2 - a_{11}}{a_{12}}\right)\gamma_2\alpha_2^t$. Let us bound the two terms separately, starting with the first term $\left(\frac{\alpha_1 - a_{11}}{a_{12}}\right)\gamma_1\alpha_1^t$. To this end, we can bound $\left(\frac{\alpha_1 - a_{11}}{a_{12}}\right)\gamma_1$ as follows.

$$\begin{aligned}
\left(\frac{\alpha_1 - a_{11}}{a_{12}}\right)\gamma_1 &\leqslant \left(\frac{a_{21}}{a_{11} - a_{22}}\right)\gamma_1 \\
(\text{From (59)}) &\leqslant \left(\frac{20\varepsilon b}{d}\right)\gamma_1 \\
&= \left(\frac{20\varepsilon b}{d}\right)\left(\left(\frac{a_{11} - \alpha_2}{\alpha_1 - \alpha_2}\right) y^{(0)} + \left(\frac{a_{12}}{\alpha_1 - \alpha_2}\right) z^{(0)}\right) \\
(\text{From (61)}) &\leqslant \left(\frac{20\varepsilon b}{d}\right)\left(y^{(0)} + \left(\frac{a_{12}}{a_{11} - a_{22}}\right) z^{(0)}\right) \\
(\text{From (58)}) &\leqslant \left(\frac{20\varepsilon b}{d}\right)\left(y^{(0)} + \left(\frac{40\varepsilon b}{dn}\right) z^{(0)}\right).
\end{aligned}$$

Note that the first inequality comes from Observation D.2 and from $\gamma_1 \geqslant 0$. From the above bound on $\left(\frac{\alpha_1 - a_{11}}{a_{12}}\right)\gamma_1$ and our bound on α_1 from (61), we have

$$\left(\frac{\alpha_1 - a_{11}}{a_{12}}\right)\gamma_1\alpha_1^t \leqslant \left(\frac{20\varepsilon b}{d}\right)\left(y^{(0)} + \left(\frac{40\varepsilon b}{dn}\right) z^{(0)}\right)\left(1 - \frac{8\delta b}{dn} + \frac{336\delta^2 b}{dn^2}\right)^t. \tag{62}$$

Let us next bound $\left(\frac{\alpha_2 - a_{11}}{a_{12}}\right)\gamma_2\alpha_2^t$. Again, we first rearrange the coefficient $\left(\frac{\alpha_2 - a_{11}}{a_{12}}\right)\gamma_2$ as

$$\begin{aligned}
\left(\frac{\alpha_2 - a_{11}}{a_{12}}\right)\gamma_2 &= \left(\frac{\alpha_2 - a_{11}}{a_{12}}\right)\left(\left(\frac{\alpha_1 - a_{11}}{\alpha_1 - \alpha_2}\right) y^{(0)} - \left(\frac{a_{12}}{\alpha_1 - \alpha_2}\right) z^{(0)}\right) \\
\left(\text{Since } \frac{\alpha_2 - a_{11}}{a_{12}} \leqslant 0 \text{ and } \frac{\alpha_1 - a_{11}}{\alpha_1 - \alpha_2} \geqslant 0\right) &\leqslant \left(\frac{a_{11} - \alpha_2}{\alpha_1 - \alpha_2}\right) z^{(0)} \\
(\text{From (61)}) &\leqslant z^{(0)}.
\end{aligned}$$

Hence, from the above inequality and (61), we have

$$\left(\frac{\alpha_2 - a_{11}}{a_{12}}\right)\gamma_2\alpha_2^t \leqslant z^{(0)}a_{22}^t = z^{(0)}\left(1 - \frac{4\delta(1-\delta)\lambda_3}{n}\right)^t. \tag{63}$$

Combining (62) and (63) indeed yields the desired bound on $z^{(t)}$. □

D.1.6 Proofs of Lemma D.6 and Lemma D.7

Proof of Lemma D.6. Since $\kappa = 1 + 40\varepsilon b\beta/d$ and since $\varepsilon b\beta/d \leqslant 1$, it suffices to show that $(\xi_1/\xi)^t \leqslant 1 + \varepsilon b\beta/d$. To show this, let us rearrange $(\xi/\xi_1)^t$ as follows.

$$\begin{aligned}
\left(\frac{\xi}{\xi_1}\right)^t &= \left(1 - \frac{\xi_1 - \xi}{\xi_1}\right)^t \\
&\geqslant \left(1 - \frac{\frac{336\delta^2 b}{dn^2}}{\xi_1}\right)^t \\
\text{(Since } \xi \geqslant 1/2 \text{ when } n \geqslant 8) &\geqslant \left(1 - \frac{672\delta^2 b}{dn^2}\right)^t \\
\text{(From Bernoulli's inequality)} &\geqslant 1 - \frac{672\delta^2 bt}{dn^2} \\
\text{(Since } t \leqslant \frac{n^2\beta}{1344\delta(\lambda_3 - \lambda_2)}) &\geqslant 1 - \frac{\varepsilon b\beta}{2d}.
\end{aligned}$$

Note that we can apply Bernoulli's inequality since $\frac{672\delta^2 b}{dn^2} \leqslant 1$ for any sufficiently large n (i.e. $n \geqslant 30$). Finally, note that the above inequality implies that $(\xi_1/\xi)^t \leqslant 1 + \varepsilon b\beta/d$ since $\frac{1}{1-\frac{\varepsilon b\beta}{2d}} \leqslant 1 + \varepsilon b\beta/d$ because $\varepsilon b\beta/d \leqslant 1$. □

Proof of Lemma D.7. Observe that, in (57), we have already proved that $\xi - \xi_2 \geqslant \frac{2\delta^2}{\varepsilon n} = \frac{2\delta(\lambda_3-\lambda_2)}{n}$. Moreover, observe that $\xi_2 \leqslant 1$. Hence, we have

$$\begin{aligned}
\left(\frac{\xi}{\xi_2}\right)^t &= \left(1 + \frac{\xi - \xi_2}{\xi_2}\right)^t \\
&\geqslant \left(1 + \frac{2\delta(\lambda_3 - \lambda_2)}{n}\right)^t. \\
\text{(From Bernoulli's inequality)} &\geqslant 2^{\frac{2\delta(\lambda_3-\lambda_2)t}{n}} \\
&\geqslant 2^{4\log\left(\frac{nd\beta}{\varepsilon b}\right)} \\
&= \left(\frac{nd\beta}{\varepsilon b}\right)^4,
\end{aligned}$$

which implies the inequality stated in the lemma. □

E Omitted Proofs from Section 5.2

The main goal of this section is to prove Theorem 5.3. The actual proof deviates in a couple of subtle ways from the outline in Section 5.2.1:

- Firstly, we use a slightly different notion of "good at time t". In the outline, we say that a node is good at time t if $\chi_u(\mathbf{x}_u^{(t)} - a_{||}) \approx \mu(t)/n$. However, since $\chi_u \cdot \mathbf{x}_u^{(T_u(\tau_u^{\mathrm{s}}))} > \chi_u \cdot \mathbf{x}_u^{(T_u(\tau_u^{\mathrm{e}}))}$ suffices to conclude that $\mathbf{h}_u^{jump} = \chi_u$, it is enough for us to pick $\eta \in \mathbb{R}$ as a cutoff threshold and says that u such that $\chi_u(\mathbf{x}_u^{(T_u(t))} - a_{||}) \geqslant \eta$ is *good for stored time* t and u such that $\chi_u(\mathbf{x}_u^{(T_u(t))} - a_{||}) < \eta$ is *good for end time* t. More formally, for each $t \in \mathbb{N}$, let $R_t^\eta \triangleq \{u \in V \mid \chi_u(\mathbf{x}_u^{(t)} - a_{||}) \geqslant \eta\}$ be the set of good nodes for stored time t and $\bar{R}_t^\eta \triangleq V \setminus R_t^\eta$ be the set of good nodes for end time t.
- Secondly, instead of arguing that $[T_u(\tau^{\mathrm{s}}), T_u(\tilde{\tau}^{\mathrm{s}})] \cap [0.5n\tau^{\mathrm{s}}, 0.5n\tilde{\tau}^{\mathrm{s}}]$ is large for most u (and similarly for the ending time), we will argue that $[T_u(\tau^{\mathrm{s}}), T_u(\tilde{\tau}^{\mathrm{s}})] \subseteq [0.4n\tau^{\mathrm{s}}, 0.6n\tilde{\tau}^{\mathrm{s}}]$ for most u, which suffices for our purpose.

More precisely, the main steps of the proof are as follows. After selecting appropriate values of $\tau^{\mathrm{s}}, \tilde{\tau}^{\mathrm{s}}, \tau^{\mathrm{e}}, \tilde{\tau}^{\mathrm{e}}, \eta$, our proof consists of three main steps as follows. For brevity, let us focus on the stored time here as the statements for the end time are analogous.

1. We start by using the concentration result from the previous section to argue that, for each $t \in [0.4n\tau^{\mathrm{s}}, 0.6n\tilde{\tau}^{\mathrm{s}}]$, most nodes are good for stored time t, i.e., $\mathbb{E}_{\mathcal{E}} |R_t^\eta|$ is large. In other words, we will show that $\mathbb{E}_{\mathcal{E}} |\bar{R}_t^\eta|$ is small for such t's.
2. We next argue that, for most nodes u, $T_u(\tau^{\mathrm{s}}), \ldots, T_u(\tilde{\tau}^{\mathrm{s}})$ are "sufficiently uniform" in the following sense: $[T_u(\tau^{\mathrm{s}}), T_u(\tilde{\tau}^{\mathrm{s}})] \subseteq [0.4n\tau^{\mathrm{s}}, 0.6n\tilde{\tau}^{\mathrm{s}}]$ and, for most $\tau \in [\tau^{\mathrm{s}}, \tilde{\tau}^{\mathrm{s}}]$, $T_u(\tau+1) - T_u(\tau)$ is not too much smaller than its expected value, $n/2$.
3. Finally, we show that, if most nodes are uniform, then using local time is not much worse than using global time. In other words, we show that, if the average size of the sets of bad nodes $\bar{R}_t^\eta$ is small over $t \in [0.4n\tau^{\mathrm{s}}, 0.6n\tilde{\tau}^{\mathrm{s}}]$, then the average size of $\bar{R}_{T_u(\tau)}^\eta$ is also small over all $\tau \in [\tau^{\mathrm{s}}, \tilde{\tau}^{\mathrm{s}}]$. The latter indeed implies that most u is unlikely to be bad for random stored time $\tau_u^{\mathrm{s}} \in [\tau^{\mathrm{s}}, \tilde{\tau}^{\mathrm{s}}]$.

The values of the parameters that we will be using throughout this section are as follows.

- Pick $\tau^{\mathrm{s}} \triangleq \frac{100\log\left(\frac{nd}{\varepsilon b}\right)}{\delta(\lambda_3 - \lambda_2)}$, $\tilde{\tau}^{\mathrm{s}} \triangleq 2\tau^{\mathrm{s}}$, $\tau^{\mathrm{e}} \triangleq 3\tilde{\tau}^{\mathrm{s}} + \frac{10d}{\delta b}$ and $\tilde{\tau}^{\mathrm{e}} \triangleq 2\tau^{\mathrm{e}}$.
- Let $\eta \triangleq 0.25\mu(0.6n\tilde{\tau}^{\mathrm{s}})/n$.

E.1 The proof

Let us now proceed to the proof. The two main lemmas of the first step can be stated as follows. Since both lemmas follow easily from our concentration result, we defer their proofs to Appendix E.2.

Lemma E.1. *For any $\mathbf{x}^{(0)}$ such that $\|\mathbf{z}^{(0)}\|^2 \leqslant n\sqrt{d/(\varepsilon b)}\|\mathbf{y}^{(0)}\|^2$ and every $t \in [0.4n\tau^s, 0.6n\tilde{\tau}^s]$, we have $\mathbb{E}_{\mathcal{E}} |\bar{R}_t^\eta| \leqslant O\left(n\sqrt{\varepsilon b/d}\right)$.*

Lemma E.2. *For any $\mathbf{x}^{(0)}$ such that $\|\mathbf{z}^{(0)}\|^2 \leqslant n\sqrt{d/(\varepsilon b)}\|\mathbf{y}^{(0)}\|^2$ and every $t \in [0.4n\tau^e, 0.6n\tilde{\tau}^e]$, we have $\mathbb{E}_{\mathcal{E}} |R_t^\eta| \leqslant O\left(n\sqrt{\varepsilon b/d}\right)$.*

As stated in the proof overview, the second step is to argue that a random sequence of edges $\mathcal{E}$ is "sufficiently uniform" for most node u with high probability. The notion of uniformity needed here is formalized below. Note that the parameters a, b below will later be set to $\frac{\tau^{\mathrm{s}}}{\log n}, \frac{\tilde{\tau}^{\mathrm{s}}}{\log n}$ to achieve uniformity for stored time and $\frac{\tau^{\mathrm{e}}}{\log n}, \frac{\tilde{\tau}^{\mathrm{e}}}{\log n}$ to achieve uniformity for end time.

Definition E.1. *Let a, b, ζ be any positive real number such that $b \geqslant 2a$. We say that a node $u \in V$ is* (a, b, ζ)-uniform *(with respect to a sequence of edges $\mathcal{E}$) if*

- $T_u(a \log n) > 0.4an \log n$ *and* $T_u(b \log n + 1) \leqslant 0.6bn \log n$*, and,*
- $\mathbb{P}_{\tau \in [a \log n, b \log n]} \left[T_u(\tau + 1) < T_u(\tau) + \sqrt{\zeta} n\right] \leqslant 4\sqrt{\zeta}$.

We can argue that, for a random $\mathcal{E}$, with high probability, most nodes are uniform, as stated below. Since this follows from standard Chernoff bound, we defer the proof to Appendix E.3.

Lemma E.3. *With probability $1 - n^{-\Omega(\sqrt{\zeta}a)}$, at least $n - n^{1-\Omega(\sqrt{\zeta}a)}$ nodes are (a, b, ζ)-uniform.*

To state the main lemma of the final step of the proof, let us defined an additional notation: we call a sequence $\{S_t\}_{t\in\mathbb{N}}$ of subsets $S_t \subseteq V$ *compatible with $\mathcal{E}$* if, for every $t \in \mathbb{N}$, $S_t \triangle S_{t+1} \subseteq \{u_{t+1}, v_{t+1}\}$, i.e., S_{t+1} can only differ from S_t on the endpoints of the edge in step t. Observe that the sequences $\{R_t^\eta\}_{t\in N}$and $\{\bar{R}_t^\eta\}_{t\in N}$ are compatible with $\mathcal{E}$ since $\mathbf{x}_u^{(t+1)}$ can change only when $u \in \{u_{t+1}, v_{t+1}\}$. The main lemma of this part is stated below.

Lemma E.4. *For any sequence of edges $\mathcal{E}$ such that at least $(1-\sqrt{\zeta})n$ nodes are (a, b, ζ)-uniform and for any sequence of subsets $\{S_t\}_{t\in\mathbb{N}}$ that is compatible with $\mathcal{E}$ and that $\mathbb{E}_{t\in[0.4an \log n, 0.6bn \log n]} [|S_t|] \leqslant \zeta n$, we have*

$$\mathbb{P}_{u\in V, \tau \in [a \log n, b \log n]} \left[u \in S_{T_u(\tau)}\right] \leqslant O(\sqrt{\zeta}).$$

S_t should be thought of as the set of bad nodes for t; for stored time, we should think of S_t as $\bar{R}_t^\eta$ whereas, for end time, we should think of S_t as R_t^η. The above lemma asserts that, by randomly choosing a local time from $[a \log n, b \log n]$, most nodesill likely end up in a global time step where it is good. The proof of Lemma E.4 is deferred to Subsection E.4.

Let us now show how to use these lemmas to prove Theorem 5.3.

Proof of Theorem 5.3. First, note that the fact that every node is labeled at time $O(\frac{n \log n}{\delta(\lambda_3 - \lambda_2)} + \frac{nd}{\delta b})$ w.h.p. follows easily from applying a union bound on top of a Chernoff bound on $\mathbb{P}_{\mathcal{E}}[T_u(\tilde{\tau}^{\mathrm{e}}) \leqslant 100n\tilde{\tau}^{\mathrm{e}}]$ for each node $u \in V$; this latter probability is simply the same as the probability that sum of $100n\tilde{\tau}^{\mathrm{e}}$ i.i.d. Bernoulli random variables each with mean $2/n$ is less than $\tilde{\tau}^{\mathrm{e}}$.

To we prove the reconstruction guarantee, let us define the following notations for brevity:

- $\Theta^{\mathrm{initial}}$ denotes the event that $\|\mathbf{z}^{(0)}\|^2 \leqslant n\sqrt{d/(\varepsilon b)}\|\mathbf{y}^{(0)}\|^2$.
- Θ^{s} denotes the event that $\mathbb{E}_{t\in[0.4n\tau^{\mathrm{s}}, 0.6n\tau^{\mathrm{s}}]} |\bar{R}_t^\eta| \leqslant n\sqrt[4]{\varepsilon b/d}$.
- Θ^{e} denotes the event that $\mathbb{E}_{t\in[0.4n\tau^{\mathrm{e}}, 0.6n\tau^{\mathrm{e}}]} |R_t^\eta| \leqslant n\sqrt[4]{\varepsilon b/d}$.
- Let $a^{\mathrm{s}} \triangleq \frac{\tau^{\mathrm{s}}}{\log n}, b^{\mathrm{s}} \triangleq \frac{\tilde{\tau}^{\mathrm{s}}}{\log n}, a^{\mathrm{e}} \triangleq \frac{\tau^{\mathrm{e}}}{\log n}$ and $b^{\mathrm{e}} \triangleq \frac{\tilde{\tau}^{\mathrm{e}}}{\log n}$.
- Let $\zeta \triangleq \max\left\{\sqrt{\frac{\varepsilon b}{d}}, \frac{1}{\log n}\right\}$.
- $\Theta^{\mathrm{uniform,s}}$ denotes the event that at least $(1 - \sqrt{\zeta})n$ nodes are $(a^{\mathrm{s}}, b^{\mathrm{s}}, \zeta)$-uniform.
- $\Theta^{\mathrm{uniform,e}}$ denotes the event that at least $(1 - \sqrt{\zeta})n$ nodes are $(a^{\mathrm{e}}, b^{\mathrm{e}}, \zeta)$-uniform.
- Θ denotes the event that $\Theta^{\mathrm{initial}}, \Theta^{\mathrm{s}}, \Theta^{\mathrm{e}}, \Theta^{\mathrm{uniform,s}}$ and $\Theta^{\mathrm{uniform,e}}$ all occur.

Note that Θ here is the "nice" event, where the conditions required in Lemma E.1, Lemma E.2 and Lemma E.4 are satisfied and we can invoke them. Our proof will proceed in two steps: we will first show that the probability that Θ occurs is large and, then, we will use our auxiliary lemmas to show that, conditioned on Θ happening, we achieve the desired reconstruction most of the time.

To bound $\mathbb{P}_{\mathbf{x}^{(0)},\mathcal{E}}[\neg\Theta]$, first note that, from Proposition 1, we have $\mathbb{P}_{\mathbf{x}^{(0)}}[\neg\Theta^{\text{initial}}] \leqslant O(\sqrt[4]{\varepsilon b/d})$. Moreover, from Lemma E.1, we have

$$\mathop{\mathbb{E}}_{\mathbf{x}^{(0)},\mathcal{E}}\left[\mathop{\mathbb{E}}_{t\in[0.4n\tau^{\text{s}},0.6n\tau^{\text{s}}]}|\bar{R}_t^{\eta}| \;\middle|\; \Theta^{\text{initial}}\right] \leqslant O\left(n\sqrt{\varepsilon b/d}\right).$$

From Markov's inequality, this implies that $\mathbb{P}_{\mathbf{x}^{(0)},\mathcal{E}}[\neg\Theta^{\text{s}} \mid \Theta^{\text{initial}}] \leqslant O\left(\sqrt[4]{\varepsilon b/d}\right)$. Similarly, Lemma E.2 implies that $\mathbb{P}_{\mathbf{x}^{(0)},\mathcal{E}}[\neg\Theta^{\text{e}} \mid \Theta^{\text{initial}}] \leqslant O\left(\sqrt[4]{\varepsilon b/d}\right)$. Now, note that, since $\zeta \geqslant 1/\log n$ and $a^{\text{s}} \geqslant 1$, we have that $\sqrt{\zeta}n \geqslant n^{1-\Omega(\sqrt{\zeta}a^{\text{s}})}$ for sufficiently large n. Hence, we can apply Lemma E.3, which implies that $\mathbb{P}_{\mathcal{E}}[\neg\Theta^{\text{uniform,s}}] \leqslant n^{-\Omega(\sqrt{\zeta}a)} \leqslant O(\sqrt{\zeta})$. Similarly, we also have $\mathbb{P}_{\mathcal{E}}[\neg\Theta^{\text{uniform,e}}] \leqslant O(\sqrt{\zeta})$. Finally, by combining these four bounds, we have

$$\begin{aligned}\mathbb{P}[\neg\Theta] &\leqslant \mathbb{P}[\neg\Theta^{\text{initial}}] + \mathbb{P}[\neg\Theta^{\text{s}} \mid \Theta^{\text{initial}}] \\ &\quad + \mathbb{P}[\neg\Theta^{\text{e}} \mid \Theta^{\text{initial}}] + \mathbb{P}[\neg\Theta^{\text{uniform,s}}] + \mathbb{P}[\neg\Theta^{\text{uniform,e}}] \\ &\leqslant O(\sqrt{\zeta}). \end{aligned} \tag{64}$$

We now proceed to the second part of the proof. Let us denote the set of nodes incorrectly labeled by $V^{\text{incorrect}}$, i.e., $V^{\text{incorrect}} = \left\{u \in V \mid \chi_u\left(\mathbf{x}_{T_u(\tau_u^{\text{s}})} - \mathbf{x}_{T_u(\tau_u^{\text{e}})}\right) < 0\right\}$. We will show that

$$\mathop{\mathbb{E}}_{\mathbf{x}^{(0)},\mathcal{E},\{\tau_u^{\text{s}}\}_{u\in V},\{\tau_u^{\text{e}}\}_{u\in V}}\left[|V^{\text{incorrect}}| \mid \Theta\right] \leqslant O\left(n\sqrt{\zeta}\right). \tag{65}$$

Before we prove the above inequality, let us first show how this implies the desired reconstruction property. By applying Markov's inequality to (65), we can conclude that $\mathbb{P}\left[|V^{\text{incorrect}}| \geqslant n\sqrt[4]{\zeta}/2 \;\middle|\; \Theta\right] \leqslant O(\sqrt[4]{\zeta})$. From this and from (64), we can conclude that $\mathbb{P}\left[|V^{\text{incorrect}}| \geqslant n\sqrt[4]{\zeta}/2\right] \leqslant O(\sqrt[4]{\zeta})$. Note that, when $|V^{\text{incorrect}}| \leqslant n\sqrt[4]{\zeta}/2$, the protocol achieves a $\sqrt[4]{\zeta}$-weak reconstruction. Hence, with probability $1 - O(\sqrt[4]{\zeta})$, our protocol achieves a $\sqrt[4]{\zeta}$-weak reconstruction of the graph. Since $\zeta = \max\{\sqrt{\varepsilon b/d}, 1/\log n\}$, this indeed implies the reconstruction property as stated in Theorem 5.3.

Finally, let us next prove (65) and complete our proof of Theorem 5.3. Since $\{\tau_u^{\text{s}}\}_{u\in V}, \{\tau_u^{\text{e}}\}_{u\in V}$ are independent of Θ, we can write the left-hand side of (65) as

$$\begin{aligned}&\mathop{\mathbb{E}}_{\mathbf{x}^{(0)},\mathcal{E},\{\tau_u^{\text{s}}\}_{u\in V},\{\tau_u^{\text{e}}\}_{u\in V}}\left[|V^{\text{incorrect}}| \mid \Theta\right] \\ &= \mathop{\mathbb{E}}_{\mathbf{x}^{(0)},\mathcal{E},\{\tau_u^{\text{s}}\}_{u\in V},\{\tau_u^{\text{e}}\}_{u\in V}}\left[n \cdot \mathop{\mathbb{P}}_{u\in V}[u \in V^{\text{incorrect}}] \;\middle|\; \Theta\right] \\ &= \mathop{\mathbb{E}}_{\mathbf{x}^{(0)},\mathcal{E}}\left[n \cdot \mathop{\mathbb{P}}_{u\in V,\tau_u^{\text{s}},\tau_u^{\text{e}}}[u \in V^{\text{incorrect}}] \;\middle|\; \Theta\right].\end{aligned}$$

Hence, it suffices for us to show that, assuming that Θ happens, $\mathbb{P}_{u\in V,\tau_u^{\text{s}},\tau_u^{\text{e}}}[u \in V^{\text{incorrect}}] \leqslant O(\sqrt{\zeta})$. To prove this, first observe that if $u \in V^{\text{incorrect}}$, then either $u \in \bar{R}^{\eta}_{T_u(\tau_u^{\text{s}})}$ or $u \in R^{\eta}_{T_u(\tau_u^{\text{e}})}$ or both. Thus, we have

$$\mathop{\mathbb{P}}_{u\in V,\tau_u^{\text{s}},\tau_u^{\text{e}}}[u \in V^{\text{incorrect}}] \leqslant \mathop{\mathbb{P}}_{u\in V,\tau_u^{\text{s}}}[u \in \bar{R}^{\eta}_{T_u(\tau_u^{\text{s}})}] + \mathop{\mathbb{P}}_{u\in V,\tau_u^{\text{e}}}[u \in R^{\eta}_{T_u(\tau_u^{\text{e}})}].$$

Since we assume that Θ occurs, $\Theta^{\text{uniform,s}}$ also occurs, which means that we can apply Lemma E.4 for the sequence $S_t = \bar{R}_t^\eta, a = a^{\text{s}}$ and $b = b^{\text{s}}$. This implies that

$$\mathop{\mathbb{P}}_{u \in V, \tau_u^{\text{s}}}[u \in \bar{R}^\eta_{T_u(\tau_u^{\text{s}})}] \leqslant O(\sqrt{\zeta}).$$

Similarly, applying Lemma E.4 with $S_t = R_t^\eta, a = a^{\text{e}}$ and $b = b^{\text{e}}$, we also have

$$\mathop{\mathbb{P}}_{u \in V, \tau_u^{\text{s}}}[u \in R^\eta_{T_u(\tau_u^{\text{e}})}] \leqslant O(\sqrt{\zeta}).$$

By combining the above three inequalities, we indeed arrive at the desired bound. □

E.2 Bounding $|R_t^\eta|$ and $|\bar{R}_t^\eta|$: Proofs of Lemma E.1 and Lemma E.2

Proof of Lemma E.1. Consider any $t \in [0.4n\tau^{\text{s}}, 0.6n\tilde{\tau}^{\text{s}}]$. For brevity, let Θ denote an event that $a_y(t) \in [0.5\mu(t), 1.5\mu(t)]$ and $\|\mathbf{z}(t)\| \leqslant \left(0.5\sqrt[4]{\varepsilon b/d}\right)\mu(t)$. Applying Theorem D.1 with $\beta = \sqrt{\frac{d}{\varepsilon b}}$, we have $\mathbb{P}_{\mathcal{E}}[\Theta] \geqslant 1 - O(\sqrt{\varepsilon b/d})$.

Now, let us bound the size of $\bar{R}_t^\eta$ conditioned on Θ happening. To do so, first recall that, since $\mathbf{x}(t) = a_{||} \cdot (\mathbf{1}/\sqrt{n}) + a_y(t) \cdot (\chi/\sqrt{n}) + \mathbf{z}(t)$, we have $\mathbf{x}_u(t) = a_{||} + \frac{a_y(t)\chi_u}{\sqrt{n}} + \mathbf{z}_u(t)$. Hence, we have

$$(\mathbf{x}_u(t) - a_{||})\chi_u - \eta = \frac{a_y(t)}{n} + \chi_u \mathbf{z}_u(t) - \eta \geqslant \frac{0.25\mu(t)}{n} + \chi_u \mathbf{z}_u(t).$$

where the inequality comes from $a_y(t) \geqslant 0.5\mu(t)$ and from $\mu(t) \geqslant \mu(0.6n\tilde{\tau}^{\text{s}})$. This inequality implies that, if $u \in \bar{R}_t^\eta$, then $|\mathbf{z}_u(t)| > \frac{0.25\mu(t)}{\sqrt{n}}$. As a result, we can conclude that

$$|\bar{R}_t^\eta| < \frac{\|\mathbf{z}(t)\|^2}{(0.25\mu(t)/\sqrt{n})^2} \leqslant O\left(n\sqrt{\varepsilon b/d}\right)$$

where the latter comes from $\|\mathbf{z}(t)\| \leqslant \left(0.5\sqrt[4]{\varepsilon b/d}\right)\mu(t)$.

Thus, we have

$$\mathop{\mathbb{E}}_{\mathcal{E}} |\bar{R}_t^\eta| = \mathop{\mathbb{E}}_{\mathcal{E}}[|\bar{R}_t^\eta| \mid \Theta]\,\mathbb{P}[\Theta] + \mathop{\mathbb{E}}_{\mathcal{E}}[|\bar{R}_t^\eta| \mid \neg\Theta]\,\mathbb{P}[\neg\Theta] \leqslant O(n\sqrt{\varepsilon b/d}) \cdot 1 + n \cdot O(\sqrt{\varepsilon b/d}) \leqslant O(n\sqrt{\varepsilon b/d})$$

as desired. □

The proof of Lemma E.1 is analogous to the above proof and is presented below.

Proof of Lemma E.2. Consider any $t \in [0.4n\tau^{\text{e}}, 0.6n\tilde{\tau}^{\text{e}}]$. First of all, let us note that

$$\frac{\mu(t)}{\eta} = 4\left(1 - \frac{4\delta b}{dn}\right)^{t - 0.6n\tilde{\tau}^{\text{s}}} \leqslant 4\left(1 - \frac{4\delta b}{dn}\right)^{\frac{4dn}{\delta b}} \leqslant 4 \cdot e^{-16} \leqslant 0.25 \tag{66}$$

where the second inequality comes from the fact that $1 + x \leqslant e^x$ for all $x \in \mathbb{R}$.

The rest of the proof proceeds similar to the proof of Lemma E.1. Again, let Θ denote an event that $a_y(t) \in [0.5\mu(t), 1.5\mu(t)]$ and $\|\mathbf{z}(t)\| \leqslant \left(0.5\sqrt[4]{\varepsilon b/d}\right)\mu(t)$. From Theorem D.1 with $\beta = \sqrt{\frac{d}{\varepsilon b}}$, we have $\mathbb{P}_{\mathcal{E}}[\Theta] \geqslant 1 - O(\sqrt{\varepsilon b/d})$.

Conditioned on Θ, observe that

$$(\mathbf{x}_u(t) - a_{||})\chi_u - \eta = \frac{a_y(t)}{n} + \chi_u \mathbf{z}_u(t) - \eta \leqslant -\frac{0.5\mu(t)}{n} + \chi_u \mathbf{z}_u(t).$$

where the inequality comes from $a_y(t) \leqslant 1.5\mu(t)$ and (66). Hence, if $u \in R_t^\eta$, then $|\mathbf{z}_u(t)| > \frac{0.5\mu(t)}{\sqrt{n}}$. Since $\|\mathbf{z}(t)\| \leqslant \left(0.5\sqrt[4]{\varepsilon b/d}\right)\mu(t)$, this implies that $|R_t^\eta| < \frac{\|\mathbf{z}(t)\|^2}{(0.5\mu(t)/\sqrt{n})^2} \leqslant n\sqrt{\varepsilon b/d}$.

Thus, we have

$$\mathop{\mathbb{E}}_{\mathcal{E}} |R_t^\eta| = \mathop{\mathbb{E}}_{\mathcal{E}}[|R_t^\eta| \mid \Theta]\,\mathbb{P}[\Theta] + \mathop{\mathbb{E}}_{\mathcal{E}}[|R_t^\eta| \mid \neg\Theta]\,\mathbb{P}[\neg\Theta] \leqslant n\sqrt{\varepsilon b/d} + n \cdot O(\sqrt{\varepsilon b/d}) \leqslant O(n\sqrt{\varepsilon b/d})$$

as desired. □

E.3 Most Vertices are Uniform: Proof of Lemma E.3

Proof of Lemma E.3. Let us fix a vertex $u \in V$. We will compute the probability that u is (a, b, ζ)-uniform. First, we will bound the probability that the first condition is not satisfied. To do so, let us introduce an additional notation; we use X_t to denote an indicator variable of the event $u \in \{u_t, v_t\}$. Note each X_t is an i.i.d. Bernoulli random variable which is one with probability $2/n$. The probability that $T_u(a \log n) < 0.4an \log n$ can now be written in terms of X_t's as follows.

$$\mathop{\mathbb{P}}_{\mathcal{E}}[T_u(a\log n) \leqslant 0.4an\log n] = \mathop{\mathbb{P}}_{\mathcal{E}}[X_1 + \cdots + X_{0.4an\log n} \geqslant a\log n] \leqslant 2^{-\Omega(a\log n)} \tag{67}$$

where the inequality comes from an application of Chernoff bound. Similarly, we get the following bound for $\mathbb{P}_{\mathcal{E}}[T_u(b\log n + 1) > 0.6b\log n]$:

$$\mathop{\mathbb{P}}_{\mathcal{E}}[T_u(b\log n + 1) < 0.6b\log n] = \mathop{\mathbb{P}}_{\mathcal{E}}[X_1 + \cdots + X_{0.6bn\log n} \leqslant b\log n] \leqslant 2^{-\Omega(b\log n)}. \tag{68}$$

Next, we proceed to bound the probability that the second condition fails. Let Y_τ denote an indicator variable of the event $T_u(\tau+1) < T_u(\tau) + \sqrt{\zeta}n$. Observe that Y_τ's are i.i.d. Moreover, the probability that $Y_\tau = 1$ can be bounded as follows.

$$\mathbb{P}[Y_\tau = 1] = \mathbb{P}\left[\bigwedge_{i=T_u(\tau)}^{T_u(\tau)+\sqrt{\zeta}n-1} u \in \{u_t, v_t\}\right] \leqslant \sum_{i=T_u(\tau)}^{T_u(\tau)+\sqrt{\zeta}n-1} \mathbb{P}\left[u \in \{u_t, v_t\}\right] = 2\sqrt{\zeta}.$$

Hence, by Chernoff bound, we have

$$\mathbb{P}\left[\sum_{\tau=a\log n}^{b\log n} Y_\tau \leqslant 4\sqrt{\zeta}(b-a)\log n\right] \leqslant 2^{-\Omega(\sqrt{\zeta}(b-a)\log n)}. \tag{69}$$

Observe that $\sum_{\tau=a\log n}^{b\log n} Y_\tau \leqslant 4\sqrt{\zeta}(b-a)\log n$ is equivalent to

$$\mathop{\mathbb{P}}_{\tau\in[a\log n, b\log n]}\left[T_u(\tau+1) < T_u(\tau) + \sqrt{\zeta}n\right] \leqslant 4\sqrt{\zeta}.$$

Thus, (67), (68) and (69) together with the fact that $b = 2a$ imply that the probability that u is (a, b, ζ)-standard is at least $1 - 2^{-\Omega(\sqrt{\zeta}a\log n)}$, which is at least $1 - n^{-C\sqrt{\zeta}a}$ for some global constant C.

As a result, the expected number of vertices that are not (a, b, ζ)-uniform is at most $n^{1-C\sqrt{\zeta}a}$. Hence, an application of Markov's inequality implies that, with probability at most $n^{C\sqrt{\zeta}a/2}$, the number of non-uniform vertices is at most $n^{1-C\sqrt{\zeta}a/2}$, which concludes the proof of this lemma. □

E.4 From Global to Local Time: Proof of Lemma E.4

In this subsection, we present the proof of Lemma E.4.

Proof of Lemma E.4. Let $St \subseteq V$ denote the set of (a, b, ζ)-uniform vertices. We can first write the left-handside term so that we separate out the uniform u's from the non-uniform ones as follows.

$$\begin{aligned}\mathop{\mathbb{P}}_{u\in V,\tau\in[a\log n,b\log n]}\left[u \in S_{T_u(\tau)}\right] &\leqslant \mathop{\mathbb{P}}_{u\in V,\tau\in[a\log n,b\log n]}\left[u \in St \wedge u \in S_{T_u(\tau)}\right] + \mathop{\mathbb{P}}_{u\in V}[u \notin St] \\ &\leqslant \mathop{\mathbb{P}}_{u\in V,\tau\in[a\log n,b\log n]}\left[u \in St \wedge u \in S_{T_u(\tau)}\right] + \sqrt{\zeta}\end{aligned}$$

where the last inequality comes from our assumption that there are only $\sqrt{\zeta}n$ non-uniform vertices. For each vertex $u \in V$, denote the set of $\tau \in [a \log n, b \log n]$ such that $T_u(\tau + 1) - T_u(\tau) < \sqrt{\zeta}n$ by R_u. We can further bound the term $\mathbb{P}_{u\in V,\tau\in[a\log n,b\log n]}\left[u \in St \wedge u \in S_{T_u(\tau)}\right]$ by

$$\begin{aligned}\mathop{\mathbb{P}}_{u\in V,\tau\in[a\log n,b\log n]}\left[u \in St \wedge u \in S_{T_u(\tau)}\right] &\leqslant \mathop{\mathbb{P}}_{u\in V,\tau\in[a\log n,b\log n]}\left[u \in St \wedge \tau \notin R_u \wedge u \in S_{T_u(\tau)}\right] \\ &\quad + \mathop{\mathbb{P}}_{u\in V,\tau\in[a\log n,b\log n]}\left[u \in St \wedge \tau \in R_u\right] \\ &\leqslant \mathop{\mathbb{P}}_{u\in V,\tau\in[a\log n,b\log n]}\left[u \in St \wedge \tau \notin R_u \wedge u \in S_{T_u(\tau)}\right] + O(\sqrt{\zeta})\end{aligned}$$

where the second inequality comes from the fact that, if u is (a, b, ζ)-uniform, then $\mathbb{P}_{\tau\in[a\log n,b\log n]}\left[\tau \in R_u\right] \leqslant 4\sqrt{\zeta}$. Hence, to prove the intended inequality, it suffices to show that

$$\mathop{\mathbb{P}}_{u\in V,\tau\in[a\log n,b\log n]}\left[u \in St \wedge \tau \notin R_u \wedge u \in S_{T_u(\tau)}\right] \leqslant \sqrt{\zeta}.$$

Observe that this probability can be further rearranged as

$$\begin{aligned}&\mathop{\mathbb{P}}_{u\in V,\tau\in[a\log n,b\log n]}\left[u \in St \wedge \tau \notin R_u \wedge u \in S_{T_u(\tau)}\right] \\ &= \frac{1}{n(b\log n - a\log n + 1)}\left(\sum_{u\in V}\sum_{\tau\in[a\log n,b\log n]} \mathbb{1}\left[u \in St \wedge \tau \notin R_u \wedge u \in S_{T_u(\tau)}\right]\right) \\ &\leqslant \frac{1}{na\log n}\left(\sum_{u\in St}\sum_{\tau\in[a\log n,b\log n]\setminus R_u} \mathbb{1}\left[u \in S_{T_u(\tau)}\right]\right). \qquad (70)\end{aligned}$$

Let us fix $u \in St$ and $\tau \in [a \log n, b \log n] \setminus R_u$. Recall that, since $\{S_t\}_{t\in\mathbb{N}}$ is compatible with $\mathcal{E}$, we have $\mathbb{1}[u \in S_t] = \mathbb{1}[u \in S_{T_u(\tau)}]$ for every $t \in [T_u(\tau), T_u(\tau + 1))$. Moreover, because $\tau \notin R_u$, we have $T_u(\tau + 1) - T_u(\tau) \geqslant \sqrt{\zeta}n$. Thus, we have

$$\begin{aligned}\mathbb{1}\left[\tau \notin R_u \wedge u \in S_{T(\tau)}\right] &= \mathbb{1}\left[\tau \notin R_u\right]\mathbb{1}\left[u \in S_{T_u(\tau)}\right] \\ &= \mathbb{1}\left[\tau \notin R_u\right]\left(\frac{\sum_{t=T_u(\tau)}^{T_u(\tau+1)-1}\mathbb{1}\left[u \in S_t\right]}{T_u(\tau+1) - T_u(\tau)}\right) \\ &\leqslant \mathbb{1}\left[\tau \notin R_u\right]\left(\frac{\sum_{t=T_u(\tau)}^{T_u(\tau+1)-1}\mathbb{1}\left[u \in S_t\right]}{\sqrt{\zeta}n}\right)\end{aligned}$$

$$\leqslant \frac{\sum_{t=T_u(\tau)}^{T_u(\tau+1)-1} \mathbb{1}\left[u \in S_t\right]}{\sqrt{\zeta} n}.$$

Plugging the above inequality back into (70), the probability

$$\underset{u \in V, \tau \in [a \log n, b \log n]}{\mathbb{P}} \left[u \in St \wedge \tau \notin R_u \wedge u \in S_{T_u(\tau)}\right]$$

can be upper bounded by

$$\begin{aligned}
&\frac{1}{\sqrt{\zeta} n^2 a \log n} \left(\sum_{u \in St} \sum_{\tau \in [a \log n, b \log n] \setminus R_u} \sum_{t=T_u(\tau)}^{T_u(\tau+1)-1} \mathbb{1}\left[u \in S_t\right] \right) \\
&\leqslant \frac{1}{\sqrt{\zeta} n^2 a \log n} \left(\sum_{u \in St} \sum_{\tau \in [a \log n, b \log n]} \sum_{t=T_u(\tau)}^{T_u(\tau+1)-1} \mathbb{1}\left[u \in S_t\right] \right) \\
&= \frac{1}{\sqrt{\zeta} n^2 a \log n} \left(\sum_{u \in St} \sum_{t=T_u(a \log n)}^{T_u(b \log n+1)-1} \mathbb{1}\left[u \in S_t\right] \right).
\end{aligned}$$

Finally, recall from definition of uniform vertices that, if u is (a, b, ζ)-uniform, then $T_u(a \log n) > 0.4an \log n$ and $T_u(b \log n + 1) \leqslant 0.6b \log n$. Combining this with the above inequality, we have

$$\begin{aligned}
\underset{u \in V, \tau \in [a \log n, b \log n]}{\mathbb{P}} \left[u \in St \wedge \tau \notin R_u \wedge u \in S_{T_u(\tau)}\right] &\leqslant \frac{1}{\sqrt{\zeta} n^2 a \log n} \left(\sum_{u \in St} \sum_{t=0.4an \log n}^{0.6b \log n-1} \mathbb{1}\left[u \in S_t\right] \right) \\
&\leqslant \frac{1}{\sqrt{\zeta} n^2 a \log n} \left(\sum_{u \in V} \sum_{t=0.4an \log n}^{0.6b \log n-1} \mathbb{1}\left[u \in S_t\right] \right) \\
&= \frac{1}{\sqrt{\zeta} n^2 a \log n} \left(\sum_{t=0.4an \log n}^{0.6b \log n-1} \sum_{u \in V} \mathbb{1}\left[u \in S_t\right] \right) \\
&= \frac{1}{\sqrt{\zeta} n^2 a \log n} \left(\sum_{t=0.4an \log n}^{0.6b \log n-1} |S_t| \right) \\
\left(\text{Since} \underset{t \in [0.4an \log n, 0.6b \log n]}{\mathbb{E}} |S_t| \leqslant \zeta n\right) &\leqslant \frac{1}{\sqrt{\zeta} n^2 a \log n} \left((0.6bn \log n - 0.4an \log n + 1)\zeta n\right) \\
&= O(\sqrt{\zeta}),
\end{aligned}$$

which concludes our proof. □