Find Your Place: Simple Distributed Algorithms for Community Detection

Luca Becchetti, Andrea Clementi, Emanuele Natale, Francesco Pasquale, Luca Trevisan

FIND YOUR PLACE: SIMPLE DISTRIBUTED ALGORITHMS FOR COMMUNITY DETECTION*

LUCA BECCHETTI†, ANDREA CLEMENTI‡, EMANUELE NATALE§, FRANCESCO PASQUALE‡ , AND LUCA TREVISAN¶

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

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

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

Key words. Distributed Algorithms, Averaging Dynamics, Community Detection, Spectral Analysis, Stochastic Block Models.

1. Introduction.

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

Distributed community detection is an active research topic, motivated by the analysis of the structure of increasingly large networks [42]. A number of large-scale graph processing systems, including Google's Pregel [51] and Apache Giraph [65], rely on suitable simulations of decentralized link-based algorithms, where each node of the network is viewed as a local computing unit.

We investigate the effectiveness of the algorithm above in recovering an underlying cluster structure under a variety of models, which include the *stochastic block model* [36]. Informally speaking, in the scenario we consider, a hidden partition of the vertex set $V = (V_1, V_2)$ exists, so that the subgraph induced by each *community* V_i is a good expander, whereas V_1 and V_2 are "loosely" connected by a sparse cut. Under this setting, we show that the process resulting from the above simple local rule converges, in a logarithmic number of rounds, to a coloring that exactly or approximately (depending on the model) reflects the underlying community structure. We further

*A preliminary version of this paper has been presented at ACM-SIAM SODA 2017 [10].

Funding: This work is partly supported by the National Science Foundation under Grants No. CCF 1540685 and CCF 1655215, by the University of Rome "Tor Vergata" under research programme "Mission: Sustainability" project ISIDE (grant no. E81I18000110005), by Sapienza University of Rome under research programme "Dynamics in complex systems and distributed computing", by ERC Advanced Grants 788893 AMDROMA "Algorithmic and Mechanism Design Research in Online Markets" and 834861 SOReCoDi "Spectral and Optimization Techniques for Robust Recovery, Combinatorial Constructions, and Distributed Algorithms," by the EC H2020RIA project "SoBigData++" (871042), and by MIUR PRIN project ALGADIMAR "Algorithms, Games, and Digital Markets".

†Sapienza Università di Roma, Rome, Italy (becchetti@dis.uniroma1.it)

‡Università di Roma "Tor Vergata", Rome, Italy ([clementi,pasquale]@mat.uniroma2.it)

§CNRS, I3S, UCA, INRIA, Sophia Antipolis, France (natale@i3s.unice.fr)

¶Università Bocconi, Milan, Italy (l.trevisan@unibocconi.it)

show that our approach simply and naturally extends to more communities, providing a quantitative analysis for a regularized version of the stochastic block model with multiple communities.

Besides the general motivations described above, the major theoretical goal of this work is to show, in a rigorous and well-established framework, that the algorithm above turns out to be one of the few examples of a *dynamics* [6, 4, 29, 60] that solves a computational problem that is non-trivial in a centralized setting. By *dynamics* we here mean synchronous distributed algorithms characterized by a very simple structure, whereby the state of a node at round t depends only on its state and a symmetric function of the multiset of states of its neighbors at round $t-1$, while the update rule is the same for every graph and every node and does not change over time. Note that this definition implies that the network is *anonymous*, that is, nodes do not possess distinguished identities. Examples of dynamics include update rules in which every node updates its state to the plurality[1] or the median it sees in its neighborhood,[2] or, as is the case in this paper, every node holds a value, which it updates to the average of the values held by its neighbors. In contrast, an algorithm that, say, proceeds in two phases, using averaging during the first $10 \log n$ rounds and plurality from round $1 + 10 \log n$ onward, with n the number of nodes, is not a dynamics according to our definition, since its update rule depends on the size of the graph. As another example, an algorithm that starts by having the lexicographically first vertex elected as "leader" and then propagates its state to all other nodes again does not meet our definition of dynamics, since it assigns roles to nodes and requires them to possess distinguishable identities.

The AVERAGING dynamics, in which each node updates its value to the average of its neighbors', is perhaps one of the simplest and most interesting examples of linear dynamics, and it always converges when G is connected and not bipartite: It converges to the global average of the initial values if the graph is regular and to a weighted global average if it is not [18, 66]. Important applications of linear dynamics have been proposed in the recent past [41, 7, 72, 43], for example to perform basic tasks such as self-stabilizing *consensus* in faulty distributed systems [12, 74, 62]. The convergence time of the AVERAGING dynamics is the mixing time of a random walk on G [66]. It is logarithmic in $|V|$ if the underlying graph is a good *expander* [37], while it is slower on graphs that exhibit sparse cuts.

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

1.1. Our contributions. Consider a graph $G = (V, E)$. Let A denote its adjacency matrix and D the diagonal matrix such that $D_{u,u}$ is the degree of node u. We show that if a partition (V_1, V_2) of G exists such that $\mathbf{1}_{V_1} - \mathbf{1}_{V_2}$ is[3] (or is close to)

[1] Given a node $v \in V$, the plurality is defined as the most frequent state among those taken on by the neighbors of v (with ties broken by some fixed rule).

[2] When states correspond to rational values.

[3] As explained in more detail later, $\mathbf{1}_{V_i}$ is the vector with $|V|$ components, such that the j-th component is 1 if $j \in V_i$ and is 0 otherwise.

a right eigenvector of the second largest eigenvalue of the graph's transition matrix[4] $P = D^{-1}A$ of G, and the gap between the second and the third largest eigenvalues is sufficiently large, our algorithm recovers (V_1, V_2), or a close approximation thereof, in a logarithmic number of rounds. Though the algorithm we propose does not entail any explicit eigenvector computation, it does exploit the spectral structure of the underlying graph, based on the intuition that our dynamics is a distributed version of the power method [1].

Our approach and analysis present a few novelties. Among these are the completely local and decentralized criterion whereby nodes assign themselves to clusters, and the spectral bounds we prove for certain graph classes. A conceptual contribution is to have each node, at each round t, assign itself to a cluster ("find its place") according to the difference between its values at rounds t and $t-1$. Globally, this criterion amounts to removing the component of the state vector in the eigenspace of the first eigenvalue, without computing it. This idea presents two advantages: it results in a particularly simple algorithm, and it gives a running time that depends on the gap between the second and third eigenvalues of the transition matrix of the graph. In graphs consisting of two expanders connected by a sparse cut, the corresponding running time only depends on the intra-cluster expansion (i.e., on the gap between the second and third eigenvalues of the transition matrix), and it can be $\mathcal{O}(\log |V|)$, while the mixing time of the overall graph (that depends on the gap between the first and the second eigenvalues) can be $\Omega(|V|^{\alpha})$, for some constant $\alpha > 0$. To the best of our knowledge, this is the first distributed block reconstruction algorithm that, in the above specified graphs, converges faster than the mixing time of the underlying random walk.

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

Strong reconstruction for regular clustered graphs. A $(2n, d, b)$-clustered regular graph $G = ((V_1, V_2), E)$ is a connected graph over the vertex set $V_1 \cup V_2$, with $|V_1| = |V_2| = n$, adjacency matrix A, and such that every node has degree d and (exactly) b neighbors outside its cluster. If the two subgraphs induced by V_1 and V_2 are good expanders and b is sufficiently small, the second and third eigenvalues of the graph's transition matrix $P = (1/d) \cdot A$ are separated by a large gap. In this case, we can prove that the following happens with high probability (for short *w.h.p.*[5]): If the AVERAGING dynamics is initialized by having every node choose a value uniformly and independently at random in $\{-1, 1\}$, within a logarithmic number of rounds the system enters a regime in which nodes' values are monotonically increasing or decreasing, depending on the community they belong to. As a consequence, every

[4] This is the transition matrix of a simple random walk over G (see Subsection 2.2 for more on the properties of P).

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

node can apply a simple and completely local clustering rule at each round, which eventually results in a strong reconstruction (Theorem 3.3).

We then show that, under mild assumptions, a graph selected from the *regular stochastic block model* [19] is a $(2n, d, b)$-clustered regular graph that satisfies the above spectral gap hypothesis, w.h.p. We thus obtain a fast and extremely simple dynamics for strong reconstruction, over the full range of parameters of the regular stochastic block model for which this is known to be possible using polynomial-time centralized algorithms [57, 19] (Section 1.2 and Corollary 3.6).

We further show that a natural extension of our algorithm, in which nodes maintain an array of values and an array of colors, correctly identifies a hidden balanced k-partition in a regular clustered graph with a gap between λ_k and λ_{k+1}. Graphs sampled from the regular stochastic block model with k communities satisfy such conditions, w.h.p. (Theorem 5.2).

Weak reconstruction for non-regular clustered graphs. As a main technical contribution, we extend the above analysis to show that our dynamics also ensures weak reconstruction in clustered graphs having two clusters that satisfy an approximate regularity condition and a gap between the second and third eigenvalues of the transition matrix P (Theorem 4.7). As an application, we further prove that these conditions are met by the *stochastic block model* [2, 24, 27, 30, 36, 40, 54], a random graph model that offers a popular framework for the probabilistic modelling of graphs that exhibit good clustering or community properties. We here consider its simplest version, with two equal-sized communities, defined as follows.

DEFINITION 1.1. *The random graph $\mathcal{G}_{2n,p,q}$ consists of $2n$ nodes and the following edge probability: The node set is partitioned into two subsets V_1 and V_2, each of size n; edges linking nodes belonging to the same partition appear in E independently at random with probability $p = p(n)$, while edges connecting nodes from different partitions appear with probability $q = q(n) < p$. In the remainder, we let $a = pn$ and $b = qn$ respectively denote the expected inner degree and the expected outer degree of a node.*

In this paper, we prove that graphs sampled from $\mathcal{G}_{2n,p,q}$ satisfy the above approximate regularity and spectral gap conditions, w.h.p., whenever $a - b > c\sqrt{(a+b) \cdot \log n}$, for a suitable, absolute constant $c > 0$ (Theorem 4.10).

We remark that the latter result for the stochastic block model follows from an analysis that applies to general *non-random* clustered graphs and hence does not exploit crucial properties of random graphs. A further technical contribution of this paper is a refined, ad-hoc analysis of the AVERAGING dynamics for the $\mathcal{G}_{2n,p,q}$ model, showing that this protocol achieves weak reconstruction, correctly classifying a $1 - \varepsilon$ fraction of vertices, in logarithmic time whenever $a - b > \Omega(\sqrt{(a+b)})$ and the expected degree $d = a + b$ grows at least logarithmically (notice that the factor hidden in the notation $\Omega(\cdot)$ depends on ε - see Theorem 4.12). This refined analysis requires a deeper understanding of the eigenvectors of the *transition matrix* of G. Coja-Oghlan [24] defined certain graph properties that guarantee that a near-optimal bisection can be found based on eigenvector computations of the *adjacency matrix*. Similarly, we show simple sufficient conditions under which a right eigenvector of the second largest eigenvalue of the transition matrix of a graph approximately identifies the hidden partition. We give a tight analysis of the spectrum of the transition matrix of graphs sampled from the stochastic block model in Subsection 4.3. Notice that the analysis of the transition matrix is somewhat harder than that of the adjacency matrix, since the entries are not independent of each other; we were not able to find

comparable results in the existing literature.

Final remarks. While we do think that improving performance is an important goal, this was not the main driver behind this work. Rather, we believe the main contribution of this paper is providing rigorous evidence for the effectiveness of completely decentralized and elementary dynamics in addressing important mining tasks, such as community detection in clustered non-dense graphs. The complexity of such a task appears to lie far beyond the tasks to which this kind of dynamics have been applied in the area of distributed computing (for example, consensus problems).

1.2. Related work.

Dynamics for block reconstruction. Dynamics received considerable attention in the recent past across different research communities, both as efficient distributed algorithms [6, 12, 62, 55] and as abstract models of *natural* interaction mechanisms inducing emergent behavior in complex systems [4, 20, 29, 32, 60]. For instance, simple averaging dynamics have been considered to model opinion formation mechanisms [28, 33], while a number of other dynamics have been proposed to describe different social phenomena [31]. *Label propagation algorithms* [64] are dynamics based on majority updating rules [6] and have been applied to some computational problems including clustering. Several papers present experimental results for such protocols on specific classes of clustered graphs [8, 49, 64]. The only available rigorous analysis of label propagation algorithms on planted partition graphs is the one presented in [44], where the authors propose and analyze a label propagation protocol on $\mathcal{G}_{2n,p,q}$ for dense topologies. In particular, their analysis considers the case where $p = \Omega(1/n^{1/4-\varepsilon})$ and $q = \mathcal{O}(p^2)$, a parameter range in which very dense clusters of constant diameter separated by a sparse cut occur w.h.p. In this setting, characterized by a polynomial gap between p and q, simple combinatorial and concentration arguments show that the protocol converges in constant expected time. They also conjecture a logarithmic bound for sparser topologies.

Belief propagation algorithms. Because of their relevance for the reconstruction problem, we also briefly discuss the class of *belief propagation algorithms*, whose simplicity is close to that of dynamics. These algorithms are best known as message-passing algorithms for performing inference in graphical models [50]. Belief propagation cannot be considered a dynamics: At each round, every node sends different messages to each neighbor. As a result, the update rule is not symmetric w.r.t. the neighbors, requiring thus port numbering [70], while the required amount of local memory grows linearly in the degree of the node. There is non-rigorous, strong supporting evidence that some *belief propagation algorithms* might be optimal for the reconstruction problem [27]. A rigorous analysis is a major challenge; in particular, the convergence to the correct value of belief propagation is far from being fully understood on graphs which are not trees [73, 56]. A complex algorithm based on belief propagation has been presented in [58] by Mossel et al.: Among other results, they show that this algorithm achieves a weak reconstruction for $\mathcal{G}_{2n,p,q}$ which is optimal for certain parameters.

General algorithms for block reconstruction. We next compare our dynamics to previous general algorithms for block reconstruction. Several algorithms for community detection are *spectral*: They typically consider the eigenvector associated with the second eigenvalue of the adjacency matrix A of G, or the eigenvector corresponding to the largest eigenvalue of the matrix $A - \frac{d}{n}J$ [14, 23, 24, 54] [6], since these are

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

correlated with the hidden partition. More recently spectral algorithms have been proposed [3, 24, 17, 59, 45, 63] that find a weak reconstruction even in the sparse, tight regime.

Even though the above mentioned algorithms have been presented in a centralized setting, spectral algorithms turn out to be a feasible approach also for distributed models. Indeed, Kempe and McSherry [42] show that eigenvalue computations can be performed in a distributed fashion, yielding distributed algorithms for community detection in various models, including the stochastic block model. However, the algorithm of Kempe and McSherry as well as any distributed version of the above mentioned centralized algorithms are not dynamics. Actually, adopting the effective concept from Hassin and Peleg in [35], such algorithms are not even *light-weight*: Different and non-simple operations are executed at different rounds, nodes have identities, messages are treated differently depending on the originator, and so on. Moreover, a crucial aspect is convergence time: The mixing time of the simple random walk on the graph is a bottleneck for the distributed algorithm of Kempe and McSherry and for any algorithm that performs community detection in a graph G by employing the power method or the Lanczos method [46] as a subroutine to compute the eigenvector associated with the second eigenvalue of the adjacency matrix of G. Notice that the mixing time of graphs sampled from $\mathcal{G}_{2n,p,q}$ is at least of the order of $\frac{a+b}{2b}$: Hence, it can be super-logarithmic and even $n^{\Omega(1)}$.

In general, the reconstruction problem has been studied extensively using a multiplicity of techniques, which include combinatorial algorithms [30], belief propagation [27] and variants of it [58], spectral-based techniques [54, 24], Metropolis approaches [40], and semidefinite programming [2], among others.

Stochastic Block Models have been studied in a number of areas, including computer science [14, 54, 53], probability theory [57], statistical physics [27], and social sciences [36]. Unlike the distributed setting, where the existence of *light-weight protocols* [35] is the main issue (even in non-sparse regimes), in centralized settings, strong attention has been devoted to establishing sharp thresholds for weak and strong reconstruction. Define $a = np$ as the expected *internal degree* (the number of neighbors that each node has on the same side of the partition) and $b = nq$ as the expected *external degree* (the number of neighbors that each node has on the opposite side of the partition). Decelle et al. [27] conjectured that weak reconstruction is possible if and only if $a - b > 2\sqrt{a+b}$. This was proved by Massoulie and Mossel et al. [59, 53, 57]. Strong recovery is instead possible if and only if $a - b > 2\sqrt{a+b} + \log n$ [2].

Versions of the stochastic block model in which the random graph is regular have also been considered [57, 19]. In particular, Brito et al. [19] show that strong reconstruction is possible in polynomial time when $a - b > 2\sqrt{a+b-1}$.

1.3. Roadmap. The rest of this paper is organized as follows. In Section 2, we formalize the main concepts concerning block reconstruction and the AVERAGING dynamics. Section 3 is devoted to the analysis of our protocol in the case of regular graphs with two communities. This analysis is then generalized in Section 4 for the case of almost-regular graphs with two communities. In particular, the important case of the stochastic block model is described in Subsections 4.2 and 4.3. In Section 5, we show how our analysis and the consequent weak reconstruction can be extended to the case of regular graphs having more communities. Some useful but standard results in linear algebra and probability are deferred to the Appendix.

2. Preliminaries.

2.1. Distributed block reconstruction. The block reconstruction of a clustered graph is a two-coloring of the nodes that separates the hidden communities, and it can be defined in two versions depending on whether a small fraction of outliers is allowed or not.

DEFINITION 2.1. *Let $G = ((V_1, V_2), E)$ be a graph with $V_1 \cap V_2 = \emptyset$. An ε-weak reconstruction is a map $f : V_1 \cup V_2 \to \{\texttt{red}, \texttt{blue}\}$ such that there are two subsets $W_1 \subseteq V_1$ and $W_2 \subseteq V_2$ with $|W_1 \cup W_2| \geqslant (1-\varepsilon)|V_1 \cup V_2|$ and $f(W_1) \cap f(W_2) = \emptyset$. When $\varepsilon = 0$ we say that f is a strong reconstruction.*

Given a graph $G = ((V_1, V_2), E)$, the block reconstruction problem requires computing an ε-reconstruction of G. In this paper, we propose the following distributed protocol for this problem. It is based on the averaging dynamics and produces a coloring of the nodes at the end of every round. In the next two sections we show that, within $\mathcal{O}(\log n)$ rounds, the coloring computed by the algorithm we propose achieves *strong reconstruction* of the two blocks in the case of clustered regular graphs and *weak reconstruction* in the case of clustered non-regular graphs.

AVERAGING protocol:
Rademacher initialization: At round $t = 0$ every node $u \in V$ independently samples its initial value from $\{-1, +1\}$ uniformly at random;
Updating rule: At each subsequent round $t \geqslant 1$, every node $u \in V$

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

Distributed implementation of AVERAGING protocol and its properties. The protocol above can be implemented in the popular synchronous LOCAL communication model [48], where nodes share a global clock that is initialized at the beginning of the execution. At each round determined by the global clock, every node can perform local computations and send/receive one message of arbitrary length to/from each of its neighbors. We assume that there is no fault or message corruption: If a message is sent at round t, then it arrives safely before round $t + 1$ begins.

However, we remark that the AVERAGING protocol requires no unique labeling of the nodes, and it is completely oblivious to time, being a dynamics in the strictest sense. Namely, after initialization, the protocol iterates over and over at every node. Convergence to a (possibly weak) reconstruction is a property of the protocol, of which nodes are not aware; it is something that eventually occurs.

Finally, we remark that the clustering criterion is completely *local*, in the sense that a decision is individually and independently made by each node at each round, only on the basis of its state in the current and previous rounds. This may seem counterintuitive at first, but it is only superficially so. Despite being local, the clustering criterion uses information that reflects the global structure of the network, since nodes' values are related to the second eigenvector of the network's transition matrix.

Notational remark. In the remainder of the paper, we use $\mathbf{x}^{(t)}$ to denote the vector of values collectively held by the nodes at the end of round t and, for simplicity of notation, we use $\mathbf{x}^{(0)} = \mathbf{x}$ to denote the vector of binary values held upon initialization.

2.2. The AVERAGING dynamics and random walks on G. The analysis of the AVERAGING dynamics on a graph G is closely related to the behavior of random

walks on G, which are best studied using tools from linear algebra that we briefly summarize below.

Let $G = (V, E)$ be an undirected graph (possibly with multiple edges and self loops), A its adjacency matrix and d_u the degree of node u. The *transition matrix* of (the random walk on) G is the matrix $P = D^{-1}A$, where D is the diagonal matrix such that $D_{u,u} = d_u$. The entry $P_{u,v} = (1/d_u) \cdot A_{u,v}$ thus corresponds to the probability of going from u to v in one step of the random walk on G. So P operates as the random walk process on G by left multiplication, and as the AVERAGING dynamics by right multiplication.

DEFINITION 2.2. *For $i = 1, 2$, define $\mathbf{1}_{V_i}$ as the $|V|$-dimensional vector, whose u-th component is 1 if $u \in V_i$ and is 0 otherwise. If (V_1, V_2) is a bipartition of the nodes with $|V_1| = |V_2| = n$, we define the partition indicator vector $\boldsymbol{\chi} = \mathbf{1}_{V_1} - \mathbf{1}_{V_2}$.*

It is easy to show that, after t rounds of the AVERAGING dynamics, the vector of values at time t can be written as

$$\mathbf{x}^{(t)} = P^t \mathbf{x} \,.$$

The product of the power of a matrix times a vector is best understood in terms of the spectrum of the matrix, which is what we exploit in the next section. In what follows we always denote by

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_{2n}$$

the eigenvalues of P. Recall that P is a stochastic matrix, hence $\lambda_1 = 1$ and $\lambda_{2n} \geqslant -1$. Moreover, for all graphs that are connected and not bipartite, $\lambda_2 < 1$ and $\lambda_{2n} > -1$. We denote by λ the largest, in absolute value, among all but the first two eigenvalues, namely

$$\lambda = \max \{|\lambda_i| \,:\, i = 3, 4, \ldots, 2n\} \,.$$

Unless otherwise specified, the norm of a vector $\mathbf{x}$ is the ℓ_2 norm $\|\mathbf{x}\| = \sqrt{\sum_i (x(i))^2}$, and the norm of a matrix A is the spectral norm

$$\|A\| = \sup_{\mathbf{x}:\|\mathbf{x}\|=1} \|A\mathbf{x}\| \,.$$

For a diagonal matrix, this is the largest diagonal entry in absolute value.

Cuts and conductance. The notion of *conductance* plays an important role in the convergence results we prove in Section 4, in particular in the proofs of Theorems 4.7 and 4.12.

Let $G = (V, E)$ be an undirected graph with $|E| = m$ and consider a subset $S \subseteq V$. Then, the corresponding cut $(S, V - S)$ is defined as the subset of the edges of the graph with one endpoint in S and the other in $V - S$. The *volume* $vol(S)$ of a subset $S \subseteq V$ is defined as

$$vol(S) = \sum_{u \in S} d_u \,.$$

DEFINITION 2.3. *Given a graph $G = (V, E)$ and a subset $S \subseteq V$, we define the associated conductance $\Phi_G(S)$ as*

$$\Phi_G(S) = \frac{|(S, V - S)|}{vol(S)} \,.$$

The conductance *of G is defined as*

$$\Phi_G = \min_{S:vol(S)\leqslant m/2} \Phi_G(S)\,.$$

We simply write Φ in the remainder, whenever G is clear from context.[7]

3. Strong reconstruction for regular graphs. If the graph G is d-regular, then $P = (1/d)A$ is a real symmetric matrix, P and A have the same set of eigenvectors, and $\mathbf{1}$ is an eigenvector with eigenvalue 1. We denote by

$$\mathbf{v}_1 = (1/\sqrt{2n})\mathbf{1}, \mathbf{v}_2, \ldots, \mathbf{v}_{2n}$$

a basis of orthonormal eigenvectors, where each $\mathbf{v}_i$ is the eigenvector associated with eigenvalue λ_i. Then, we can write a vector $\mathbf{x}$ as a linear combination $\mathbf{x} = \sum_i \alpha_i \mathbf{v}_i$, and the averaging process, starting from $\mathbf{x}$, can be described as

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

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

We say that a d-regular graph G is $(2n, d, b)$-*regular* if a partition of the nodes (V_1, V_2) exists such that every node in V_1 has b neighbors in V_2 and every node in V_2 has b neighbors in V_1.

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

Proof. Every node u has b neighbors w on the opposite side of the partition, for which $\chi(w) = -\chi(u)$, and $d-b$ neighbors w on the same side, for which $\chi(w) = \chi(u)$, so

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

□

We next show that, if the regular graph is "well" clustered, then the AVERAGING protocol produces a strong reconstruction of the two clusters, w.h.p. By a *well-clustered graph* we mean a $(2n, d, b)$-regular graph where (V_1, V_2) represents the unique smallest cut.

DEFINITION 3.1 (Clustered regular graphs). *A $(2n, d, b)$-clustered regular graph $G = ((V_1, V_2), E)$ is a $(2n, d, b)$-regular graph such that $1 - 2b/d$ is the second largest eigenvalue of P, that is, $\lambda_2 = 1 - 2b/d$ and $\lambda < \lambda_2$. We also call $a = d - b$ the* inner *degree of each node.*

In the next lemma, we show that, after t rounds of the AVERAGING dynamics over a $(2n, d, b)$-clustered regular graph, the configuration $\mathbf{x}^{(t)}$ is close to a linear combination of $\mathbf{1}$ and $\boldsymbol{\chi}$.

LEMMA 3.2. *Assume that we run the* AVERAGING *dynamics on a $(2n, d, b)$-clustered regular graph G with an arbitrary initial vector $\mathbf{x} \in \{-1, 1\}^{2n}$. Then there are reals α_1, α_2 such that, at every round $t \geqslant 0$,*

$$\mathbf{x}^{(t)} = \alpha_1\mathbf{1} + \alpha_2\lambda_2^t\boldsymbol{\chi} + \mathbf{e}^{(t)}\,, \quad \textit{where} \quad \left\|\mathbf{e}^{(t)}\right\|_\infty \leqslant \lambda^t\sqrt{2n}\,. \tag{3.1}$$

[7] The notions introduced here refer to unweighted graphs, which are the ones addressed in this paper, but they seamlessly extend to weighted, undirected graphs.

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

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

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

$$\begin{aligned}\mathbf{x}^{(t)} &= \frac{1}{2n}\langle \mathbf{x}, \mathbf{1}\rangle \cdot \mathbf{1} + \lambda_2^t \frac{1}{2n}\langle \mathbf{x}, \boldsymbol{\chi}\rangle \cdot \boldsymbol{\chi} + \sum_{i=3}^{2n} \lambda_i^t \alpha_i \mathbf{v}_i \\ &= \alpha_1 \mathbf{1} + \alpha_2 \lambda_2^t \cdot \boldsymbol{\chi} + \sum_{i=3}^{2n} \lambda_i^t \alpha_i \mathbf{v}_i,\end{aligned}$$

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

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

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

$$x^{(t-1)}(u) - x^{(t)}(u) = \alpha_2 \lambda_2^{t-1}(1 - \lambda_2)\chi(u) + e^{(t-1)}(u) - e^{(t)}(u)\,. \tag{3.2}$$

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

THEOREM 3.3 (Strong reconstruction). *Let $G = ((V_1, V_2), E)$ be a connected $(2n, d, b)$-clustered regular graph with $\lambda_2 = 1 - 2b/d > (1 + \delta)\lambda$ for an arbitrarily small constant $\delta > 0$. Then the* AVERAGING *protocol produces a strong reconstruction within $\mathcal{O}(\log n)$ rounds, w.h.p. Moreover, the overall number of messages produced by the protocol (until its completion time) is $\mathcal{O}(m \log n)$ (where $m = |E|$), and each message has size $\Theta(\log n)$, w.h.p.*

Proof. From (3.2) we have that $\mathtt{sgn}\left(x^{(t-1)}(u) - x^{(t)}(u)\right) = \mathtt{sgn}\left(\alpha_2\chi(u)\right)$ whenever

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

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

$$t - 1 \geqslant \log\left(\frac{2\sqrt{2n}}{|\alpha_2|(1 - \lambda_2)}\right) \cdot \frac{1}{\log\left(\lambda_2/\lambda\right)}\,. \tag{3.4}$$

Next, note that the right-hand side of (3.4) can be upper bounded as follows:

$$\begin{aligned}\log\left(\frac{2\sqrt{2n}}{|\alpha_2|(1-\lambda_2)}\right)\cdot\frac{1}{\log(\lambda_2/\lambda)} &= \log\left(\frac{d\sqrt{2n}}{b|\alpha_2|}\right)\cdot\frac{1}{\log(\lambda_2/\lambda)}\\ &< \log\left(\frac{d\sqrt{2n}}{b|\alpha_2|}\right)\cdot\frac{1}{\log(1+\delta)}\\ &\leqslant \frac{2}{\delta}\log\left(\frac{n\sqrt{2n}}{|\alpha_2|}\right), \qquad (3.5)\end{aligned}$$

where the first equality follows since $\lambda_2 = 1 - 2b/d$ in the $(2n, d, b)$-clustered regular case, the second inequality follows from the hypothesis on the spectral gap between λ_2 and λ, while the third inequality follows since $\log(1+\delta) \geqslant \delta/2$, for a sufficiently small δ, and from the trivial bound $d/b \leqslant n$.

The second key step of the proof relies on the randomness of the initial vector. Indeed, since $\mathbf{x}$ is a vector of independent and uniformly distributed random variables in $\{-1, 1\}$, the absolute difference between the two partial averages in the two communities, that is $|\alpha_2|$, is "sufficiently" large, w.h.p. More precisely, from Lemma B.1, if R is the sum of $2n$ Rademacher random variables, for every $0 < \eta < 1$,

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

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

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

for some positive constant γ, w.h.p. In this case, we can upper bound (3.5) as follows:

$$\frac{2}{\delta}\log\left(\frac{n\sqrt{2n}}{|\alpha_2|}\right) \leqslant \frac{2}{\delta}\log\left(\sqrt{2}n^{\gamma+3/2}\right)\,.$$

Hence, for some positive γ, w.h.p. (3.3) is satisfied whenever

$$t - 1 \geqslant \frac{2}{\delta}\log\left(\sqrt{2}n^{\gamma+3/2}\right) = \frac{2}{\delta}\left(\gamma + \frac{3}{2}\right)\log n + \frac{2}{\delta}\log\sqrt{2}\,. \qquad (3.6)$$

As for the communication complexity of the protocol, we observe that, at every round, every node receives one message (i.e., its current state) from each neighbor. So, since we proved that the completion time is w.h.p. $\mathcal{O}(\log n)$, the overall number of messages is $\mathcal{O}(m \log n)$, w.h.p. Finally, in order to correctly apply the coloring rule, the values $x^{(t)}(u)$ must be encoded with $\Theta(\log n)$ bits. ∎

Not surprisingly, a typical case in which the spectral condition required by Theorem 3.3 is satisfied is when the input graph consists of two regular expanders connected by a sparse regular cut. This is formalized in the next corollary.

COROLLARY 3.4. *Let $G = ((V_1, V_2), E)$ be a $(2n, d, b)$-clustered regular graph and let $\lambda_A = \max\{\lambda_2(A_1), \lambda_2(A_2)\}$, where A_1 and A_2 are the adjacency matrices of the subgraphs induced by V_1 and V_2, respectively. If $a - b \geqslant (1+\delta)(\lambda_A + b)$ then G satisfies the hypothesis of Theorem 3.3.*

Proof. We can write the transition matrix as $P = (1/d)(M + E)$ where

$$M = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}, \quad E = \begin{pmatrix} 0 & B \\ B^\intercal & 0 \end{pmatrix},$$

and B is a matrix with exactly b non-zero entries in each row and in each column, and each non-zero entry being 1. Observe that for every $i = 1, \ldots, 2n$ (see e.g. Corollary 4.10 in Chapter IV in [68])

$$|\lambda_i (M + E) - \lambda_i(M)| \leqslant \|E\|_2 \,. \tag{3.7}$$

Since $\lambda_1(A_1) = \lambda_1(A_2) = a$, we also have that $\lambda_1(M) = \lambda_2(M) = a$ and $\lambda_3(M) = \max\{\lambda_2(A_1), \lambda_2(A_2)\} = \lambda_A$ (notice that a value μ is an eigenvalue of M if and only if μ is an eigenvalue of A_1 or an eigenvalue of A_2). Hence,

$$\lambda_3(P) \leqslant (1/d)(\lambda_3(M) + \|E\|_2) \leqslant \frac{\lambda_A + b}{d} \leqslant \frac{1}{1+\delta} \cdot \frac{a-b}{d} = \frac{1}{1+\delta}\,\lambda_2 \,,$$

where in the first inequality we used (3.7), in the second one we used $\lambda_3(M) = \lambda_A$ and $\|E\|_2 \leqslant b$, and in the third inequality we used the hypothesis $a - b \geqslant (1+\delta)(\lambda_A + b)$. □

REMARK 1. *While we defined $(2n, d, b)$-clustered regular graphs as simple graphs, Definition 3.1 could be easily extended to include regular* multigraphs*, where regularity clearly refers to the standard definition of multiedges. Then Theorem 3.3 holds for this class of multigraphs as well.*

3.1. Regular stochastic block model. We can use Theorem 3.3 to prove that the AVERAGING protocol achieves strong reconstruction in the regular stochastic block model. In the case of two communities, a graph on $2n$ vertices is obtained as follows: Given two parameters $a(n)$ and $b(n)$ (*internal* and *external* degrees, respectively), partition the vertices into two equal-sized subsets V_1 and V_2 and then sample a random $a(n)$-regular graph over each of V_1 and V_2 and a random $b(n)$-regular graph between V_1 and V_2. This model can be instantiated in different ways depending on how one samples the random regular graphs (for example, via the uniform distribution over regular graphs, or by taking the disjoint union of random matchings) [57, 19]. We here consider a regular stochastic block model obtained as *random lift* [5].

If G is a multigraph on n vertices, then a *random k-lift* of G is a distribution over graphs G' on kn vertices sampled as follows: every vertex v of G is replaced by k vertices $v_1, \ldots, v_k$ in G', every edge (u, v) in G is replaced by a random bipartite matching between $u_1, \ldots, u_k$ and $v_1, \ldots, v_k$ (if there are multiple edges, each edge is replaced by an independently sampled matching) and every self loop over u is replaced by a random degree-2 graph over $u_1, \ldots, u_k$ which is sampled by taking a random permutation $\pi : \{1, \ldots, k\} \to \{1, \ldots, k\}$ and connecting u_i to $u_{\pi(i)}$ for every i. The regular stochastic block model we consider here is a random n-lift of the graph that has only two vertices v_1 and v_2: it has b parallel edges between v_1 and v_2, and it has $a/2$ self-loops on v_1 and $a/2$ self-loops on v_2.

If G is a graph sampled according to any random regular stochastic block model with internal and external degrees a and b respectively (notice that those sampled according to random lifts are, in general, multi-graphs), then G is a $(2n, d, b)$-clustered graph with largest eigenvalue of the transition matrix 1 and corresponding eigenvector $\mathbf{1}$, while $\boldsymbol{\chi}$ is also an eigenvector, with eigenvalue $1 - 2b/d$, where $d = a + b$. By using a

general result of Friedman and Kohler [34] on *random k-lifts* of a graph, we can derive the following upper bound on the maximal absolute value achieved by the other $2n-2$ eigenvalues corresponding to eigenvectors orthogonal to $\mathbf{1}$ and $\boldsymbol{\chi}$.

LEMMA 3.5. *Let G be a graph sampled from the regular stochastic block model with internal and external degrees a and b, respectively, such that $a - b > 2\sqrt{a+b} + o(1)$. Then, w.h.p.,*

$$\lambda \leqslant \frac{2}{a+b}(\sqrt{a+b-1} + o(1))\,. \tag{3.8}$$

Proof. The lemma is a consequence of the following general results of Friedman and Kohler [34] (see also Bordenave [16]).

For every lift of any d-regular graph, the lifted graph is still d-regular, and every eigenvalue of the adjacency matrix of the base graph is still an eigenvalue of the lifted graph. Our base graph consists of only two nodes and it has two eigenvalues, that is, 1 and $1 - 2b/d$. When we consider the random n-lift, the obtained multi-graph has $2n - 2$ new eigenvalues.

In this setting, Friedman and Kohler [34] prove that, if $d \geqslant 3$, then with probability $1-\mathcal{O}(1/n)$ over the choice of a random n-lift, the new eigenvalues of the adjacency matrix of the lifted graph are at most $2\sqrt{d-1} + o(1)$ in absolute value. [8] ∎

Since $\lambda_2 = \frac{a-b}{a+b}$, using (3.8) in Theorem 3.3 for clustered regular multigraphs (see Remark 1), we get a strong reconstruction for the regular stochastic block model.

COROLLARY 3.6. *Let η be an arbitrary positive constant and let G be a random graph sampled from the regular stochastic block model with the parameters a and b such that $a - b > 2(1+\eta)\sqrt{a+b}$. Then the* AVERAGING *protocol produces a strong reconstruction in $\mathcal{O}(\log n)$ rounds, w.h.p.*

Proof. The hypothesis on $a - b$ implies that

$$\lambda_2 = \frac{a-b}{a+b} > \frac{2(1+\eta)}{\sqrt{a+b}}\,. \tag{3.9}$$

On the other hand, for sufficiently large n,

$$\lambda \leqslant \frac{2}{a+b}(\sqrt{a+b-1} + o(1)) < \frac{2}{\sqrt{a+b}}\,. \tag{3.10}$$

Together, (3.9) and (3.10) imply $\lambda_2 > (1+\eta)\lambda$. Then, the claim follows from Theorem 3.3. In particular, plugging $\delta = \eta$ into (3.6),

$$t - 1 \geqslant \frac{2}{\eta}\left(\gamma + \frac{3}{2}\right)\log n + \frac{2}{\eta}\log\sqrt{2}\,.$$

∎

4. Non-regular graphs.

[8] Bordenave [16, Corollary 21] has considerably simplified the proof of Friedman and Kohler; although he does not explicitly state the probability of the above event, his argument also bounds the failure probability by $1/k^{\Omega(1)}$ [15].

The AVERAGING *process on non-regular graphs.* The results of Section 3 rely on very clear spectral properties of regular, clustered graphs, immediately reflecting their underlying topological structure. If G is not regular then the matrix $P = D^{-1}A$ is not symmetric in general. However it is possible to relate its eigenvalues and eigenvectors to those of a symmetric matrix as follows. We consider the *normalized adjacency matrix* of G

$$N = D^{-1/2}AD^{-1/2} = D^{1/2}PD^{-1/2}\,.$$

Notice that N is symmetric, and P and N have the same eigenvalues $\lambda_1, \ldots, \lambda_{2n}$. We also recall that λ is defined as the largest, in absolute value, among all but the first two eigenvalues, namely

$$\lambda = \max\{|\lambda_i| \,:\, i = 3, 4, \ldots, 2n\}\,.$$

Moreover, $\mathbf{v}$ is an eigenvector of P if and only if $D^{1/2}\mathbf{v}$ is an eigenvector of N. Finally, $P = N$ when G is regular.

Let $\mathbf{w}_1, \ldots, \mathbf{w}_{2n}$ be a basis of orthonormal eigenvectors of the normalized adjacency matrix N of G, with $\mathbf{w}_i$ defined as the eigenvector associated with the eigenvalue λ_i, for each i. Notice that $\mathbf{w}_1 = \frac{D^{1/2}\mathbf{1}}{\|D^{1/2}\mathbf{1}\|}$. If we set $\mathbf{v}_i = D^{-1/2}\mathbf{w}_i$, we obtain a set of eigenvectors for the transition matrix P, and we can express the initial vector as a linear combination of them, i.e.,

$$\mathbf{x} = \sum_i \tilde{\alpha}_i \mathbf{v}_i\,. \tag{4.1}$$

Then, the AVERAGING process can again be described by the following linear equation

$$\mathbf{x}^{(t)} = P^t\mathbf{x} = \sum_i \lambda_i^t \tilde{\alpha}_i \mathbf{v}_i = \tilde{\alpha}_1\mathbf{v}_1 + \sum_{i=2}^{2n} \lambda_i^t \tilde{\alpha}_i \mathbf{v}_i. \tag{4.2}$$

We remark that (4.1) uniquely defines the coefficients $\tilde{\alpha}_i$, and, since P is not symmetric, they cannot be written as simple inner products. The next lemma summarizes some useful properties of the AVERAGING process for general graphs.

LEMMA 4.1. *Let G be connected and not bipartite and let $\mathbf{x}$ be any initial vector. Then*

1. *The coefficients in the spectral decomposition of $\mathbf{x}$ in (4.1) are*

$$\tilde{\alpha}_i = \mathbf{w}_i^\intercal D^{1/2}\mathbf{x} = \mathbf{v}_i^\intercal D\mathbf{x}\,, \quad \text{for } i = 1, \ldots, 2n.$$

2. *For every $t \geqslant 1$, the difference between vector $\mathbf{x}^{(t)}$ in (4.2) in two consecutive rounds is*

$$\mathbf{x}^{(t-1)} - \mathbf{x}^{(t)} = \tilde{\alpha}_2(1-\lambda_2)\lambda_2^{t-1}\mathbf{v}_2 + \mathbf{e}^{(t-1)} - \mathbf{e}^{(t)}, \quad \text{where } \mathbf{e}^{(t)} = \sum_{i=3}^{2n} \lambda_i^t \tilde{\alpha}_i \mathbf{v}_i. \tag{4.3}$$

3. *For every $t \geqslant 0$, the norm of vector $\mathbf{e}^{(t)}$ in (4.3) is bounded by*

$$\|\mathbf{e}^{(t)}\| \leqslant \sqrt{\frac{d_{\max}}{d_{\min}}}\,\lambda^t\|\mathbf{x}\| = \lambda^t\sqrt{\frac{d_{\max}}{d_{\min}}2n}\,. \tag{4.4}$$

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

Proof. The first claim follows from the definition of the $\tilde{\alpha}_i$'s. Indeed, since $\mathbf{x} = \sum_i \tilde{\alpha}_i \mathbf{v}_i$,

$$D^{1/2}\mathbf{x} = \sum_i \tilde{\alpha}_i D^{1/2}\mathbf{v}_i = \sum_i \tilde{\alpha}_i \mathbf{w}_i \,.$$

On the other hand, $\mathbf{w}_i$'s are orthonormal, so $\tilde{\alpha}_i$ is the projection of $D^{1/2}\mathbf{x}$ along $\mathbf{w}_i$, i.e.,

$$\tilde{\alpha}_i = \mathbf{w}_i^{\intercal} D^{1/2}\mathbf{x} \,.$$

As for the second claim, (4.3) follows straightforwardly from (4.2), when one considers $\mathbf{x}^{(t-1)} - \mathbf{x}^{(t)}$.

As for the third claim,

$$\begin{aligned}
\|\mathbf{e}^{(t)}\|^2 &= \left\| \sum_{i=3}^{2n} \tilde{\alpha}_i \lambda_i^t D^{-1/2}\mathbf{w}_i \right\|^2 \leqslant \left\|D^{-1/2}\right\|^2 \left\| \sum_{i=3}^{2n} \tilde{\alpha}_i \lambda_i^t \mathbf{w}_i \right\|^2 = \left\|D^{-1/2}\right\|^2 \sum_{i=3}^{2n} \tilde{\alpha}_i^2 \lambda_i^{2t} \\
&\leqslant \left\|D^{-1/2}\right\|^2 \lambda^{2t} \sum_{i=3}^{2n} \tilde{\alpha}_i^2 \leqslant \left\|D^{-1/2}\right\|^2 \lambda^{2t} \left\|D^{1/2}\mathbf{x}\right\|^2 \\
&\leqslant \left\|D^{-1/2}\right\|^2 \left\|D^{1/2}\right\|^2 \lambda^{2t}\|\mathbf{x}\|^2 \leqslant \frac{d_{\max}}{d_{\min}} \lambda^{2t}\|\mathbf{x}\|^2 = \frac{d_{\max}}{d_{\min}} \lambda^{2t} 2n,
\end{aligned}$$

where the last equality follows since, by definition, $\|\mathbf{x}\|^2 = 2n$. ∎

Lemma 4.1 states that, if G is connected and not bipartite, the AVERAGING dynamics converges to $\tilde{\alpha}_1 \mathbf{v}_1$, which is parallel to $\mathbf{1}$. Moreover, as in the regular case, if the second eigenvalue λ_2 of the transition matrix P of a clustered almost-regular graph is close to 1 and $|\lambda_3|, \ldots, |\lambda_{2n}|$ are sufficiently small, the AVERAGING dynamics exhibits a long phase in which $\mathbf{x}^{(t)} = P^t\mathbf{x}$ is close to $\tilde{\alpha}_1\mathbf{1} + \tilde{\alpha}_2\mathbf{v}_2$. The next result establishes a crucial, general relationship between the coloring rule used by the AVERAGING protocol and the sign of $\tilde{\alpha}_2 v_2(u)$.

COROLLARY 4.2. *Let G be connected and not bipartite, and let $\mathbf{x}$ be any initial vector. If $\tilde{\alpha}_2 \neq 0$, then, for every component (i.e., node) u such that $v_2(u) \neq 0$,*

(4.5)

$$\mathtt{sgn}(x^{(t-1)}(u) - x^{(t)}(u)) = \mathtt{sgn}(\tilde{\alpha}_2 v_2(u)) \,, \text{ whenever } t - 1 \geqslant \frac{\log\left(\frac{2\sqrt{2n(d_{\max}/d_{\min})}}{|\tilde{\alpha}_2|\cdot|v_2(u)|(1-\lambda_2)}\right)}{\log(\lambda_2/\lambda)} \,.$$

Proof. From Claim 3 of Lemma 4.1 (in particular, using (4.4)),

$$|e^{(t-1)}(u) - e^{(t)}(u)| \leqslant \|\mathbf{e}^{(t-1)} - \mathbf{e}^{(t)}\| \leqslant \|\mathbf{e}^{(t-1)}\| + \|\mathbf{e}^{(t)}\| < 2\sqrt{\frac{d_{\max}}{d_{\min}}}\lambda^{t-1}\|\mathbf{x}\| \,.$$

Hence, (4.3) implies that $\mathtt{sgn}(x^{(t-1)}(u) - x^{(t)}(u)) = \mathtt{sgn}(\tilde{\alpha}_2 v_2(u))$, whenever

$$|\tilde{\alpha}_2 v_2(u)|\lambda_2^{t-1}(1-\lambda_2) \geqslant 2\sqrt{\frac{d_{\max}}{d_{\min}}}\lambda^{t-1}\|\mathbf{x}\|,$$

that is, whenever

$$\left(\frac{\lambda_2}{\lambda}\right)^{t-1} \geqslant \frac{2\sqrt{2n(d_{\max}/d_{\min})}}{|\tilde{\alpha}_2|\cdot|v_2(u)|(1-\lambda_2)} \,.$$

Taking logarithms of both sides yields (4.5). ∎

The result stated by Corollary 4.2 will be leveraged in the following subsections to prove weak reconstruction for the following class of clustered, almost regular graphs.

DEFINITION 4.3 (Clustered γ-regular graphs). *A $(2n, d, b, \gamma)$-clustered graph $G = ((V_1, V_2), E)$ is a graph over the vertex set $V = V_1 \cup V_2$, where $|V_1| = |V_2| = n$ such that: i) every node $u \in V$ has degree $d_u = d \pm \gamma d$, and ii) every node in V_1 has $b \pm \gamma d$ neighbors in V_2, and every node in V_2 has $b \pm \gamma d$ neighbors in V_1.*

We observe that, in any $(2n, d, b, \gamma)$-clustered graph, the ratio $d_{\max}/d_{\min}$ is upper bounded by $(1+\gamma)/(1-\gamma)$. Moreover, since

$$\|D^{1/2}\mathbf{1}\| = \|D^{1/2}\boldsymbol{\chi}\| = \sqrt{\sum_{u\in[2n]} d_u}\,,$$

from the definition of $(2n, d, b, \gamma)$-clustered graph,

$$\sqrt{(1-\gamma)2nd} \leqslant \|D^{1/2}\mathbf{1}\| = \|D^{1/2}\boldsymbol{\chi}\| \leqslant \sqrt{(1+\gamma)2nd}\,. \tag{4.6}$$

In what follows we will also name ν the quantity

$$\nu = 1 - \frac{2b}{d},$$

since it will appear several times in the remainder of this section.[9]

Conductance and Cheeger's inequalities. Cheeger's inequalities are a key result in spectral graph theory, relating the second eigenvalue of N to the conductance Φ of a graph G (see Definition 2.3). In particular, we will make use of the following version of them.

THEOREM 4.4 ([21, 22, 39]). *Let $G = (V, E)$ be a graph. Then*

$$\frac{\Phi^2}{2} \leqslant 1 - \lambda_2 \leqslant 2\Phi\,. \tag{4.7}$$

Roadmap of Section 4. The rest of this section in organized as follows: In Subsection 4.1 we show that the AVERAGING dynamics produces a "weak" reconstruction for the family of $(2n, d, b, \gamma)$-clustered graphs (see Theorem 4.7). In Subsection 4.2 we show that graphs sampled according to the stochastic block model $\mathcal{G}_{2n,p,q}$ belong to the above family, w.h.p., for large ranges of the parameters p and q (see Theorem 4.10). In Section 4.3 we provide an ad-hoc analysis for $\mathcal{G}_{2n,p,q}$ over a parameter range that matches the weak reconstruction threshold (up to a constant factor) in the case of logarithmic node degree (see Theorem 4.12).

4.1. Analysis of the AVERAGING dynamics on almost-regular clustered graphs. In this subsection, we generalize (3.1) to $(2n, d, b, \gamma)$-clustered graphs, showing that for this class, $\mathbf{x}^{(t)}$ is still close to a linear combination of $\mathbf{1}$ and $\boldsymbol{\chi}$. This is the key ingredient to prove weak reconstruction for this class (see Theorem 4.7).

Differently from the regular case, however, the partition indicator vector $\boldsymbol{\chi}$ no longer is an eigenvector of P. As a consequence, the argument needed to prove the above result is considerably harder than in the regular case. In the following Lemma 4.5 we prove that $D^{1/2}\boldsymbol{\chi}$ is close to its projection on $\mathbf{w}_2$. We will use this fact in Lemma 4.6, to prove that $\boldsymbol{\chi}$ is close to its projection on the second eigenvector $\mathbf{v}_2$ of P.

[9] Recall from Fact 1 that $\lambda_2 = \nu$ in clustered regular graphs.

LEMMA 4.5. *Let G be a connected $(2n, d, b, \gamma)$-clustered graph with $\gamma \leqslant 1/10$. If $\lambda_3 < \nu$ then*

$$\left\| D^{1/2}\boldsymbol{\chi} - \beta_2 \mathbf{w}_2 \right\| \leqslant \frac{52\,\gamma}{\nu - \lambda_3}\sqrt{2nd}\,,$$

where $\beta_2 = \boldsymbol{\chi}^\intercal D^{1/2}\mathbf{w}_2$ is the length of $D^{1/2}\boldsymbol{\chi}$'s projection onto $\mathbf{w}_2$.

Proof. For every node v, we denote by a_v and b_v the numbers of v's neighbors in its own cluster and in the other cluster, respectively, so that its degree is $d_v = a_v + b_v$. From the definition of $(2n, d, b, \gamma)$-clustered graphs, we know that

$$(1-\gamma)d \leqslant d_v \leqslant (1+\gamma)d \quad \text{and} \quad b - \gamma d \leqslant b_v \leqslant b + \gamma d\,.$$

Then, for any node v, we get

$$|a_v - b_v - \nu d_v| = \left| 2b\frac{d_v}{d} - 2b_v \right| \leqslant |2b(1+\gamma) - 2(b - \gamma d)| \leqslant 4\gamma d\,. \tag{4.8}$$

Using the above inequality, we obtain

$$\begin{aligned}
\|A\boldsymbol{\chi} - \nu D\boldsymbol{\chi}\|^2 &= \sum_{v\in[2n]} \left(\sum_{w\in \mathrm{Neigh}(v)} \chi(w) - \nu d_v \chi(v) \right)^2 \\
&= \sum_{v\in[2n]} \left(a_v\chi(v) - b_v\chi(v) - \nu d_v \chi(v) \right)^2 \\
&= \sum_{v\in[2n]} \left(a_v - b_v - \nu d_v \right)^2 \\
\text{(from (4.8))} \quad &\leqslant 32nd^2\gamma^2\,. \qquad (4.9)
\end{aligned}$$

Thus,

$$\begin{aligned}
\left\| ND^{1/2}\boldsymbol{\chi} - \nu D^{1/2}\boldsymbol{\chi} \right\| &= \left\| D^{-1/2}A\boldsymbol{\chi} - \nu D^{1/2}\boldsymbol{\chi} \right\| = \left\| D^{-1/2}\left(A\boldsymbol{\chi} - \nu D\boldsymbol{\chi}\right) \right\| \\
&\leqslant \left\| D^{-1/2} \right\| \cdot \|A\boldsymbol{\chi} - \nu D\boldsymbol{\chi}\| \\
\text{(from Hyp. } \gamma \leqslant 1/10 \text{ and (4.9))} \quad &\leqslant \frac{\sqrt{10/9}}{\sqrt{d}} \cdot \sqrt{2n} 4d\gamma \\
&\leqslant 8\sqrt{2nd}\,\gamma\,. \qquad (4.10)
\end{aligned}$$

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

$$|\mathbf{1}^\intercal D\boldsymbol{\chi}| = \left| \sum_{v\in[2n]} \chi(v) d_v \right| \leqslant (1+\gamma)dn - (1-\gamma)dn = 2nd\,\gamma\,. \tag{4.11}$$

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

$$D^{1/2}\boldsymbol{\chi} = \frac{\mathbf{1}^\intercal D\boldsymbol{\chi}}{\|D^{1/2}\mathbf{1}\|^2} D^{1/2}\mathbf{1} + \mathbf{y}\,. \tag{4.12}$$

From (4.12) and triangle inequality we get

$$\begin{aligned}\left\|D^{1/2}\boldsymbol{\chi}-\beta_2\mathbf{w}_2\right\| &= \left\|\frac{\mathbf{1}^\intercal D\boldsymbol{\chi}}{\|D^{1/2}\mathbf{1}\|^2}D^{1/2}\mathbf{1}+\mathbf{y}-\beta_2\mathbf{w}_2\right\| \\ &\leqslant \frac{|\mathbf{1}^\intercal D\boldsymbol{\chi}|}{\|D^{1/2}\mathbf{1}\|}+\|\mathbf{y}-\beta_2\mathbf{w}_2\|\,. \end{aligned} \tag{4.13}$$

As for the first term in (4.13), from (4.11) and the lower bound on $\|D^{1/2}\mathbf{1}\|$ in (4.6) it follows that

$$\begin{aligned}\frac{|\mathbf{1}^\intercal D\boldsymbol{\chi}|}{\|D^{1/2}\mathbf{1}\|} &\leqslant \frac{\gamma}{\sqrt{1-\gamma}}\sqrt{2nd} \\ \text{(from Hyp. } \gamma\leqslant 1/10) \quad &\leqslant 2\gamma\sqrt{2nd}\,. \end{aligned} \tag{4.14}$$

As for the second term in (4.13), we first remark that

$$b\leqslant d/2 \;\text{ and }\; \left\|D^{1/2}\mathbf{1}\right\| \geqslant (1/2)\sqrt{2nd}, \tag{4.15}$$

where the second inequality above holds since $\gamma\leqslant 1/10$. Then

$$\begin{aligned}\|N\mathbf{y}-\nu\mathbf{y}\| &= \left\|N\left(D^{1/2}\boldsymbol{\chi}-\frac{\mathbf{1}^\intercal D\boldsymbol{\chi}}{\|D^{1/2}\mathbf{1}\|^2}D^{1/2}\mathbf{1}\right)-\nu\left(D^{1/2}\boldsymbol{\chi}-\frac{\mathbf{1}^\intercal D\boldsymbol{\chi}}{\|D^{1/2}\mathbf{1}\|^2}D^{1/2}\mathbf{1}\right)\right\| \\ &\leqslant \left\|ND^{1/2}\boldsymbol{\chi}-\nu D^{1/2}\boldsymbol{\chi}\right\|+\frac{|\mathbf{1}^\intercal D\boldsymbol{\chi}|}{\|D^{1/2}\mathbf{1}\|^2}\left\|ND^{1/2}\mathbf{1}-\nu D^{1/2}\mathbf{1}\right\| \\ &= \left\|ND^{1/2}\boldsymbol{\chi}-\nu D^{1/2}\boldsymbol{\chi}\right\|+\frac{|\mathbf{1}^\intercal D\boldsymbol{\chi}|}{\|D^{1/2}\mathbf{1}\|}\,\frac{2b}{d} \\ &\leqslant 8\sqrt{2nd}\,\gamma+4\sqrt{2nd}\,\gamma\,, \end{aligned} \tag{4.16}$$

where: in the first equality we used (4.12), in the second inequality we used the triangular inequality, the third equality follows by the definition of ν, and in the last inequality we used (4.10), (4.14), and (4.15).
We can now bound $\|\mathbf{y}\|$ as follows.

$$\begin{aligned}\|\mathbf{y}\| &\geqslant \left\|D^{1/2}\boldsymbol{\chi}\right\|-\frac{\mathbf{1}^\intercal D\boldsymbol{\chi}}{\left\|D^{1/2}\mathbf{1}\right\|} \geqslant (1-\gamma)\sqrt{2nd}-2\gamma\sqrt{2nd} \\ &= (1-3\gamma)\sqrt{2nd} > (1/2)\sqrt{2nd}\,, \end{aligned} \tag{4.17}$$

where: in the first equality we used (4.12) and the triangular inequality, in the second inequality we used (4.6) and (4.14), while the fourth inequality follows from the hypothesis $\gamma\leqslant 1/10$.
Now, let us write $\mathbf{y}$ as a linear combination of the orthonormal eigenvectors of N

$$\mathbf{y}=\beta_2\mathbf{w}_2+\cdots+\beta_{2n}\mathbf{w}_{2n}$$

(recall that $\mathbf{y}^\intercal\mathbf{w}_1=0$ by definition of $\mathbf{y}$ in (4.12)). By comparing (4.16) and (4.17),

$$(24\gamma)^2\|\mathbf{y}\|^2 \geqslant \|N\mathbf{y}-\nu\mathbf{y}\|^2 = \left\|\sum_{i=2}^{2n}(\lambda_i-\nu)\beta_i\mathbf{w}_i\right\|^2 = \sum_{i=2}^{2n}(\lambda_i-\nu)^2\beta_i^2\,. \tag{4.18}$$

Moreover, from the hypothesis $\lambda_3 < \nu$,

$$\sum_{i=2}^{2n}(\lambda_i-\nu)^2\beta_i^2 \geqslant \sum_{i=3}^{2n}(\lambda_i-\nu)^2\beta_i^2 \geqslant (\lambda_3-\nu)^2\sum_{i=3}^{2n}\beta_i^2 = (\lambda_3-\nu)^2\|\mathbf{y}-\beta_2\mathbf{w}_2\|^2. \tag{4.19}$$

Thus, by combining (4.18) and (4.19),

$$\|\mathbf{y}-\beta_2\mathbf{w}_2\| \leqslant \frac{24\,\gamma}{\nu-\lambda_3}\|\mathbf{y}\|, \tag{4.20}$$

where $\beta_2 = \mathbf{y}^\intercal\mathbf{w}_2 = \left(D^{1/2}\boldsymbol{\chi}\right)^\intercal\mathbf{w}_2$.
Then, since $\mathbf{y}$ is the projection of $D^{\frac{1}{2}}\boldsymbol{\chi}$ on $D^{\frac{1}{2}}\mathbf{1}$,

$$\|\mathbf{y}\| \leqslant \|D^{\frac{1}{2}}\boldsymbol{\chi}\| \leqslant 2\sqrt{2nd}\,. \tag{4.21}$$

Finally, from (4.13),

$$\begin{aligned}\left\|D^{1/2}\boldsymbol{\chi}-\beta_2\mathbf{w}_2\right\| &\leqslant \frac{|\mathbf{1}^\intercal D\boldsymbol{\chi}|}{\|D^{1/2}\mathbf{1}\|} + \|\mathbf{y}-\beta_2\mathbf{w}_2\| \leqslant 4\gamma\sqrt{2nd} + \frac{24\,\gamma}{\nu-\lambda_3}\|\mathbf{y}\| \\ &\leqslant 4\,\gamma\sqrt{2nd} + \frac{48\,\gamma}{\nu-\lambda_3}\sqrt{2nd} \leqslant \frac{52\,\gamma}{\nu-\lambda_3}\sqrt{2nd}\,,\end{aligned}$$

where the second inequality follows from (4.14) and (4.20), in the third inequality we use (4.21), and the last inequality follows since $\nu-\lambda_3 < 2$. ∎

The next lemma essentially states that the second eigenvector $\mathbf{v}_2$ of P is almost parallel to $\boldsymbol{\chi}$, up to an additive "error" which is comparatively small in norm.

LEMMA 4.6. *Let G be a connected $(2n,d,b,\gamma)$-clustered graph with $\gamma \leqslant 1/10$ and let $\mathbf{x} = \sum_i \tilde{\alpha}_i\mathbf{v}_i$ be the decomposition of an arbitrary initial vector $\mathbf{x}$, according to the basis $\{\mathbf{v}_i\}_i$ of eigenvectors of the transition matrix P. If $\lambda < \nu$ then*

$$\tilde{\alpha}_2\mathbf{v}_2 = \alpha_2\left(\boldsymbol{\chi}+\mathbf{z}\right)\,, \tag{4.22}$$

where we set

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

and

$$\|\mathbf{z}\| \leqslant \frac{104\,\gamma}{\nu-\lambda_3}\sqrt{2n}\,.$$

Proof. We can write

$$\tilde{\alpha}_2\mathbf{v}_2 = \frac{\tilde{\alpha}_2}{\beta_2}\beta_2\mathbf{v}_2 = \frac{\tilde{\alpha}_2}{\beta_2}\left(\boldsymbol{\chi}+(\beta_2\mathbf{v}_2-\boldsymbol{\chi})\right) = \alpha_2(\boldsymbol{\chi}+\mathbf{z})$$

with $\mathbf{z} = \beta_2\mathbf{v}_2-\boldsymbol{\chi}$. As for the norm of $\mathbf{z}$ observe that

$$D^{1/2}\mathbf{z} = \beta_2 D^{1/2}\mathbf{v}_2 - D^{1/2}\boldsymbol{\chi} = \beta_2\mathbf{w}_2 - D^{1/2}\boldsymbol{\chi}\,.$$

Thus, from Lemma 4.5

$$\|\mathbf{z}\| = \|D^{-1/2}D^{1/2}\mathbf{z}\| \leqslant \|D^{-1/2}\|\cdot\|D^{1/2}\mathbf{z}\| \leqslant \frac{2}{\sqrt{d}}\cdot\frac{52\,\gamma}{\nu-\lambda_3}\sqrt{2nd} = \frac{104\,\gamma}{\nu-\lambda_3}\sqrt{2n}\,.$$ ∎

The above lemma allows us to generalize our approach to achieve efficient, weak reconstruction in clustered, almost regular graphs.

THEOREM 4.7 (Weak reconstruction). *Let G be a connected $(2n, d, b, \gamma)$-clustered graph such that: $\gamma \leqslant (\nu - \lambda_3)/208$, $\lambda < \nu$, and $\lambda_2 \geqslant (1+\delta)\lambda$ for an arbitrarily small constant $\delta > 0$. Then the* AVERAGING *protocol produces an $\mathcal{O}(\gamma^2/(\nu-\lambda_3)^2)$-weak reconstruction within $\mathcal{O}(\log n)$ rounds, w.h.p.*[10] *Moreover, the overall number of messages exchanged by the protocol (until its completion time) is $\mathcal{O}(m \log n)$ and each message has size $\Theta(\log n)$, w.h.p.*

Proof. From (4.3) in Lemma 4.1

$$\begin{aligned} x^{(t-1)}(u) - x^{(t)}(u) &= \tilde{\alpha}_2 \lambda_2^{t-1}(1-\lambda_2)v_2(u) + e^{(t-1)}(u) - e^{(t)}(u) \\ &= \alpha_2 \lambda_2^{t-1}(1-\lambda_2)\left(\chi(u) + z(u)\right) + e^{(t-1)}(u) - e^{(t)}(u), \end{aligned}$$

where the second equality follows from Lemma 4.6. Corollary 4.2 (in particular, (4.5)) implies that, for every initial state $\mathbf{x}$ such that $\tilde{\alpha}_2 \neq 0$ and for every node u such that $v_2(u) \neq 0$,

$$\mathtt{sgn}\left(x^{(t-1)}(u) - x^{(t)}(u)\right) = \mathtt{sgn}\left(\tilde{\alpha}_2 v_2(u)\right), \tag{4.23}$$

whenever $t - 1 \geqslant \dfrac{\log\left(\frac{2\sqrt{2n(1+\gamma)/(1-\gamma)}}{|\tilde{\alpha}_2|\cdot|v_2(u)|(1-\lambda_2)}\right)}{\log(\lambda_2/\lambda)}$.

We next prove the following three claims:

1. For constant $\hat{c} = \left(\frac{208\gamma}{\nu-\lambda_3}\right)^2 < 1$, a subset S of nodes, with $|S| \geqslant 2(1-\hat{c})n$, exists such that for each node $u \in S$,
$$|v_2(u)| \geqslant \frac{\varepsilon}{\sqrt{2nd}} \quad \text{and} \quad \mathtt{sgn}\left(\tilde{\alpha}_2 v_2(u)\right) = \mathtt{sgn}\left(\alpha_2\right)\mathtt{sgn}\left(\boldsymbol{\chi}\right),$$
where $\varepsilon \leqslant 1/2$ is a positive constant.
2. $|\tilde{\alpha}_2| \geqslant \frac{1}{\sqrt{2nd}}$, w.h.p. Thus, for the subset S in the previous claim the AVERAGING protocol produces the right reconstruction.
3. $1 - \lambda_2 \geqslant \frac{1}{2((1+\gamma)dn)^2}$.

Together with (4.23), these claims imply that weak reconstruction is achieved within $\mathcal{O}(\log n)$ rounds, w.h.p.

Proof of Claim 1. Recall from Lemma 4.6 that $\|\mathbf{z}\| \leqslant \tilde{c}\sqrt{2n}$, where $\tilde{c} = \frac{112\gamma}{\nu-\lambda_3}$. Hence, if we set $S = \{u \in V : |z(u)| < 1 - \hat{\varepsilon}\}$, for some $0 < \hat{\varepsilon} < 1/2$, and $h = |V \setminus S|$ we get

$$(1-\hat{\varepsilon})^2 h \leqslant \|\mathbf{z}\|^2 \leqslant 2\tilde{c}^2 n\,.$$

Thus, $h \leqslant 2\hat{c}n$, where

$$\hat{c} = \frac{\tilde{c}^2}{(1-\hat{\varepsilon})^2} = \left(\frac{112\gamma}{(1-\hat{\varepsilon})(\nu-\lambda_3)}\right)^2 \leqslant \left(\frac{208\gamma}{\nu-\lambda_3}\right)^2.$$

Finally, observe that $|\beta_2 v_2(u) - z(u)| = 1$ for all nodes u, since $\beta_2 \mathbf{v}_2 = \boldsymbol{\chi} + \mathbf{z}$. Hence, $|\beta_2 v_2(u)| > \hat{\varepsilon}$ for each node u with $|z(u)| < 1 - \hat{\varepsilon}$. Thus, for all nodes $u \in S$,

[10] Consistently, Theorem 3.3 is a special case of this one when $\gamma = 0$.

(a) $|v_2(u)| \geqslant \frac{\hat{\varepsilon}}{|\beta_2|} \geqslant \frac{\hat{\varepsilon}}{\sqrt{(1+\gamma)2nd}} = \frac{\varepsilon}{\sqrt{2nd}}$, where $\varepsilon = \hat{\varepsilon}/\sqrt{1+\gamma} < \hat{\varepsilon} \leqslant 1/2$.

(b) $\mathtt{sgn}\left(\tilde{\alpha}_2 v_2(u)\right) = \mathtt{sgn}\left(\alpha_2(\chi(u) + z(u))\right) = \mathtt{sgn}\left(\alpha_2\right)\mathtt{sgn}\left(\chi(u)\right)$ where in the first equality we used Lemma 4.6 and in the second one the fact that $|z(u)| \leqslant 1 - \hat{\varepsilon} < 1$.

Proof of Claim 2. We rely on Lemma B.2. In detail, we apply Lemma B.2 with $\mathbf{y} = D(\boldsymbol{\chi} + \mathbf{z}) = \beta_2 D\mathbf{v}_2$. By definition of S in the proof of the previous claim, $|y(u)| = |D(\boldsymbol{\chi} + \mathbf{z})(u)| \geqslant \hat{\varepsilon} d_u$ for all $u \in S$, since $|z(u)| \leqslant 1 - \hat{\varepsilon}$. For the same reason, we have $|y(u)| \leqslant (2-\hat{\varepsilon})d_u$. These considerations and the fact that G is a $(2n, d, b, \gamma)$-clustered graph (see Definition 4.3) imply that, for every $u \in S$ we have $(1-\gamma)\hat{\varepsilon}d \leqslant |y(u)| \leqslant (1+\gamma)(2-\hat{\varepsilon})d$. So, we can apply Lemma B.2 with the following setting: $\mathbf{y} = D(\boldsymbol{\chi}+\mathbf{z})$, S is defined as in Claim 1 above, $r = (1-\gamma)\hat{\varepsilon}d$, $c = \frac{(1+\gamma)(2-\hat{\varepsilon})}{(1-\gamma)\hat{\varepsilon}}$, $k = 1/(1-\hat{c})$, $km = 2n$, and $\delta = \sqrt{\frac{1-\hat{c}}{2n}}$. We then obtain

$$\begin{aligned}\mathbf{P}\left(|\langle(1/\sqrt{km})\mathbf{y}, \mathbf{x}\rangle| \leqslant \delta\right) &= \mathbf{P}\left(|\langle \mathbf{y}, \mathbf{x}\rangle| \leqslant \sqrt{\frac{1-\hat{c}}{2n}}\sqrt{\frac{2n}{1-\hat{c}}}\right) = \mathbf{P}\left(|\langle \mathbf{y}, \mathbf{x}\rangle| \leqslant 1\right) \\ &\leqslant \sqrt{\frac{2k}{\pi}} \cdot \frac{\delta}{r} + \frac{4c}{\sqrt{m}} = \mathcal{O}\left(\frac{1}{d\sqrt{n}}\right) + \mathcal{O}\left(\frac{1}{n}\right) = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

Next, we have:

$$|\tilde{\alpha}_2| = |\mathbf{v}_2^\intercal D\mathbf{x}| = \frac{|(\boldsymbol{\chi} + \mathbf{z})^\intercal D\mathbf{x}|}{|\beta_2|},$$

where the first equality follows from Claim 1 of Lemma 4.1, while the second equality follows from (4.22). As for β_2, recall that

$$|\beta_2| = |\mathbf{w}_2^\intercal D^{1/2}\boldsymbol{\chi}| \leqslant \|D^{1/2}\boldsymbol{\chi}\| = \sqrt{\sum_{i=1}^{2n} d_i} \leqslant \sqrt{(1+\gamma)nd},$$

where the second step follows from Cauchy inequality and $\|\mathbf{w}_2\| = 1$, while the last inequality follows from the definition of $(2n, d, b, \gamma)$-clustered graphs. As a consequence, we have $|\tilde{\alpha}_2| \geqslant \frac{1}{\sqrt{(1+\gamma)nd}}$, w.h.p. over the randomness of $\mathbf{x}$.

Proof of Claim 3. The third claim follows directly from Cheeger's inequalities and connectedness of the graph. In particular, from (4.7) we have $1 - \lambda_2 \geqslant \frac{\Phi^2}{2}$. Connectedness in turn implies that, for every subset S of the vertices, the corresponding cut is crossed by at least one edge, so that its conductance is at least $1/vol(S)$. These considerations immediately imply that $\phi_G \geqslant \frac{1}{(1+\gamma)dn}$, whenever G is $(2n, d, b, \gamma)$-clustered. This proves the third claim and thus the theorem.

Finally, observe that we can perform the same analysis of the communication cost we made in the proof of Theorem 3.3, thus leading to the same bounds. ∎

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

4.2. Stochastic block model. In this subsection, we prove that graphs sampled according to the stochastic block model $\mathcal{G}_{2n,p,q}$, satisfy the hypotheses of Theorem 4.7, w.h.p. and, thus, the AVERAGING protocol efficiently produces a good reconstruction. In what follows, we will often use the following parameters of the model:

$$\text{expected inner degree } \ a = pn\,, \quad \text{expected outer degree } b = qn, \ \text{ and } \ d = a + b\,.$$

We need two preliminary lemmas. In the first one, since G is not regular and random, we derive some spectral properties of its adjacency matrix A by considering a "more tractable" matrix, namely the expected matrix

$$B = \mathbf{E}\,[A] = \begin{pmatrix} pJ & qJ \\ qJ & pJ \end{pmatrix}, \tag{4.24}$$

where J is the matrix with all entries equal to 1, so $B_{i,j}$ is the probability that the edge (i,j) exists in a random graph $G \sim \mathcal{G}_{2n,p,q}$. The matrix B has a very simple spectral structure, summarized in the following fact, which will be useful in the following.

FACT 2. *If B is defined as in (4.24), then $\mathbf{1}$ is an eigenvector of eigenvalue d, $\boldsymbol{\chi}$ is an eigenvector of eigenvalue $a-b$, and all vectors orthogonal to $\mathbf{1}$ and to $\boldsymbol{\chi}$ are eigenvectors of eigenvalue 0.*

In detail, the first lemma claims that G is likely to have an adjacency matrix A close to B in spectral norm.

LEMMA 4.8. *Let A be the adjacency matrix of a random graph sampled from $\mathcal{G}_{2n,p,q}$ with $d > 5\log n$. Then there is a large enough absolute constant $c_1 > 0$ such that, w.h.p.,*

$$\|A - B\| \leqslant c_1\sqrt{d}\,.$$

Proof. The lemma directly follows from Theorem 2.1 in [47] with $d' = 2d$ and the observation that, from the Chernoff bounds, all degrees are smaller than $2d$, w.h.p. □

The second lemma states that every clustered graph whose adjacency matrix is close to B has the properties required in the analysis of the AVERAGING dynamics.

LEMMA 4.9. *Let G be a $(2n,d,b,\gamma)$-clustered graph such that: (i) $\nu = 1 - 2b/d > 12\gamma$ and (ii) the adjacency matrix A of G satisfies $\|A - B\| \leqslant \gamma d$. Then*

$$\lambda \leqslant 4\gamma \quad \text{and} \quad \lambda_2 \geqslant 2\lambda \geqslant 2\lambda_3\,.$$

Proof. Recall Fact 2. In order to understand the eigenvalues and eigenvectors of N, and hence the eigenvalues and eigenvectors of P, we first prove that A approximates B and that N approximates $(1/d)A$, namely $\|dN - A\| \leqslant 3\gamma d$.

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

$$\|D^{1/2} - \sqrt{d}I\| \leqslant \gamma\sqrt{d}\,. \tag{4.25}$$

We know that $\|D\| \leqslant d + \gamma d < 2d$ and that $\|N\| = 1$, so from (4.25)

$$\begin{aligned} \|A - dN\| &= \|D^{1/2}ND^{1/2} - dN\| \leqslant \|D^{1/2}ND^{1/2} - \sqrt{d}ND^{1/2}\| + \|\sqrt{d}ND^{1/2} - dN\| \\ &= \|(D^{1/2} - \sqrt{d}I)\cdot ND^{1/2}\| + \|\sqrt{d}N\cdot(D^{1/2} - \sqrt{d}I)\| \\ &\leqslant \|D^{1/2} - \sqrt{d}I\|\cdot\|N\|\cdot\|D^{1/2}\| + \sqrt{d}\cdot\|N\|\cdot\|D^{1/2} - \sqrt{d}I\| \leqslant 3\gamma d\,. \end{aligned} \tag{4.26}$$

By using the triangle inequality and (4.26),

$$\|N - (1/d)B\| \leqslant \|N - (1/d)A\| + (1/d) \cdot \|A - B\| \leqslant 4\gamma\,. \tag{4.27}$$

Finally, we use Theorem A.2 (see Appendix A), which is a standard fact in matrix approximation theory: If two real symmetric matrices are close in spectral norm then their eigenvalues are close. From (4.27) and Fact 2,

$$|\lambda_i| = |\lambda_i - 0| \leqslant \|N - (1/d)B\| \leqslant 4\gamma, \text{ for every } i \in \{3, \dots, 2n\}\,. \tag{4.28}$$

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

$$|\lambda_2 - \nu| \leqslant \|N - (1/d)B\| \leqslant 4\gamma\,. \tag{4.29}$$

Now, (4.28) and (4.29) imply that $\lambda \leqslant 4\gamma$ and $\lambda_2 \geqslant \nu - 4\gamma$, respectively. The thesis then follows from the hypothesis $\nu > 12\gamma$. ∎

We can now prove the main result of this subsection.

THEOREM 4.10. *Let $G \sim \mathcal{G}_{2n,p,q}$ with $a - b > 72\sqrt{d \log n}$. Then w.h.p.: i) G is $(2n, d, b, 6\sqrt{\log n/d})$-clustered,* and *ii) $\lambda \leqslant \min\left\{\lambda_2/2\,,\, 24\sqrt{\log n/d}\right\}$.*

Proof. From the hypothesis $d > a - b > 72^2 \log n$, we get Claim (i) (with probability at least $1 - n^{-1}$) from an easy application of the Chernoff bound with any degree deviation $\gamma \geqslant 6\sqrt{(\log n)/d}$. As for Claim (ii), the hypothesis above and Lemma 4.8 allow[11] to apply Lemma 4.9 with $\gamma = 6\sqrt{(\log n)/d}$. This implies that $\lambda \leqslant 24\sqrt{(\log n)/d}$ and, moreover, $\lambda \leqslant \lambda_2/2$. ∎

By combining Theorem 4.10 and Theorem 4.7, we achieve weak reconstruction for the stochastic block model.

COROLLARY 4.11. *Let $G \sim \mathcal{G}_{2n,p,q}$ with $a - b > 1368\sqrt{d \log n}$ and $b = \Omega(\log n)$. Then the* AVERAGING *protocol produces an $\mathcal{O}(d \log n/(a-b)^2)$-weak reconstruction within $\mathcal{O}(\log n)$ rounds, w.h.p.*

Proof. From Theorem 4.10 we get that w.h.p. G is $(2n, d, b, \gamma)$-clustered with $\gamma = 6\sqrt{\log n/d}$, $\lambda \leqslant 4\gamma$ and $\lambda_2 \geqslant (1 + \delta)\lambda_3$ with $\delta = 1$. Given the hypotheses on a and b, we also have that the graph is connected, w.h.p. Moreover, from the hypothesis of the corollary, $d\nu = a - b > 1368\sqrt{d \log n}$. Hence,

$$\frac{\gamma}{\nu - \lambda_3} = \frac{d\gamma}{d\nu - d\lambda_3} < \frac{6\sqrt{d \log n}}{1368\sqrt{d \log n} - 24\sqrt{d \log n}} = \frac{1}{224}\,,$$

where the last inequality follows from the hypothesis $a - b > 1368\sqrt{d \log n}$ and since $\lambda_3 \leqslant \lambda \leqslant 4\gamma$. We can thus apply Theorem 4.7 which guarantees that, w.h.p., the AVERAGING protocol achieves a ε-weak reconstruction in $(2n, d, b, \gamma)$-clustered graphs, with

$$\varepsilon = \left(\frac{224\gamma}{\nu - \lambda_3}\right)^2 < 1\,,$$

from the derivations above and Claim 1 in the proof of Theorem 4.7. This concludes the proof. ∎

[11] Notice that Lemma 4.8 would work even for a smaller γ, i.e., for $\gamma = \Omega(1/\sqrt{d})$: However, in this proof, this stronger bound is not useful since, in order to get the first claim of the theorem, we need $\gamma \geqslant 6\sqrt{(\log n)/d}$. The stronger bound will be instead useful in the next subsection.

4.3. Improved analysis for the stochastic block model. In this section we assume again that the underlying graph G is sampled from $\mathcal{G}_{2n,p,q}$, and we recall that $a = pn, b = qn$ and $d = a + b$. In Lemma 4.10 we have shown that, when $(a-b) > c\sqrt{d \log n}$ for a suitable absolute constant c, a graph sampled according to $\mathcal{G}_{2n,p,q}$ satisfies the hypotheses of Theorem 4.7, w.h.p. In this setting, the AVERAGING protocol thus achieves weak reconstruction in $\mathcal{O}(\log n)$ rounds.

It is known that, when $a - b < 2\sqrt{d}$, weak recovery is impossible for any algorithm [59, 53, 57], including centralized algorithms of arbitrarily high running time. How close to this information-theoretic bound does the AVERAGING protocol get? In this section we provide an analysis specialized to the stochastic model showing that the AVERAGING protocol achieves weak reconstruction provided that $a - b > c \cdot \sqrt{d}$ where c is a sufficiently large constant, and provided that the average degree d is at least logarithmic.

THEOREM 4.12. *There is an absolute constant c such that the following holds. Let G be sampled from $\mathcal{G}_{2n,\frac{a}{n},\frac{b}{n}}$ with $9 \log n \leqslant d < n^{\frac{1}{4}}$, $b > \log n$, and $a - b > c \cdot \sqrt{d}$. Then the* AVERAGING *protocol produces an $\mathcal{O}(d/(a-b)^2)$-weak reconstruction within $\mathcal{O}(\log n)$ rounds, w.h.p.*

To achieve this improved analysis, we first show, in Lemma 4.13 that, under the assumptions of the above theorem, w.h.p

$$\|N - B/d\| \leqslant \mathcal{O}\left(\frac{1}{\sqrt{d}}\right),$$

where we recall: B is the expectation of the adjacency matrix A of $G \sim \mathcal{G}_{2n,\frac{a}{n},\frac{b}{n}}$ and $N = D^{-1/2}AD^{-1/2}$ is the normalized adjacency matrix.

As discussed in the previous section, the matrix B has only two non-zero eigenvalues, d and $a - b$, and the eigenvector of the second eigenvalue is the indicator of the cut. In Lemma 4.15 we use the Davis-Kahan theorem (see Theorem A.4) to argue that, if the distance between N and B/d in spectral norm is smaller than $(a-b)/d$, then there is a gap between the second and the third eigenvalues of N, and the eigenvector of the second eigenvalue of N is close to the indicator of the cut. Finally, we show that these spectral conditions on N suffice for the AVERAGING dynamics to achieve weak reconstruction, thus proving Theorem 4.12.

4.3.1. Concentration of the normalized adjacency matrix. In this section we prove the following concentration result for the normalized adjacency matrix of a graph sampled from $\mathcal{G}_{2n,\frac{a}{n},\frac{b}{n}}$.

LEMMA 4.13. *There is an absolute constant c_2 such that, for every $9 \log n < d < n^{\frac{1}{4}}$, w.h.p.*

$$\|N - B/d\| \leqslant c_2/\sqrt{d}\,.$$

In order to prove that $\|N - B/d\| \leqslant \mathcal{O}(1/\sqrt{d})$, our starting point is Lemma 4.8, ensuring that $\|A/d - B/d\| \leqslant \mathcal{O}(1/\sqrt{d})$, w.h.p. Thanks to the above result, to prove Lemma 4.13, it remains to argue that $\|N - A/d\| \leqslant \mathcal{O}(1/\sqrt{d})$, w.h.p. To this aim, we would like to use the fact that $\sqrt{d} \cdot D^{-1/2}$ is close to the identity, so that $N = D^{-1/2}AD^{-1/2}$ and A/d are also be close to each other.

Unfortunately, arguing about the spectral norm of $\sqrt{d}D^{-1/2} - I$ does not work, because it introduces a dependency on the maximum degree of the graph, and the best bound that we would get in this way is $\|A/d - B/d\| \leqslant \mathcal{O}\left(\sqrt{\frac{\log n}{d}}\right)$, which would

not lead to an improvement over the analysis of the previous section. Instead, we are going to use again the fact that A and B are close, and bound $\|D^{-1/2}AD^{-1/2} - A/d\|$ in terms of $\|D^{-1/2}BD^{-1/2} - B/d\|$. Since B only acts on the span of $\mathbf{1}$ and of $\boldsymbol{\chi}$, which are very smooth vectors in which all coordinates are ± 1, the latter term does not depend on the maximum degree but can be bounded in terms of the quantity $\sum_v (\sqrt{d_v} - \sqrt{d})^2$. We begin by bounding the latter quantity.

LEMMA 4.14. *If $5 \log n < d < n^{\frac{1}{4}}$ then w.h.p.*

$$\sum_{v \in V} \left(\sqrt{d} - \sqrt{d_v}\right)^2 \leqslant 5n\,.$$

Proof. We first remark that each degree d_v has the distribution of a sum of n Bernoulli random variables of expectation p plus a sum of n Bernoulli random variables of expectation q. Thus, each d_v has expectation $\mathbf{E}[d_v] = d$ and variance $\mathrm{Var}(d_v) \leqslant d$.

Let $e_{u,v}$ be the binary random variable that is 1 iff the edge (u, v) is included in the graph and define also $d_v^{(u)}$ as the sum of all random variables $e_{v,v'}$ incident on v except for $e_{v,u}$.

As for the sum of the variables $\left(\sqrt{d} - \sqrt{d_v}\right)^2$,

$$\sum_{v \in V} \left(\sqrt{d} - \sqrt{d_v}\right)^2 = 2dn + \sum_{v \in V} d_v - 2\sqrt{d} \cdot \sum_{v \in V} \sqrt{d_v}. \tag{4.30}$$

From the Chernoff bound (and the hypothesis $d < n^{\frac{1}{4}}$), w.h.p.

$$\sum_{v \in V} d_v \leqslant 2dn + n\,. \tag{4.31}$$

We will next prove that, w.h.p.,

$$\sum_{v \in V} \sqrt{d_v} \geqslant 2n\sqrt{d} - 2\frac{n}{\sqrt{d}}\,. \tag{4.32}$$

Observe that, by using (4.31) and (4.32) in (4.30),

$$\sum_{v \in V} \left(\sqrt{d} - \sqrt{d_v}\right)^2 \leqslant 5n\,,$$

which concludes the proof of the lemma. So, we will now prove that (4.32) holds (w.h.p.). Observe that if $x \geqslant 0$, then

$$\sqrt{x} \geqslant 1 + \frac{x-1}{2} - \frac{(x-1)^2}{2}$$

so that if X is a non-negative random variable of expectation 1 then[12]

$$\mathbf{E}\left[\sqrt{X}\right] \geqslant 1 - \frac{\mathrm{Var}(X)}{2}\,.$$

[12] This argument is due to Ori Gurel-Gurevich (see [38]).

By applying the above inequality to d_v/d we get

$$\mathbf{E}\left[\sqrt{\frac{d_v}{d}}\right] \geqslant 1 - \frac{\mathrm{Var}\left(\frac{d_v}{d}\right)}{2} = 1 - \frac{\mathrm{Var}\left(d_v\right)}{2d^2} \geqslant 1 - \frac{1}{2d}$$

and

$$\mathbf{E}\left[\sqrt{d_v}\right] \geqslant \sqrt{d} - \frac{1}{2\sqrt{d}}\,. \tag{4.33}$$

We will now show that $\sum_{v\in V}\sqrt{d_v}$ is concentrated around its expectation by using Chebyshev's inequality[13]. In order to do that, we will bound their covariance as

$$\mathbf{E}\left[\sqrt{d_v d_u}\right] - \mathbf{E}\left[\sqrt{d_v}\right]\mathbf{E}\left[\sqrt{d_u}\right] \leqslant \frac{8d^2}{n}\,.$$

With a slight abuse of notation, in what follows we use $\mathbf{P}\left(e_{v,u}\right)$ to denote $\mathbf{P}\left(e_{v,u}=1\right)$. By the law of total probability

$$\mathbf{E}\left[\sqrt{d_v}\right] = \mathbf{P}\left(e_{v,u}\right)\mathbf{E}\left[\sqrt{d_v^{(u)}+1}\right] + (1-\mathbf{P}\left(e_{v,u}\right))\mathbf{E}\left[\sqrt{d_v^{(u)}}\right]$$

and

$$\mathbf{E}\left[\sqrt{d_u d_v}\right] = \mathbf{P}\left(e_{v,u}\right)\mathbf{E}\left[\sqrt{d_v^{(u)}+1}\right]\mathbf{E}\left[\sqrt{d_u^{(v)}+1}\right] + (1-\mathbf{P}\left(e_{v,u}\right))\mathbf{E}\left[\sqrt{d_u^{(v)}}\right]\mathbf{E}\left[\sqrt{d_v^{(u)}}\right].$$

Then the last two equations above imply that

$$\begin{aligned}
&\mathbf{E}\left[\sqrt{d_v d_u}\right] - \mathbf{E}\left[\sqrt{d_v}\right]\mathbf{E}\left[\sqrt{d_u}\right] \\
&= \mathbf{P}\left(e_{v,u}\right)\mathbf{E}\left[\sqrt{d_v^{(u)}+1}\right]\mathbf{E}\left[\sqrt{d_u^{(v)}+1}\right] + (1-\mathbf{P}\left(e_{v,u}\right))\mathbf{E}\left[\sqrt{d_u^{(v)}}\right]\mathbf{E}\left[\sqrt{d_v^{(u)}}\right] \\
&\quad - \mathbf{P}\left(e_{v,u}\right)^2\mathbf{E}\left[\sqrt{d_u^{(v)}+1}\right]\mathbf{E}\left[\sqrt{d_v^{(u)}+1}\right] \\
&\quad - \mathbf{P}\left(e_{v,u}\right)\left(1-\mathbf{P}\left(e_{v,u}\right)\right)\mathbf{E}\left[\sqrt{d_u^{(v)}}\right]\mathbf{E}\left[\sqrt{d_v^{(u)}+1}\right] \\
&\quad - \mathbf{P}\left(e_{v,u}\right)\left(1-\mathbf{P}\left(e_{v,u}\right)\right)\mathbf{E}\left[\sqrt{d_u^{(v)}+1}\right]\mathbf{E}\left[\sqrt{d_v^{(u)}}\right] \\
&\quad - \left(1-\mathbf{P}\left(e_{v,u}\right)\right)^2\mathbf{E}\left[\sqrt{d_u^{(v)}}\right]\mathbf{E}\left[\sqrt{d_v^{(u)}}\right] \\
&\leqslant p(1-p)\Big(\mathbf{E}\left[\sqrt{d_v^{(u)}+1}\right]\mathbf{E}\left[\sqrt{d_u^{(v)}+1}\right] + \mathbf{E}\left[\sqrt{d_u^{(v)}}\right]\mathbf{E}\left[\sqrt{d_v^{(u)}}\right]\Big) \leqslant \frac{4d^2}{n}\,,
\end{aligned} \tag{4.34}$$

where in the last inequality we used that $\mathbf{E}\left[\sqrt{d_v^{(u)}+1}\right] < \sqrt{2d}$, w.h.p., (by the Chernoff bound), and that $\max\{q(1-q), p(1-p)\} = p(1-p) < p < d/n$ (from the hypothesis $d < n^{\frac{1}{4}}$). Then, (4.34) implies that

$$\mathrm{Var}\left(\sum_{v\in V}\sqrt{d_v}\right) \leqslant 2nd + 16d^2 n = \mathcal{O}\left(\frac{n^2}{dn^{\frac{1}{4}}}\right). \tag{4.35}$$

[13] A stronger bound which does not require the hypothesis $d \leqslant n^{1/4}$ may be obtained with some concentration techniques compatible with the stochastic dependence among the $\sqrt{d_v}$s.

Finally, by combining (4.35) and (4.33) with Chebyshev's inequality,

$$\begin{aligned}\mathbf{P}\left(\sum_{v\in V}\sqrt{d_v} < 2n\sqrt{d} - 2\frac{n}{\sqrt{d}}\right) &\leqslant \mathbf{P}\left(\Big|\sum_{v\in V}\sqrt{d_v} - \mathbf{E}\Big[\sum_{v\in V}\sqrt{d_v}\Big]\Big| > \frac{n}{\sqrt{d}}\right)\\ &= \mathcal{O}\left(\frac{1}{n^{1/4}}\right).\end{aligned}$$

∎

We now can prove Lemma 4.13.

Proof of Lemma 4.13. The main idea is to use the triangle inequality to upper bound $\|N - B/d\|$ in terms of $\|A - B\|$ and $\|B - dD^{-1/2}BD^{-1/2}\|$. The former can be bounded with Lemma 4.8 while the latter can be bounded by bounding $\|\sqrt{d}\mathbf{1} - D^{1/2}\mathbf{1}\|$ and $\|\sqrt{d}\boldsymbol{\chi} - D^{1/2}\boldsymbol{\chi}\|$ using Lemma 4.14.

We first write

$$\|N - B/d\| \leqslant \left\|N - D^{-\frac{1}{2}}BD^{-\frac{1}{2}}\right\| + \left\|D^{-\frac{1}{2}}BD^{-\frac{1}{2}} - B/d\right\|. \tag{4.36}$$

We can bound the first term on the right-hand side as

$$\begin{aligned}\left\|N - D^{-\frac{1}{2}}BD^{-\frac{1}{2}}\right\| &= \left\|D^{-\frac{1}{2}}(A-B)D^{-\frac{1}{2}}\right\|\\ &\leqslant \|D^{-\frac{1}{2}}\| \cdot \|A - B\| \cdot \|D^{-\frac{1}{2}}\|. \qquad (4.37)\end{aligned}$$

A simple application of the Chernoff bound and the union bound shows that w.h.p.

$$\|D^{-1/2}\| \leqslant \sqrt{\frac{2}{d}}. \tag{4.38}$$

Indeed, by definition $\|D^{-1/2}\| = 1/\min_i\{\sqrt{d_i}\}$. For every $i = 1, \ldots, 2n$, $\mathbf{E}[d_i] = d$. Therefore, from the Lemma's hypothesis that $d > 9\log n$ and from the fact that the edges are generated independently, a straightforward application of Chernoff's bounds yields:

$$\mathbf{P}\left(d_i < \frac{d}{2}\right) \leqslant e^{-\frac{d}{8}} \leqslant \frac{1}{n^{9/8}}.$$

Applying a union bound on the nodes yields (4.38).

The assumptions of the lemma together with Lemma 4.8 imply that

$$\|A - B\| \leqslant \mathcal{O}(\sqrt{d}).$$

By applying the above bound and (4.38) in (4.37), we conclude

$$\left\|N - D^{-\frac{1}{2}}BD^{-\frac{1}{2}}\right\| \leqslant \mathcal{O}\left(\frac{1}{\sqrt{d}}\right).$$

Regarding the second error term in (4.36), we write

$$\begin{aligned}\left\|D^{-\frac{1}{2}}BD^{-\frac{1}{2}} - B/d\right\| &\leqslant \left\|D^{-\frac{1}{2}}BD^{-\frac{1}{2}} - D^{-\frac{1}{2}}\frac{B}{\sqrt{d}}\right\| + \left\|D^{-\frac{1}{2}}\frac{B}{\sqrt{d}} - B/d\right\|\\ &= \left\|D^{-\frac{1}{2}}\left(B - \frac{B}{\sqrt{d}}D^{\frac{1}{2}}\right)D^{-\frac{1}{2}}\right\| + \left\|\frac{1}{\sqrt{d}}D^{-\frac{1}{2}}\left(B - D^{\frac{1}{2}}\frac{B}{\sqrt{d}}\right)\right\|\\ &\leqslant \|D^{-\frac{1}{2}}\|^2 \cdot \left\|B - \frac{B}{\sqrt{d}}D^{\frac{1}{2}}\right\| + \frac{\|D^{-\frac{1}{2}}\|}{\sqrt{d}} \cdot \left\|B - D^{\frac{1}{2}}\frac{B}{\sqrt{d}}\right\|\\ &\leqslant \mathcal{O}\left(\frac{1}{d}\right) \cdot \left\|B - D^{\frac{1}{2}}\frac{B}{\sqrt{d}}\right\|, \qquad (4.39)\end{aligned}$$

where, in the second-to-last step we used the fact that $\|M\| = \|M^\intercal\|$ for every matrix M and the fact that B is symmetric. The last step holds w.h.p.

Recall that

$$B = \frac{d}{2n}\mathbf{1}\mathbf{1}^\intercal + \frac{a-b}{2n}\boldsymbol{\chi}\boldsymbol{\chi}^\intercal\,.$$

So

$$\begin{aligned}\left\|B - D^{\frac{1}{2}}\frac{B}{\sqrt{d}}\right\| &\leqslant \left\|\frac{d}{2n}\mathbf{1}\mathbf{1}^\intercal - \frac{\sqrt{d}D^{\frac{1}{2}}}{2n}\mathbf{1}\mathbf{1}^\intercal\right\| + \left\|\frac{a-b}{2n}\boldsymbol{\chi}\boldsymbol{\chi}^\intercal - \frac{(a-b)D^{\frac{1}{2}}}{2n\sqrt{d}}\boldsymbol{\chi}\boldsymbol{\chi}^\intercal\right\| \\ &= \frac{\sqrt{d}}{2n}\|(\sqrt{d}\mathbf{1} - D^{\frac{1}{2}}\mathbf{1})\mathbf{1}^\intercal\| + \frac{a-b}{2n\sqrt{d}}\|(\sqrt{d}\boldsymbol{\chi} - D^{\frac{1}{2}}\boldsymbol{\chi})\boldsymbol{\chi}^\intercal\| \\ &\leqslant \frac{\sqrt{d}}{2n}\|\sqrt{d}\mathbf{1} - D^{\frac{1}{2}}\mathbf{1}\|\cdot\|\mathbf{1}\| + \frac{\sqrt{d}}{2n}\|\sqrt{d}\boldsymbol{\chi} - D^{\frac{1}{2}}\boldsymbol{\chi}\|\cdot\|\boldsymbol{\chi}^\intercal\| \\ &\leqslant \mathcal{O}(\sqrt{d})\,, \qquad (4.40)\end{aligned}$$

where the last step holds w.h.p. and uses the fact that, under our assumptions, Lemma 4.14 implies that w.h.p.

$$\sum_i (\sqrt{d} - \sqrt{d_i})^2 \leqslant \mathcal{O}(n)\,.$$

Then, for any vector $\mathbf{x} \in \{\pm 1\}^n$,

$$\|\sqrt{d}\mathbf{x} - D^{\frac{1}{2}}\mathbf{x}\|^2 = \sum_i (x_i\sqrt{d} - x_i\sqrt{d_i})^2 = \sum_i x_i^2(\sqrt{d} - \sqrt{d_i})^2 \leqslant \mathcal{O}(n)\,,$$

so that

$$\|\sqrt{d}\mathbf{1} - D^{\frac{1}{2}}\mathbf{1}\| = |\sqrt{d}\boldsymbol{\chi} - D^{\frac{1}{2}}\boldsymbol{\chi}\| = \mathcal{O}(\sqrt{n})\,.$$

From (4.39) and (4.40) it follows that

$$\left\|D^{-\frac{1}{2}}BD^{-\frac{1}{2}} - B/d\right\| \leqslant \mathcal{O}\left(\frac{1}{\sqrt{d}}\right)\,,$$

which concludes the proof. ∎

Remark 2 (Comparison with the work of Le and Vershynin). *In [47] Le and Vershynin's show the following result. Let G be sampled from $\mathcal{G}_{2n,\frac{a}{n},\frac{b}{n}}$ with average degree $d = a + b$ and adjacency matrix A, and consider the modified matrix $A_\tau = A + \frac{\tau}{2n}J$, where τ is a parameter of the order of d (for example what we state below holds for $\tau = 3d$), and let*

$$N_\tau = (D + \tau I)^{-\frac{1}{2}}A_\tau(D + \tau I)^{-\frac{1}{2}}$$

be the normalized version of the modified adjacency matrix, and let B_τ be the expectation of the modified adjacency matrix

$$B_\tau = B + \frac{\tau}{2n}J\,.$$

Then, with constant probability,

$$\left\|N_\tau - \frac{1}{d+\tau}B_\tau\right\| \leqslant \mathcal{O}\left(\frac{1}{\sqrt{d}}\right)\,,$$

even when a, b, and d are constants. Our Lemma 4.13 above shows that when d is at least order of $\log n$, then the above bound holds without correction, for $\tau = 0$. It would also be possible to modify the proof of Le and Vershynin to work for $\tau = 0$ in the regime of logarithmic or higher degree. Such a proof would be similar to our argument above, but instead of a straightforward reduction to bounding $\sum_i (\sqrt{d} - \sqrt{d_i})^2$, which requires some work, one would derive a somewhat more complicated reduction to the easier task of bounding $\sum_i (d - d_i)^2$.

4.3.2. Analysis of the AVERAGING Dynamics. We first use the Davis-Kahan theorem (Theorem A.4) to infer spectral properties of N based on the fact that N and B/d are close with high probability, as established above.

LEMMA 4.15 (Spectral properties of N). *There is an absolute constant $\bar{c}$ such that the following holds. Let G be a graph sampled from $\mathcal{G}_{2n,\frac{a}{n},\frac{b}{n}}$, with $\bar{c}\log n \leqslant d \leqslant n^{\frac{1}{4}}$ and $a - b > 200^2(4c_2 + 4)\sqrt{d}$, where c_2 is the constant of Lemma 4.13. Consider the normalized adjacency matrix N of G.[14] Then the following properties hold w.h.p.:*

1. $\lambda_2 \geqslant 1 - \frac{2b}{d} - \frac{c_1}{\sqrt{d}}$;
2. $\lambda_2 \geqslant 2\lambda$;
3. *A subset of nodes S exists with $|S| = 2n\left(1 - 200^2 \cdot (4c_2 + 4) \cdot \frac{\sqrt{d}}{a-b}\right)$ such that, for every $u \in S$,*

$$|\sqrt{2nd}(D^{-1/2}w_2)(u) - \chi(u)| \leqslant \frac{1}{100}\,.$$

Proof. Recall that B/d is such that its largest eigenvalue is 1, its second largest is $(a-b)/d = 1 - 2b/d$, with eigenvector $\boldsymbol{\chi}$, and all other eigenvalues are zero. From Lemma 4.13, w.h.p.

$$\|N - B/d\| \leqslant c_2/\sqrt{d}\,. \tag{4.41}$$

From Theorem A.2, we have $\lambda_2 \geqslant 1 - 2b/d - c_2/\sqrt{d}$ and $\lambda \leqslant c_2/\sqrt{d}$, which in turn implies $\lambda_2 \geqslant 2\lambda$, because $a - b \geqslant 3c_2\sqrt{d}$.

As for the third claim, let us write $\mathbf{w}_2 = \mathbf{w}_\parallel + \mathbf{w}_{\boldsymbol{\chi}} + \mathbf{w}_\perp$ where $\mathbf{w}_\parallel$ and $\mathbf{w}_{\boldsymbol{\chi}}$ are the projections of $\mathbf{w}_2$ onto $\mathbf{1}$ and $\boldsymbol{\chi}$ respectively, and $\mathbf{w}_\perp$ is the projection of $\mathbf{w}_2$ onto the subspace orthogonal to $\mathbf{1}$ and $\boldsymbol{\chi}$. We are going to argue that $\mathbf{w}_2$ is close to $\mathbf{w}_{\boldsymbol{\chi}}$ and hence to $\boldsymbol{\chi}$.

First let us see that $\mathbf{w}_\parallel$ is small. We know that $\langle \mathbf{w}_2, D^{1/2}\mathbf{1}\rangle = 0$, thus

$$\|\mathbf{w}_\parallel\| = \frac{1}{\sqrt{2n}}|\langle \mathbf{w}_2, \mathbf{1}\frac{\sqrt{d}}{\sqrt{d}} - \frac{1}{\sqrt{d}}D^{\frac{1}{2}}\mathbf{1}\rangle| \leqslant \frac{1}{\sqrt{2nd}}\|\mathbf{w}_2\|\|\mathbf{1}\sqrt{d} - D^{\frac{1}{2}}\mathbf{1}\| \leqslant \sqrt{\frac{5}{2d}} \leqslant \frac{2}{\sqrt{d}}\,,$$

where in the second-to-last inequality we used Lemma 4.14.

From (4.41) and the definition of spectral norm,

$$\mathbf{w}_2^\intercal N\mathbf{w}_2 - \mathbf{w}_2^\intercal \frac{B}{d}\mathbf{w}_2 \leqslant \frac{c_2}{\sqrt{d}}\,.$$

Now let us compute the two quadratic forms in the above expression. We can write

$$\mathbf{w}_2^\intercal N\mathbf{w}_2 = \lambda_2 \geqslant \frac{a-b}{d} - \frac{c_2}{\sqrt{d}}$$

[14] Please refer to the beginning of Section 4 for definition and notation.

$$\mathbf{w}_2^\intercal \frac{B}{d} \mathbf{w}_2 = \|\mathbf{w}_\parallel\|^2 + \frac{a-b}{d}\|\mathbf{w}_{\boldsymbol{\chi}}\|^2$$

and, putting it all together,

$$\frac{a-b}{d}\|\mathbf{w}_{\boldsymbol{\chi}}\|^2 \geqslant \frac{a-b}{d} - 2\frac{c_2}{\sqrt{d}} - \frac{2}{\sqrt{d}} = \frac{a-b-(2c_2+2)\sqrt{d}}{d}\,,$$

which gives

$$\|\mathbf{w}_{\boldsymbol{\chi}}\|^2 \geqslant \frac{a-b-(2c_2+2)\sqrt{d}}{a-b} = 1 - \frac{(2c_2+2)\sqrt{d}}{a-b}\,.$$

Moreover,

$$\left\langle \mathbf{w}_2, \frac{\boldsymbol{\chi}}{\sqrt{2n}} \right\rangle = \|\mathbf{w}_{\boldsymbol{\chi}}\| \geqslant \|\mathbf{w}_{\boldsymbol{\chi}}\|^2 \geqslant 1 - \frac{(2c_2+2)\sqrt{d}}{a-b}\,, \tag{4.42}$$

where we used the fact that $\|\mathbf{w}_\chi\| \leqslant 1$.

We are now able to bound the distance between $\mathbf{w}_2$ and $\boldsymbol{\chi}/\sqrt{2n}$:

$$\begin{aligned}\left\|\mathbf{w}_2 - \frac{1}{\sqrt{2n}}\boldsymbol{\chi}\right\|^2 &= \|\mathbf{w}_2\|^2 + \left\|\frac{1}{\sqrt{2n}}\boldsymbol{\chi}\right\|^2 - 2\langle \mathbf{w}_2, \frac{1}{\sqrt{2n}}\boldsymbol{\chi}\rangle \\ &= 2 - 2\|\mathbf{w}_{\boldsymbol{\chi}}\| \leqslant \frac{(4c_2+4)\sqrt{d}}{a-b}\,, \end{aligned} \tag{4.43}$$

where in the last inequality we used (4.42).

We now define the set O of "outlier" nodes u such that

$$\left|\sqrt{2n}\, w_2(u) - \chi(u)\right| > \frac{1}{200}\,. \tag{4.44}$$

Since each node $u \in O$ contributes at least $1/(200^2 \cdot 2n)$ to $\|\mathbf{w}_2 - \frac{1}{\sqrt{2n}}\boldsymbol{\chi}\|^2$ and, by (4.43),

$$\|\mathbf{w}_2 - \frac{1}{\sqrt{2n}}\boldsymbol{\chi}\|^2 \leqslant \frac{(4c_2+4)\sqrt{d}}{a-b},$$

it follows that

$$|O| \leqslant 2n \cdot 200^2 \cdot \frac{(4c_2+4)\sqrt{d}}{a-b}.$$

Notice that, from the Chernoff bound, w.h.p. $\sqrt{d/d_u} = 1 \pm 1/200$, for each $u \in V$. Thus, (4.44) and the last fact imply that, for each $u \in S = V \setminus O$, w.h.p.

$$\left|\sqrt{2nd}D^{-\frac{1}{2}} w_2(u) - \chi(u)\right| \leqslant \frac{1}{100},$$

and the third claim of Lemma 4.15 is proved. ∎

We are now ready to show that the AVERAGING dynamics achieves weak reconstruction when $a - b \gg \sqrt{d}$.

Proof of Theorem 4.12. The proof proceeds along the same lines as the one of Theorem 4.7. To begin, recall (4.3) from Lemma 4.1:

$$x^{(t-1)}(u) - x^{(t)}(u) = \tilde{\alpha}_2 \lambda_2^{t-1}(1-\lambda_2)v_2(u) + e^{(t-1)}(u) - e^{(t)}(u).$$

Moreover, Corollary 4.2 (in particular, (4.5)) ensures that, if $\tilde{\alpha}_2 \neq 0$, for every $u \in V$ such that $v_2(u) \neq 0$,

(4.45)

$$\mathtt{sgn}(x^{(t-1)}(u) - x^{(t)}(u)) = \mathtt{sgn}(\tilde{\alpha}_2 v_2(u))\,, \quad \text{whenever } \; t-1 \geqslant \frac{\log\left(\frac{2\sqrt{2nd_{\max}/d_{\min}}}{|\tilde{\alpha}_2|\cdot|v_2(u)|(1-\lambda_2)}\right)}{\log(\lambda_2/\lambda)}\,.$$

We next prove the following three claims:

1. Set $\hat{c} = 200^2 \cdot (4c_2+4) \cdot \frac{\sqrt{d}}{a-b}$, with c_2 the absolute constant appearing in Claim 3 of Lemma 4.15, and notice that $\hat{c} < 1$, whenever $a - b > 200^2 \cdot (4c_2+4) \cdot \sqrt{d}$. From this claim, a subset S of nodes exists, such that $|S| \geqslant 2(1-\hat{c})n$ and, for each node $u \in S$,
$$|v_2(u)| \geqslant \frac{99}{100\sqrt{2nd}} \quad \text{and} \quad \mathtt{sgn}\left(\tilde{\alpha}_2 v_2(u)\right) = \mathtt{sgn}\left(\alpha_2\right)\mathtt{sgn}\left(\boldsymbol{\chi}\right).$$
2. $|\tilde{\alpha}_2| \geqslant \frac{1}{\sqrt{2nd}}$, w.h.p. Thus, for the subset S in the previous claim the AVERAGING protocol produces the right reconstruction.
3. $1 - \lambda_2 \geqslant \frac{1}{n^4}$, w.h.p. over the randomness of G.

Together with (4.45) and the fact that $d_{\max}/d_{\min} = \mathcal{O}(1)$, w.h.p., these claims imply that weak reconstruction is achieved within $\mathcal{O}(\log n)$ rounds, w.h.p.

Proof of Claim 1. Since $\mathbf{v}_2 = D^{-1/2}\mathbf{w}_2$ and $(a-b)^2 > 10^6 c_2^2 d$, the third claim of Lemma 4.15 implies that, w.h.p. over the randomness of the graph, for a subset S of at least $2(1-\hat{c})n$ entries i of $\mathbf{v}_2$,

$$|\sqrt{2nd}\, v_2(i) - \chi(i)| \leqslant \frac{1}{100}\,,$$

for a suitable constant $\hat{c} < 1$. This is equivalent to:

$$v_2(i) \geqslant \frac{99}{100\sqrt{2nd}} \text{ if } i \in V_1 \cap S \;\text{ and }\; v_2(i) \leqslant -\frac{99}{100\sqrt{2nd}} \text{ if } i \in V_2 \cap S\,.$$

This proves Claim 1.

Proof of Claim 2. To begin, recall that $\tilde{\alpha}_2 = \mathbf{w}_2^{\intercal} D^{1/2}\mathbf{x}$ from Claim 1 of Lemma 4.1. Let $\mathbf{y} = \sqrt{2nd}D^{1/2}\mathbf{w}_2 = \sqrt{2nd}D\mathbf{v}_2$ for the remainder of this proof, so that

$$|\tilde{\alpha}_2| = \frac{1}{\sqrt{2nd}}|\mathbf{y}^{\intercal}\mathbf{x}|\,.$$

Then, recall that, for every $i \in S$,

$$|\sqrt{2nd}\, v_2(i) - \chi(i)| \leqslant \frac{1}{100}.$$

When $i \in V_1 \cap S$ this implies

$$\frac{99}{100} \leqslant \sqrt{2nd}v_2(i) \leqslant \frac{101}{100}.$$

Likewise,

$$-\frac{101}{100} \leqslant \sqrt{2nd} v_2(i) \leqslant -\frac{99}{100}, i \in V_2 \cap S\,,$$

whenever $i \in V_1 \cap S$. Next, the assumptions of Theorem 4.12 (in particular, $d > 9\log n$) imply that, w.h.p., $(1-\alpha)d \leqslant d_u \leqslant (1+\alpha)d$, for every $u \in V$, with $\alpha < 1$ a (small) constant. As a consequence, w.h.p.

$$\begin{cases} \frac{99}{100}(1-\alpha)d \leqslant y(i) \leqslant \frac{101}{100}(1+\alpha)d, i \in V_1 \cap S, \\ -\frac{101}{100}(1+\alpha)d \leqslant y(i) \leqslant -\frac{99}{100}(1-\alpha)d, i \in V_2 \cap S\,. \end{cases}$$

We can now apply Lemma B.2 with the following setting: $\mathbf{y}$ and S defined above, $r = \frac{99}{100}(1-\alpha)d$, $c = \frac{101}{99}\frac{1+\alpha}{1-\alpha}$, $k = 1/(1-\hat{c})$, $km = 2n$ and $\delta = \frac{1}{\sqrt{2n}}$. Then

$$\mathbf{P}\left(|(1/\sqrt{2n})\mathbf{y}^\intercal \mathbf{x}| \leqslant \frac{1}{\sqrt{2n}}\right) = \mathbf{P}\left(|(1/\sqrt{km})\mathbf{y}^\intercal \mathbf{x}| \leqslant \delta\right) \leqslant \sqrt{\frac{2k}{\pi}} \cdot \frac{\delta}{r} + \frac{4c}{\sqrt{m}} = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right).$$

Proof of Claim 3. We begin by noting that, from our hypotheses, G is connected w.h.p. In particular, each community induces a random graph with parameter $p = a/n \geqslant d/2n \geqslant 1.5\log n/n$, which is well above the connectivity threshold. Moreover, our assumptions on q imply that the expected number of edges connecting the two communities is at least $\log n$, which in turn implies that at least one such edge will be present, w.h.p. Together, these arguments imply that G is connected, w.h.p. Conditioned on this event, the third claim follows directly from Cheeger's inequalities. In more detail, (4.7) implies $1-\lambda_2 \geqslant \frac{\Phi^2}{2}$. Connectedness in turn implies that, for every subset S of the vertices, the corresponding cut is crossed by at least one edge, so that its conductance is at least $1/vol(S)$. But $vol(S) \leqslant |S|\cdot(|S|-1) < n^2$ deterministically. This immediately implies $1-\lambda_2 > \frac{1}{n^4}$ and thus Claim 3, which completes the proof. □

REMARK 3. *After looking at Lemma 4.15, one may wonder whether it could be enough to generalize Definition 4.3 to include "quasi-$(2n,d,b,\gamma)$-clustered graphs", that is, graphs that are $(2n,d,b,\gamma)$-clustered except for a small number of nodes which may have a much higher degree. In fact, this would be rather surprising: These higher-degree nodes may connect to the other nodes in a way that would greatly perturb the eigenvalues and eigenvectors of the graph. In $\mathcal{G}_{2n,p,q}$, besides the fact that there are few nodes with degree much larger than d, it is also crucial that they are connected in a* non-adversarial *way, that is, randomly.*

5. Moving beyond two communities: An outlook. The AVERAGING protocol can be naturally extended to address the case of more communities. One way to achieve this is by performing a suitable number of independent, parallel runs of the protocol. We next outline the analysis for a natural generalization of the regular block model. We formally state the result in Theorem 5.2, and we give a proof in the following subsection.

Let $G = (V, E)$ be a d-regular graph in which V is partitioned into k equal-sized communities $V_1, \ldots, V_k$, while every node in V_i has exactly a neighbors within V_i and exactly b neighbors in each V_j, for $j \neq i$. Note that $d = a+(k-1)\cdot b$. We observe that the transition matrix P of the random walk on G has an eigenvalue $1 - kb/d$ with multiplicity at least $k-1$, since all stepwise vectors that are constant within each community V_i and whose entries sum to zero are eigenvectors of P with eigenvalue $1 - kb/d$. Moreover, if

$$\max\{\lambda_{k+1}, |\lambda_{kn}|\} \;<\; (1-\varepsilon)\cdot\left(1-\frac{kb}{d}\right),$$

then P has eigenvalues

$$\lambda_1 = 1, \lambda_2 = \cdots = \lambda_k = \left(1-\frac{kb}{d}\right),$$

with all other eigenvalues strictly smaller by a $(1-\varepsilon)$ factor.

The properties discussed above can be leveraged to achieve strong reconstruction in the regular case we consider in this section. In more detail, let $\mathbf{x} \in \mathbb{R}^V$ be a vector. We say that a node v is of *positive* (respectively, *negative*) type with respect to $\mathbf{x}$ if a threshold $T = T(\mathbf{x}, v)$ exists such that the value $(P^t\mathbf{x})(v)$ *increases* (respectively, *decreases*) with t, for all $t \geqslant T$. Note that, if $\mathbf{x}$ were orthogonal to the span of the first k eigenvectors, we might in principle have nodes of neither type, because $(P^t\mathbf{x})(v)$ might not eventually become strictly monotone in t.[15] However, if $\mathbf{x}$ is chosen uniformly at random, this is unlikely to happen. On the contrary (see Lemma 5.4), w.h.p. $\mathbf{x}$ has a sufficiently large component in the span of the first k eigenvectors, so that for $t \geqslant \max\{T(\mathbf{x}, v) : v \in V\}$, the vector $P^t\mathbf{x}$ is nearly contained in the span of the first k eigenvectors and thus, from the previous paragraphs, up to a "small" additive error, it can be expressed as a linear combination of vectors that are stepwise constant with respect to the communities. This implies that, for an initial vector $\mathbf{x} \in \{-1, 1\}^{|V|}$ chosen uniformly at random, all nodes in the same community are of the same type, w.h.p. (see Lemma 5.4), while nodes from different communities are of different types, with probability γ (see Lemma 5.5), where $\gamma > 0$ is an absolute constant.

We thus consider the parallel procedure where each node initially chooses $\ell = \Theta(\log n)$ values independently and uniformly at random from $\{-1, 1\}$ and, to each of them, applies an independent copy of the AVERAGING protocol in parallel.

In the proof of Theorem 5.2, we will prove that the arguments outlined above imply that, starting from the ℓ random initial vectors $\mathbf{x}^1, \ldots, \mathbf{x}^\ell$, each in $\{-1, 1\}^V$, the parallel procedure has w.h.p. the following properties: i) every node is either of positive or negative type for each $\mathbf{x}^i$; ii) if we associate a "signature" to each node, namely, the sequence of ℓ types, then nodes within the same V_i exhibit the same signature, while nodes in different V_i, V_j have different signatures.

5.1. AVERAGING **dynamics for k-clustered regular graphs.** For $n, d, k \in \mathbb{N}$, we say that a d-regular graph G with kn nodes is (kn, d, b)-*regular* if a k-partition of the nodes $(V_1, V_2, \ldots, V_k)$ exists such that, for every $i = 1, \ldots, k$ and every $j = 1, \ldots, k$, every node in V_i has exactly b neighbors in V_j, if $j \neq i$, and thus exactly $d - (k-1)b$ neighbors in V_i. We name $a = d - (k-1)b$ the "inner" degree of each node.

For a (kn, d, b)-*regular* graph $G = (V, E)$ and k-partition $(V_1, \ldots V_k)$ we say that a vector $\mathbf{z} = (z(u) : u \in V)$ is *stepwise* if for every $i = 1, \ldots, k$ and for every $u, v \in V_i$, $z(u) = z(v)$. Notice that any stepwise vector $\mathbf{z}$ such that $\sum_{u \in V} z(u) = 0$ is an eigenvector of the transition matrix P with eigenvalue $(a-b)/d = 1 - kb/d$.

DEFINITION 5.1 (Clustered Regular Graph). *A (kn, d, b)-clustered regular graph $G = ((V_1, \ldots, V_k), E)$ is a (kn, d, b)-regular graph such that $1 - kb/d$ is the second*

[15] This, for example, might follow from the presence of negative eigenvalues associated with eigenvectors from the $(k+1)$-th onward.

largest eigenvalue of P with multiplicity $k-1$, that is, $\lambda_2 = \cdots = \lambda_k = 1 - kb/d$ and $\lambda = \max\{\lambda_{k+1}, |\lambda_{kn}|\} < \lambda_k$.

THEOREM 5.2 (More communities). *Let $G = (V, E)$ be a (kn, d, b)-clustered regular graph with $k = \mathcal{O}(n^{(1/2)-\delta})$, for an arbitrarily small constant $\delta > 0$, and assume that $\lambda = \max\{\lambda_{k+1}, |\lambda_{kn}|\} < (1-\varepsilon)\left(1 - \frac{kb}{d}\right)$, for a suitable constant $\varepsilon > 0$. Then, for $\ell = \Theta(\log n)$, the* AVERAGING *protocol with ℓ parallel runs produces a strong reconstruction within $\mathcal{O}(\log n)$ rounds, w.h.p.*

5.2. Proof of Theorem 5.2. We first give a formal definition of the *type* of a node with respect to an initial vector $\mathbf{x}$.

DEFINITION 5.3. *Let $\mathbf{x} \in \mathbb{R}^V$ be a vector. We say that a node v is of* positive *(respectively,* negative*) type with respect to $\mathbf{x}$ if a threshold $T = T(\mathbf{x}, v) \in \mathbb{N}$ exists such that, for all $t \geqslant T$, the value $(P^t\mathbf{x})(v)$* increases *(respectively,* decreases*) with t.*

From the spectral decomposition of P, it is easy to see that all nodes are either of positive or of negative type with respect to any vector $\mathbf{x}$ that is not orthogonal to the span of the first k eigenvectors of P.

In Lemmas 5.4 and 5.5 we show that, for an initial random vector $\mathbf{x} \in \{-1, 1\}^{kn}$, all nodes in the same community have the same type, w.h.p., while nodes in different communities have different types, with constant probability.

LEMMA 5.4. *For any $i \in \{1, \ldots, k\}$, if the vector $\mathbf{x} \in \{-1, 1\}^{kn}$ is chosen uniformly at random, then the nodes of V_i are either all of positive type or all of negative type, w.h.p., with threshold $T(\mathbf{x}, v) \leqslant 2\log(n)/\log(\lambda_2/\lambda)$ for all $v \in V_i$. Furthermore, the two events have equal probability.*

Proof. We can decompose the vector $\mathbf{x}$ as

$$\mathbf{x} = \mathbf{x_1} + \mathbf{x}_{V_i} + \mathbf{x}_{\perp_i} + \mathbf{x}_{\perp}\,,$$

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

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

The reason for picking the above decomposition is that $\mathbf{x}_{\perp_i}$ is zero in V_i. Indeed, since $\mathbf{x}_{\perp_i}$ is orthogonal to $\mathbf{1}$ and to $\mathbf{1}_{V_i} - k^{-1}\mathbf{1}$, from

$$\langle \mathbf{x}_{\perp_i}, \mathbf{1}\rangle = \langle \mathbf{x}_{\perp_i}, \mathbf{1}_{V_i} - k^{-1}\mathbf{1}\rangle = 0\,,$$

it follows that $\langle \mathbf{x}_{\perp_i}, \mathbf{1}_{V_i}\rangle = 0$. Thus, the entries of $\mathbf{x}_{\perp_i}$ sum to zero within V_i, but, being in the eigenspace of λ_2, the entries of $\mathbf{x}_{\perp_i}$ are constant within components, and so they must be all zero within V_i.

According to the above decomposition, we have that

$$P^t\mathbf{x} = \mathbf{x_1} + \lambda_2^t\mathbf{x}_{V_i} + \lambda_2^t\mathbf{x}_{\perp_i} + P^t\mathbf{x}_{\perp}\,.$$

Hence, for each $v \in V_i$,

$$(P^{t+1}\mathbf{x})(v) - (P^t\mathbf{x})(v) = \lambda_2^t \cdot (1-\lambda_2)(\mathbf{x}_{V_i})(v) + ((P^{t+1} - P^t)\mathbf{x}_{\perp})(v)\,. \tag{5.1}$$

For large enough t, that is, when $\lambda_2^t > n^2\lambda^t$ (where $\lambda = \max\{\lambda_{k+1}, |\lambda_{kn}|\}$, as defined in the statement of Theorem 5.2), the hypothesis $\lambda < (1-\varepsilon)\lambda_2$ implies that

$$|(P^t\mathbf{x}_{\perp})(v)| \leqslant \|P^t\mathbf{x}_{\perp}\|_\infty \leqslant \|P^t\mathbf{x}_{\perp}\| \leqslant \lambda^t\|\mathbf{x}_{\perp}\| \leqslant \sqrt{n}\cdot\lambda^t \leqslant \frac{1}{n^{1.5}}\lambda_2^t\,. \tag{5.2}$$

Now observe that

$$\|\mathbf{1}_{V_i} - k^{-1}\mathbf{1}\|^2 = \sum_{u\in V_i}\left(1-\frac{1}{k}\right)^2 + \sum_{u\in V\setminus V_i}\frac{1}{k^2} = \frac{k-1}{k}n = (1-k^{-1})n, \tag{5.3}$$

whence:

$$\|\mathbf{x}_{V_i}\| = \frac{|\langle \mathbf{x}, \mathbf{1}_{V_i} - k^{-1}\mathbf{1}\rangle|}{\|\mathbf{1}_{V_i} - k^{-1}\mathbf{1}\|} = \frac{1}{\sqrt{(1-k^{-1})n}}\left|\sum_{j\in V_i} x(j) - \sum_{j\in V}\frac{x(j)}{k}\right|. \tag{5.4}$$

Thus, for each $v \in V_i$,

$$\begin{aligned}
|(\mathbf{x}_{V_i})(v)| &= \frac{|\langle \mathbf{x}, \mathbf{1}_{V_i} - k^{-1}\mathbf{1}\rangle|}{\|\mathbf{1}_{V_i} - k^{-1}\mathbf{1}\|}\cdot\frac{1-k^{-1}}{\|\mathbf{1}_{V_i} - k^{-1}\mathbf{1}\|} \\
&= \frac{1}{(1-k^{-1})n}\left|\sum_{j\in V_i} x(j) - \sum_{j\in V}\frac{x(j)}{k}\right|(1-k^{-1}) \\
&= \frac{1}{n}\left|\sum_{j\in V_i} x(j) - \sum_{j\in V}\frac{x(j)}{k}\right| = \sqrt{\frac{1-k^{-1}}{n}}\,\|\mathbf{x}_{V_i}\|,
\end{aligned}$$

where in the second equality we used (5.3), and in the second-to-last and last steps we used (5.4). Moreover, observe that

$$\begin{aligned}
\|\mathbf{x}_{V_i}\| &= \frac{|\langle \mathbf{x}, \mathbf{1}_{V_i} - k^{-1}\mathbf{1}\rangle|}{\|\mathbf{1}_{V_i} - k^{-1}\mathbf{1}\|} = \frac{1}{\sqrt{(1-k^{-1})n}}|\langle \mathbf{x}, \mathbf{1}_{V_i} - k^{-1}\mathbf{1}\rangle| \\
&= \frac{1}{\sqrt{kn}}\left|\left\langle \mathbf{x}, \frac{\mathbf{1}_{V_i} - k^{-1}\mathbf{1}}{\sqrt{k^{-1}(1-k^{-1})}}\right\rangle\right|,
\end{aligned}$$

where in the second equality we used (5.3). If we let $\mathbf{y} = \frac{\mathbf{1}_{V_i} - k^{-1}\mathbf{1}}{\sqrt{k^{-1}(1-k^{-1})}}$, each entry of this vector is either

$$\frac{1-k^{-1}}{\sqrt{k^{-1}(1-k^{-1})}} = \sqrt{k-1} \quad \text{(this is the case for all nodes in } V_i\text{)}$$

or

$$-\frac{k^{-1}}{\sqrt{k^{-1}(1-k^{-1})}} = -\frac{1}{\sqrt{k-1}} \quad \text{(this is the case for all nodes not in } V_i\text{)}$$

In particular, for every v, $\frac{1}{\sqrt{k-1}} \leqslant |y(v)| \leqslant \sqrt{k-1}$. Hence, by applying Lemma B.2 with $m = n$, $r = \frac{1}{\sqrt{k-1}}$, $c = k-1$, and $\delta = \sqrt{\frac{1}{n}}$, we get

$$\begin{aligned}
\mathbf{P}\left(\|\mathbf{x}_{V_i}\| \leqslant \sqrt{\frac{1}{n}}\right) &= \mathbf{P}\left(\frac{1}{\sqrt{kn}}\langle \mathbf{x}, \mathbf{y}\rangle \leqslant \sqrt{\frac{1}{n}}\right) \\
&\leqslant \sqrt{\frac{2k}{\pi}}\cdot\sqrt{\frac{k-1}{n}} + \frac{4(k-1)}{\sqrt{n}} = \mathcal{O}\left(\frac{k}{\sqrt{n}}\right).
\end{aligned}$$

Hence, w.h.p.

$$|(\mathbf{x}_{V_i})(v)| \geqslant \sqrt{\frac{1-k^{-1}}{n}}\sqrt{\frac{1}{n}} = \sqrt{1-k^{-1}} \cdot \frac{1}{n}\,. \tag{5.5}$$

Using (5.2) and (5.5) in (5.1), we get that $(P^{t+1}\mathbf{x})(v) - (P^t\mathbf{x})(v)$ has the same sign as $(\mathbf{x}_{V_i})(v)$, w.h.p., for all elements of V_i simultaneously, and it is equally likely to be positive or negative. □

Observe that the above lemma also holds for all i simultaneously, w.h.p., by a union bound.

LEMMA 5.5. *Let u, v be two nodes in two distinct communities. An absolute constant $\gamma > 0$ exists such that, if $\mathbf{x} \in \{-1,1\}^{kn}$ is chosen uniformly at random, then the types of u and v are different with probability at least γ and the threshold $T = 2\log(n)/\log(\lambda_2/\lambda)$.*

Proof. Without loss of generality, assume $u \in V_1$ and $v \in V_2$. This time we decompose $\mathbf{x}$ as follows

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

where

- $\mathbf{x}_{\mathbf{1}}$ is the component parallel to $\mathbf{1}$,
- $\mathbf{x}_{V_{1\oplus 2}}$ is the component parallel to $\mathbf{1}_{V_1} + \mathbf{1}_{V_2} - \frac{2}{k}\mathbf{1}$,
- $\mathbf{x}_{V_{1\ominus 2}}$ is the component parallel to $\mathbf{1}_{V_1} - \mathbf{1}_{V_2}$,
- $\mathbf{x}_{\perp_{1,2}}$ is the component in the eigenspace of λ_2 orthogonal to $\mathbf{x}_{V_{1\oplus 2}}$ and $\mathbf{x}_{V_{1\ominus 2}}$,
- $\mathbf{x}_{\perp}$ is the component of $\mathbf{x}$ orthogonal to all the above vectors (i.e., orthogonal to both the eigenspaces of λ_1 and λ_2).

Similarly to the proof of Lemma 5.4, the important observations are that $\mathbf{x}_{V_{1\oplus 2}}$ and $\mathbf{x}_{V_{1\ominus 2}}$ are in the eigenspace of λ_2, and that $\mathbf{x}_{\perp_{1,2}}$ is zero in all the coordinates of V_1 and of V_2. Thus, for each $v \in V_1 \cup V_2$ we have that

$$(P^{t+1}\mathbf{x})(v) - (P^t\mathbf{x})(v) = \lambda_2^t(1-\lambda_2)(\mathbf{x}_{V_{1\oplus 2}} + \mathbf{x}_{V_{1\ominus 2}})(v) + ((P^{t+1} - P^t)\mathbf{x}_{\perp})(v)\,. \tag{5.6}$$

From (5.6) it is easy to see that, if the initial vector $\mathbf{x}$ is such that the two following conditions hold for every $v \in V_1 \cup V_2$,

$$|(\mathbf{x}_{V_{1\oplus 2}})(v)| \leqslant \frac{1}{2}|(\mathbf{x}_{V_{1\ominus 2}})(v)| \quad \text{and} \tag{5.7}$$

$$|((P^{t+1} - P^t)\mathbf{x}_{\perp})(v)| \leqslant \frac{1}{8}\lambda_2^t \cdot (1-\lambda_2) \cdot |(\mathbf{x}_{V_{1\ominus 2}})(v)|\,. \tag{5.8}$$

then for such an initial vector $\mathbf{x}$, all the elements in V_1 have the same type, all the elements of V_2 have the same type, and the two types are different.

Now observe that

$$(\mathbf{x}_{V_{1\oplus 2}})(v) = \frac{1}{2n}\left(\sum_{i\in V_1} x(i) + \sum_{i\in V_2} x(i) - \frac{2}{k}\sum_{i\in V} x(i)\right)$$

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

Hence, if the initial vector $\mathbf{x}$ satisfies the following three conditions

$$2\sqrt{n} \leqslant \sum_{v \in V_1} x(v) \leqslant 3\sqrt{n}, \tag{5.9}$$

$$-2\sqrt{n} \leqslant \sum_{v \in V_2} x(v) \leqslant -\frac{4}{3}\sqrt{n}\,, \text{ and} \tag{5.10}$$

$$0 \leqslant \sum_{v \in V/(V_1 \cup V_2)} x(v) \leqslant \frac{1}{10}\sqrt{kn}\,, \tag{5.11}$$

then $2n(\mathbf{x}_{V_{1 \oplus 2}})(v) \leqslant \frac{5}{3}\sqrt{n}$ and $2n(\mathbf{x}_{V_{1 \ominus 2}})(v) \geqslant \frac{10}{3}\sqrt{n}$, and we get (5.7).

Finally, observe that, for an initial vector $\mathbf{x} \in \{-1, 1\}^{kn}$ chosen uniformly at random, events (5.9), (5.10) and (5.11) are independent, and each one happens with constant probability. Thus, for such a random initial vector $\mathbf{x}$, (5.7) holds with constant probability. Moreover, notice that if (5.7) holds then (5.8) is satisfied, w.h.p., for large enough t: for example, as soon as $\lambda_2^t > n^2 \lambda^t$ (where $\lambda = \max\{\lambda_{k+1}, |\lambda_{kn}|\}$, as defined in the statement of Theorem 5.2). □

From Lemmas 5.4 and 5.5 it follows that it is enough to pick $\ell = (3/\gamma) \log n$ parallel runs to have that the signatures are well defined and they are the same within each community and different between communities, w.h.p. Indeed, Lemma 5.5 ensures that, for each pair of nodes u, v belonging to distinct communities, the probability that they are of the same type in all ℓ runs is smaller than $(1-\gamma)^\ell \leqslant e^{\gamma \ell} \leqslant n^{-3}$. A union bound over all pairs of nodes in distinct communities then proves that all pairs of nodes belonging to distinct communities have distinct type in at least one of the ℓ runs, w.h.p. Since, from Lemma 5.4, in each run all nodes in the same community have the same type, w.h.p., it follows that all nodes in the same community have the same type in all the $\ell = \mathcal{O}(\log n)$ runs, w.h.p.

6. Follow-up and open questions. The contribution of this paper represents a first important step toward a rigorous understanding of the process yielded by an elementary local rule (that is, a dynamics) when it is applied over clustered networks. In particular, our analysis essentially shows that, in this setting, the averaging dynamics possesses a metastable regime where nodes' states well-reflect the hidden communities of the network: interestingly enough, we also show that this property can be efficiently exploited by a simple and local coloring criterion that allows fully decentralized community detection.

We believe that this contribution is important since it can provide reasonable models to study self-organizing properties observed in fully-decentralized Multi-Agent Systems having natural local interaction rules [25, 26, 44]. In this setting, we emphasize three research directions that have been inspired by the conference version of our paper.

The first one is to extend the analysis of the (synchronous) AVERAGING dynamics to more general graph classes and multiple, possibly non-balanced, communities. The general underlying question is whether the temporal evolution of the power method applied to an initial random vector may provide equivalent information as a spectral method, without requiring explicit eigenvector computations. A first step in this direction was taken by Becchetti et al [11], who extended the analysis presented in Section 5. In particular, they showed that a class of graphs exhibiting a milder form of regularity than the one considered here is the largest class of undirected, possibly weighted graphs that may contain k stepwise eigenvectors in the presence of a hidden

k-partition. Graphs belonging to this class need not be regular in general, nor does the hidden partition need to be balanced. In this case, under suitable conditions, running multiple, parallel instances of the AVERAGING dynamics affords recovery of the hidden partition, w.h.p. These results use a connection between volume regularity and lumpability of Markov chains [71]. At the same time, the analysis presented in [11] highlights that simple, sign-like rules applied to the power method are unlikely to distill information equivalent to an explicit eigenvector computation. Apparently, a more sophisticated approach is needed.

The second line of research concerns the analysis of suitable versions of the averaging dynamics that work in well-established asynchronous, "sparsified" communication models. Indeed, in the AVERAGING dynamics considered in this paper, every node communicates in parallel with all its neighbors at each round. While this might be too expensive in scenarios characterized by dense topologies, it is simply infeasible in other settings (for instance, the latter is the case when links represent sporadic opportunistic meetings that occur asynchronously). Motivated by the above considerations, a first line of follow-up work considered "sparsified", asynchronous variants of the AVERAGING protocol that work on the well-known *Population Protocol* model [9, 52, 69]. In the latter model, at every round, the end-points of (only) one link, chosen uniformly at random, can exchange data. According to this model, nodes are anonymous and cannot use any fixed static subgraph of the underlying graph. In more detail, Becchetti et al. in [9] consider two averaging protocols, working on the Population Protocol model, that approximately recover the community structure in the case of a class of regular clustered graphs having good inner expansion and a sparse cut. Their local clustering criterion is similar to the one we introduce in this paper on top of the AVERAGING dynamics, that is, it is based on the sign of the nodes' state. Their second-moment analysis show that the protocol converges in $\mathcal{O}(n \log n)$ time[16] and requires a work per node of order $\mathcal{O}(\text{polylog}n)$, even in the case of dense graphs. For the same restricted class of clustered regular graphs with dense cut (i.e., when the cut between the clusters has size $\Theta(|E|)$), they also derive a more complex second-moment analysis of the AVERAGING dynamics leading to a somewhat weighted version the AVERAGING protocol, equipped with a different clustering criterion that is based on the fluctuations of the nodes' states. This second protocol converges within $\mathcal{O}(n \log n + n/\lambda_2)$ rounds and requires $\mathcal{O}(\text{polylog}n + 1/\lambda_2)$ work per node. Inspired by our work and [9], Mallmann-Trenn et al. in [52] consider Oja's classic iterative method for principal components analysis [61], to derive some asynchronous protocols that approximate the k largest eigenvectors of a graph. Then, they use them to define a more complex version of the AVERAGING protocol that gets weak reconstruction over clustered graphs including the $\mathcal{G}_{2n,p,q}$ model for a wide range of parameters p and q.

According to the aim of "sparsification" discussed above, in [69], Sun and Zanetti propose a distributed synchronous algorithm that takes a clustered graph as input and constructs a static, sparse random subgraph that, under some spectral conditions of the original graph (satisfied by the stochastic block model), preserves the original community structure. Then they apply a suitable averaging protocol on the output subgraph. The resulting protocol works even in the case of more communities and returns a weak reconstruction within $\mathcal{O}(\text{polylog}n)$ time and work per-node.

Another interesting direction inspired by our work is the rigorous analysis of well-

[16] Notice that the opportunistic communication model is sequential so a meaningful comparison with our parallel, synchronous model requires dividing the time bound by a factor of n.

known (non-linear) dynamics based on *majority rules*, when applied to graphs that exhibit community structure. In [25], Cruciani et al. consider the *2-Choices* dynamics where, at each round, every node picks two random neighbors and updates its value to the most frequent among its value and those held by its sampled neighbors. They show that if the underlying graph has a suitable core-periphery structure and the process starts in a suitable random configuration, the system reaches a metastable regime that reflects the underlying community structure. Similar results have subsequently been obtained for regular clustered graphs with dense communities in [26]. Very recently, Shimizu and Takeharu Shiraga [67] consider the *2-Choices* and another simple majority dynamics on $\mathcal{G}_{2n,p,q}$ and their ability to compute majority consensus. In short, they show that this ability undergoes a phase transition depending on the ratio q/p. One major question left open by the above works on non-linear majority dynamics is whether they can be used to get efficient community detection on $\mathcal{G}_{2n,p,q}$.

REFERENCES

[1] E. Abbe, *Community detection and stochastic block models: Recent developments*, Journal of Mach. Learn. Res., 18 (2017), pp. 177:1–177:86, http://jmlr.org/papers/v18/16-480.html.

[2] E. Abbe, A. S. Bandeira, and G. Hall, *Exact recovery in the stochastic block model*, IEEE Trans. on Information Theory, 62 (2014), pp. 471–487.

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

[4] Y. Afek, N. Alon, O. Barad, E. Hornstein, N. Barkai, and Z. Bar-Joseph, *A biological solution to a fundamental distributed computing problem*, Science, 331 (2011), pp. 183–185.

[5] A. Amit and N. Linial, *Random graph coverings I: General theory and graph connectivity*, Combinatorica, 22 (2002), pp. 1–18.

[6] D. Angluin, J. Aspnes, and D. Eisenstat, *A Simple Population Protocol for Fast Robust Approximate Majority*, Distributed Computing, 21 (2008), pp. 87–102.

[7] T. C. Aysal, M. E. Yildiz, A. D. Sarwate, and A. Scaglione, *Broadcast gossip algorithms for consensus*, IEEE Trans. on Signal Processing, 57 (2009), pp. 2748–2761.

[8] M. J. Barber and J. W. Clark, *Detecting network communities by propagating labels under constraints*, Physical Review E, 80 (2009), p. 026129.

[9] L. Becchetti, A. E. Clementi, P. Manurangsi, E. Natale, F. Pasquale, P. Raghavendra, and L. Trevisan, *Average whenever you meet: Opportunistic protocols for community detection*, in Proc. of the 26th Annual European Symposium on Algorithms, ESA 2018, 2018, pp. 7:1–7:13, https://doi.org/10.4230/LIPIcs.ESA.2018.7, https://doi.org/10.4230/LIPIcs.ESA.2018.7.

[10] L. Becchetti, A. E. Clementi, E. Natale, F. Pasquale, and L. Trevisan, *Find your place: Simple distributed algorithms for community detection*, in Proc. of the 28th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017, SIAM, 2017, pp. 940–959, https://doi.org/10.1137/1.9781611974782.59, https://doi.org/10.1137/1.9781611974782.59.

[11] L. Becchetti, E. Cruciani, F. Pasquale, and S. Rizzo, *Step-by-step community detection in volume-regular graphs*, in Proc. of 30th International Symposium on Algorithms and Computation, ISAAC 2019, 2019, pp. 20:1–20:23.

[12] F. Bénézit, P. Thiran, and M. Vetterli, *Interval consensus: from quantized gossip to voting*, in Proc. of the 34th IEEE Int. Conference on Acoustics, Speech and Signal Processing ICASSP 2009, IEEE, 2009, pp. 3661–3664.

[13] A. C. Berry, *The accuracy of the gaussian approximation to the sum of independent variates*, Trans. of the American Mathematical Society, 49 (1941), pp. 122–136.

[14] R. B. Boppana, *Eigenvalues and graph bisection: An average-case analysis*, in Proc. of the 28th Annual IEEE Symposium on Foundations of Computer Science FOCS 1987, IEEE, 1987, pp. 280–285.

[15] C. Bordenave. Personal communication, 2015.

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

[17] C. Bordenave, M. Lelarge, and L. Massoulié, *Non-backtracking spectrum of random graphs: community detection and non-regular ramanujan graphs*, in Proc. of 56rd Annual IEEE Symposium on Foundations of Computer Science FOCS 2015, IEEE, 2015, pp. 1347–1357.

[18] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, *Randomized gossip algorithms*, IEEE/ACM Trans. on Networking, 14 (2006), pp. 2508–2530, https://doi.org/10.1109/TIT.2006.874516, http://dx.doi.org/10.1109/TIT.2006.874516 (accessed 2015-10-18).

[19] G. Brito, I. Dumitriu, S. Ganguly, C. Hoffman, and L. V. Tran, *Recovery and rigidity in a regular stochastic block model*, in Proc. of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015, SIAM, 2015, pp. 371–390.

[20] L. Cardelli and A. Csikász-Nagy, *The Cell Cycle Switch Computes Approximate Majority*, Scientific Reports, Vol. 2 (2012).

[21] J. Cheeger, *A lower bound for the smallest eigenvalue of the laplacian*, in Proc. of the Princeton conference in honor of Professor S. Bochner, 1969, pp. 195–199.

[22] F. R. Chung, *Laplacians of graphs and Cheeger's inequalities*, in Combinatorics, Paul Erdös is eighty, vol. 2, János Bolyai Math. Soc., 1996, pp. 157–172, http://www.ams.org/mathscinet-getitem?mr=1395858 (accessed 2015-09-26).

[23] A. Coja-Oghlan, *Spectral techniques, semidefinite programs, and random graphs*, Habilitation thesis, Humboldt University Berlin, 2005, http://www.math.uni-frankfurt.de/~acoghlan/habil.pdf.

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

[25] E. Cruciani, E. Natale, A. Nusser, and G. Scornavacca, *Phase transition of the 2-choices dynamics on core-periphery networks*, in Proc. of the 17th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS 2018, 2018, pp. 777–785, http://dl.acm.org/citation.cfm?id=3237499.

[26] E. Cruciani, E. Natale, and G. Scornavacca, *On the metastability of quadratic majority dynamics on clustered graphs and its biological implications*, Bulletin of the EATCS, 125 (2018), http://eatcs.org/beatcs/index.php/beatcs/article/view/535.

[27] 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 (2011), p. 066106.

[28] M. H. DeGroot, *Reaching a consensus*, Journal of the American Statistical Association, 69 (1974), pp. 118–121.

[29] D. Doty, *Timing in chemical reaction networks*, in Proc. of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms SODA 2014, SIAM, 2014, pp. 772–784.

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

[31] D. Easley and J. Kleinberg, *Networks, crowds, and markets: Reasoning about a highly connected world*, Cambridge University Press, 2010.

[32] N. R. Franks, S. C. Pratt, E. B. Mallon, N. F. Britton, and D. J. Sumpter, *Information flow, opinion polling and collective intelligence in house–hunting social insects*, Philosophical Trans. of the Royal Society of London B: Biological Sciences, 357 (2002), pp. 1567–1583.

[33] N. E. Friedkin and E. C. Johnsen, *Social influence and opinions*, The Journal of Mathematical Sociology, 15 (1990), pp. 193–206.

[34] J. Friedman and D.-E. Kohler, *The relativized second eigenvalue conjecture of Alon*, arXiv preprint arXiv:1403.3462, (2014).

[35] Y. Hassin and D. Peleg, *Distributed probabilistic polling and applications to proportionate agreement*, Information and Computation, 171 (2001), pp. 248–268.

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

[37] S. Hoory, N. Linial, and A. Wigderson, *Expander graphs and their applications*, Bulletin of the American Mathematical Society, 43 (2006), pp. 439–561.

[38] O. G.-G. (http://mathoverflow.net/users/1061/ori-gurel gurevich), *Expectation of square root of binomial r.v.* MathOverflow, http://mathoverflow.net/q/121424, https://arxiv.org/abs/http://mathoverflow.net/q/121424. URL:http://mathoverflow.net/q/121424 (version: 2013-02-10).

[39] M. Jerrum and A. Sinclair, *Approximating the permanent*, SIAM Journal on Computing, 18 (1989), pp. 1149–1178.

[40] M. Jerrum and G. B. Sorkin, *The metropolis algorithm for graph bisection*, Discrete Applied Mathematics, 82 (1998), pp. 155–175, https://doi.org/10.1016/S0166-218X(97)00133-9,

http://www.sciencedirect.com/science/article/pii/S0166218X97001339.

[41] D. Kempe, A. Dobra, and J. Gehrke, *Gossip-based computation of aggregate information*, in Proc. of the 43rd Annual IEEE Symposium on Foundations of Computer Science FOCS 2003, IEEE, 2003, pp. 482–491.

[42] D. Kempe and F. McSherry, *A decentralized algorithm for spectral analysis*, in Proc. of the 36th Annual ACM Symposium on Theory of Computing STOC 2004, 2004, pp. 561–568.

[43] J. M. Kleinberg, *Authoritative sources in a hyperlinked environment*, Journal of the ACM, 46 (1999), pp. 604–632.

[44] K. Kothapalli, S. V. Pemmaraju, and V. Sardeshmukh, *On the analysis of a label propagation algorithm for community detection*, in Proc. of the 14th International Conference on Distributed Computing and Networking ICDCN 2013, 2013, pp. 255–269.

[45] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang, *Spectral redemption in clustering sparse networks*, Proc. of the National Academy of Sciences, 110 (2013), pp. 20935–20940.

[46] C. Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, United States Governm. Press Office Los Angeles, CA, 1950.

[47] C. M. Le, E. Levina, and R. Vershynin, *Concentration and regularization of random graphs*, Random Structures & Algorithms, 51 (2017), pp. 538–561, https://doi.org/10.1002/rsa.20713, https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.20713, https://arxiv.org/abs/https://onlinelibrary.wiley.com/doi/pdf/10.1002/rsa.20713.

[48] N. Linial, *Locality in distributed graph algorithms*, SIAM Journal of Computing, 21 (1992), pp. 193–201, https://doi.org/10.1137/0221015, http://dx.doi.org/10.1137/0221015.

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

[50] D. J. MacKay, *Information theory, inference and learning algorithms*, Cambridge university press, 2003.

[51] G. Malewicz, M. H. Austern, A. J. Bik, J. C. Dehnert, I. Horn, N. Leiser, and G. Czajkowski, *Pregel: A system for large-scale graph processing*, in Proc. of the ACM SIGMOD International Conference on Management of Data SIGMOD 2010, ACM, 2010, pp. 135–146, https://doi.org/10.1145/1807167.1807184, http://doi.acm.org/10.1145/1807167.1807184.

[52] F. Mallmann-Trenn, C. Musco, and C. Musco, *Eigenvector computation and community detection in asynchronous gossip models*, in Proc. of the 45th International Colloquium on Automata, Languages, and Programming, ICALP 2018, 2018, pp. 159:1–159:14, https://doi.org/10.4230/LIPIcs.ICALP.2018.159, https://doi.org/10.4230/LIPIcs.ICALP.2018.159.

[53] L. Massoulie, *Community Detection Thresholds and the Weak Ramanujan Property*, in Proc. of the 46th Annual ACM Symposium on Theory of Computing STOC 2014, 2014, pp. 694–703.

[54] F. McSherry, *Spectral partitioning of random graphs*, in Proc. of the 42nd Annual IEEE Symposium on Foundations of Computer Science FOCS 2001, 2001, pp. 529–537.

[55] Y. Métivier, J. M. Robson, N. Saheb-Djahromi, and A. Zemmari, *An optimal bit complexity randomized distributed mis algorithm*, Distributed Computing, 23 (2011), pp. 331–340.

[56] J. M. Mooij and H. J. Kappen, *Sufficient conditions for convergence of the sum–product algorithm*, IEEE Trans. on Information Theory, 53 (2007), pp. 4422–4437.

[57] E. Mossel, J. Neeman, and A. Sly, *Reconstruction and estimation in the planted partition model*, Probability Theory and Related Fields, 162 (2014), pp. 431–461, https://doi.org/10.1007/s00440-014-0576-6.

[58] E. Mossel, J. Neeman, and A. Sly, *Belief propagation, robust reconstruction and optimal recovery of block models*, Ann. Appl. Probab., 26 (2016), pp. 2211–2256, https://doi.org/10.1214/15-AAP1145, http://dx.doi.org/10.1214/15-AAP1145.

[59] E. Mossel, J. Neeman, and A. Sly, *A proof of the block model threshold conjecture*, Combinatorica, 38 (2018), pp. 665–708.

[60] E. Mossel, J. Neeman, and O. Tamuz, *Majority dynamics and aggregation of information in social networks*, Autonomous Agents and Multi-Agent Systems, 28 (2014), pp. 408–429.

[61] E. Oja, *Simplified neuron model as a principal component analyzer*, Journal of Mathematical Biology, 15 (1982), pp. 267–273, https://doi.org/10.1007/BF00275687, https://doi.org/10.1007/BF00275687.

[62] A. Olshevsky and J. N. Tsitsiklis, *Convergence speed in distributed consensus and aver-*

aging, SIAM Journal on Control and Optimization, 48 (2009), pp. 33–55.

[63] R. Peng, H. Sun, and L. Zanetti, *Partitioning well-clustered graphs: Spectral clustering works!*, SIAM Journal on Computing, 46 (2017), pp. 710–743, https://doi.org/10.1137/15M1047209, https://doi.org/10.1137/15M1047209, https://arxiv.org/abs/https://doi.org/10.1137/15M1047209.

[64] U. N. Raghavan, R. Albert, and S. Kumara, *Near linear time algorithm to detect community structures in large-scale networks*, Physical Review E, 76 (2007), p. 036106.

[65] S. Salihoglu, J. Shin, V. Khanna, B. Q. Truong, and J. Widom, *Graft: A debugging tool for apache giraph*, in Proc. of the ACM SIGMOD International Conference on Management of Data SIGMOD 2015, ACM, 2015, pp. 1403–1408, https://doi.org/10.1145/2723372.2735353, http://doi.acm.org/10.1145/2723372.2735353.

[66] D. Shah, *Gossip algorithms*, Now Publishers Inc, 2009.

[67] N. Shimizu and T. Shiraga, *Phase Transitions of Best-of-Two and Best-of-Three on Stochastic Block Models*, in Proc. of 33rd International Symposium on Distributed Computing DISC 2019, vol. 146 of Leibniz International Proceedings in Informatics (LIPIcs), Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019, pp. 32:1–32:17, https://doi.org/10.4230/LIPIcs.DISC.2019.32, http://drops.dagstuhl.de/opus/volltexte/2019/11339.

[68] P. Stewart and J.-g. Sun, *Matrix Perturbation Theory*, Academic Press, 1990.

[69] H. Sun and L. Zanetti, *Distributed graph clustering and sparsification*, CoRR, abs/1711.01262 (2017), http://arxiv.org/abs/1711.01262, https://arxiv.org/abs/1711.01262.

[70] J. Suomela, *Survey of local algorithms*, ACM Computing Surveys, 45 (2013), pp. 24:1–24:40, https://doi.org/10.1145/2431211.2431223, http://doi.acm.org/10.1145/2431211.2431223.

[71] J. P. Tian and D. Kannan, *Lumpability and commutativity of Markov processes*, Stochastic analysis and Applications, 24 (2006), pp. 685–702.

[72] J. N. Tsitsiklis, *Problems in decentralized decision making and computation.*, tech. report, DTIC Document, 1984.

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

[74] L. Xiao, S. Boyd, and S.-J. Kim, *Distributed average consensus with least-mean-square deviation*, Journal of Parallel and Distributed Computing, 67 (2007), pp. 33–46.

Appendix

Appendix A. Linear algebra toolkit.

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

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

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

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

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

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

THEOREM A.2. *(Corollary 4.10 in [68]) Let* M_1 *and* M_2 *be two Hermitian matrices, let* $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$ *be the eigenvalues of* M_1 *with multiplicities in non-increasing order, and let* $\lambda'_1 \geqslant \lambda'_2 \geqslant \cdots \geqslant \lambda'_n$ *be the eigenvalues of* M_2 *with multiplicities in non-increasing order. Then, for every* i*,*

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

THEOREM A.3 (Berry and Essen, Berry's formulation [13]). *Consider* n *independent random variables* $X_1, \ldots, X_n$ *such that for every* i*: i)* $\mathbf{E}[X_i] = 0$*, ii)* $\sigma_i^2 = \mathbf{E}\left[X_i^2\right] > 0$*, and iii)* $\rho_i = \mathbf{E}\left[|X_i|^3\right]$ *is finite. Let*

$$\sigma^2 = \sum_i \sigma_i^2 \quad \textit{and} \quad \psi = \frac{\max_i \frac{\rho_i}{\sigma_i^2}}{\sigma} \,.$$

Denote by $F(x)$*, the cumulative distribution function (for short, c.d.f.) of the variable* $X = \sum_{i=1}^n X_i$ *and by* $G(x)$ *the c.d.f. of a Gaussian distribution with mean* 0 *and variance* σ^2*. Then there exists an absolute constant* $C_0 > 0$ *such that*

$$\sup_x |F(x) - G(x)| \leqslant C_0 \psi \,.$$

THEOREM A.4 (Davis and Kahan, 1970). *Let* M_1 *and* M_2 *be two symmetric real matrices, let* $\mathbf{x}$ *be a unit length eigenvector of* M_1 *of eigenvalue* t*, and let* $\mathbf{x}_p$ *be the projection of* $\mathbf{x}$ *on the eigenspace of the eigenvectors of* M_2 *corresponding to eigenvalues* $\leqslant t - \varepsilon$*. Then*

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

Appendix B. Length of the projection of x.

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

LEMMA B.1. *Let $\mathbf{y} \in \{-1,1\}^{2n}$ an arbitrary vector with ± 1 entries. If we pick $\mathbf{x}$ uniformly at random in $\{-1,1\}^{2n}$ then, for any $\delta > 0$,*

$$\mathbf{P}\left(\left|(1/\sqrt{2n})\,\mathbf{x}^{\intercal}\,\mathbf{y}\right| \leqslant \delta\right) \leqslant \mathcal{O}(\delta)\,.$$

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

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

LEMMA B.2. *Let $k, m \in \mathbb{N}$ be two integers and let $\mathbf{y}$ be an mk-dimensional vector with real entries. Given two positive real numbers r and c, let S be the subset of coordinates of $\mathbf{y}$ whose absolute value is between r and cr, i.e.,*

$$S = \{i \in \{1,\ldots,km\} \,:\, r \leqslant |y(i)| \leqslant cr\}\,.$$

If $|S| \geqslant m$ and we pick $\mathbf{x}$ uniformly at random in $\{-1,1\}^{km}$, then for every $\delta > 0$

$$\mathbf{P}\left(\left|(1/\sqrt{km})\,\mathbf{x}^{\intercal}\,\mathbf{y}\,\right| \leqslant \delta\right) \leqslant \sqrt{\frac{2k}{\pi}}\cdot\frac{\delta}{r} + \frac{4c}{\sqrt{m}}\,.$$

Proof. Let $T = \{1,\ldots,km\} - S$. In the remainder of this proof, the restriction of $\mathbf{x}$ to entries in T will be denoted by $\mathbf{x}_T$. Clearly, $\mathbf{x}_T$ is independent of $x(i)$ for every $i \in S$. Next, we prove that, for any fixed $\mathbf{a} \in \{-1,1\}^{|T|}$,

$$\mathbf{P}\left(\,|\mathbf{x}^{\intercal}\mathbf{y}| \leqslant \delta\sqrt{km} \;\;\middle|\;\; \mathbf{x}_T = \mathbf{a}\right) \leqslant \sqrt{\frac{2k}{\pi}}\cdot\frac{\delta}{r} + \frac{4c}{\sqrt{m}}\,. \tag{B.1}$$

Observe that the lemma is a consequence of the above bound since

$$\begin{aligned}\mathbf{P}\left(|\mathbf{x}^{\intercal}\mathbf{y}| \leqslant \delta\sqrt{km}\right) &= \sum_{\mathbf{a}\in\{-1,1\}^{|T|}} \mathbf{P}\left(\,|\mathbf{x}^{\intercal}\mathbf{y}| \leqslant \delta\sqrt{km} \;\;\middle|\;\; \mathbf{x}_T = \mathbf{a}\right)\mathbf{P}\left(\mathbf{x}_T = \mathbf{a}\right)\\ &\leqslant \left(\sqrt{\frac{2k}{\pi}}\cdot\frac{\delta}{r} + \frac{4c}{\sqrt{m}}\right)\sum_{\mathbf{a}\in\{-1,1\}^{|T|}} \mathbf{P}\left(\mathbf{x}_T = \mathbf{a}\right) = \sqrt{\frac{2k}{\pi}}\cdot\frac{\delta}{r} + \frac{4c}{\sqrt{m}},\end{aligned}$$

where the last equality follows since the $\mathbf{P}\left(\mathbf{x}_T = \mathbf{a}\right)$'s obviously sum to 1. In order to prove (B.1), we define

$$t = \sum_{i\in T} a(i)y(i)\,,$$

and show that, conditioned on $\mathbf{x}_T = \mathbf{a}$, with $\mathbf{a} \in \{-1,1\}^{|T|}$,

$$\mathbf{P}\left(\,|\sum_{i\in S} x(i)y(i) + t\,| \leqslant \delta\sqrt{km} \;\;\middle|\;\; \mathbf{x}_T = \mathbf{a}\right) \leqslant \sqrt{\frac{2k}{\pi}}\cdot\frac{\delta}{r} + \frac{4c}{\sqrt{m}}\,.$$

To this purpose, we first write

$$\mathbf{P}\left(\mid \sum_{i\in S} x(i)y(i) + t \mid \leqslant \delta\sqrt{km} \quad \middle| \quad \mathbf{x}_T = \mathbf{a} \right) =$$
$$= \mathbf{P}\left(-t - \delta\sqrt{km} \leqslant \sum_{i\in S} x(i)y(i) \leqslant -t + \delta\sqrt{km} \quad \middle| \quad \mathbf{x}_T = \mathbf{a} \right).$$

Next, we apply the Berry-Esseen theorem A.3, with $X_i = x(i)y(i)$, for every $i \in S$. In particular, our hypotheses on $\mathbf{y}$ and the fact the the $x(i)$'s are Rademacher random variables imply: i) $\mathbf{E}[X_i] = 0$, ii) $\sigma_i^2 = y(i)^2$, iii) $\rho_i = |y(i)|^3$. Moreover, $\rho_i/\sigma_i^2 = |y(i)|$, so that our hypotheses on $\mathbf{y}$ in turn imply $\max_i(\rho_i/\sigma_i^2) \leqslant cr$. Finally, $\sigma^2 = \sum_i \sigma_i^2 \geqslant mr^2$. As a consequence, we can apply Theorem A.3 to $\sum_{i\in S} x(i)y(i)$, with $\psi \leqslant \frac{2c}{\sqrt{m}}$ to obtain[17]

$$\mathbf{P}\left(\left| \sum_{i\in S} x(i)y(i) + t \right| \leqslant \delta\sqrt{km} \quad \middle| \quad \mathbf{x}_T = \mathbf{a} \right) \leqslant$$
$$\leqslant \mathbf{P}\left(-t - \delta\sqrt{km} \leqslant g \leqslant -t + \delta\sqrt{km} \right) + \frac{4c}{\sqrt{m}},$$

where g is a Gaussian random variable of mean 0 and variance σ^2. Finally:

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

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

[17] In the following, we use the fact that the constant $C_0 > 0$ that appears in the claim of Theorem A.3 is less than 2 (see [13]).