

# Step-by-Step Community Detection in Volume-Regular Graphs

Luca Becchetti\*

Sapienza Universit di Roma  
Rome, Italy  
becchetti@diag.uniroma1.it

Francesco Pasquale†

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

Emilio Cruciani

Inria, I3S Lab, UCA, CNRS  
Sophia Antipolis, France  
emilio.cruciani@inria.fr

Sara Rizzo

Gran Sasso Science Institute  
L'Aquila, Italy  
sara.rizzo@gssi.it

## Abstract

Spectral techniques have proved amongst the most effective approaches to graph clustering. However, in general they require explicit computation of the main eigenvectors of a suitable matrix (usually the Laplacian matrix of the graph).

Recent work (e.g., Becchetti et al., SODA 2017) suggests that observing the temporal evolution of the power method applied to an initial random vector may, at least in some cases, provide enough information on the space spanned by the first two eigenvectors, so as to allow recovery of a hidden partition without explicit eigenvector computations. While the results of Becchetti et al. apply to perfectly balanced partitions and/or graphs that exhibit very strong forms of regularity, we extend their approach to graphs containing a hidden  $k$  partition and characterized by a milder form of volume-regularity. We show that the class of  $k$ -volume regular graphs is the largest class of undirected (possibly weighted) graphs whose transition matrix admits  $k$  “stepwise” eigenvectors (i.e., vectors that are constant over each set of the hidden partition). To obtain this result, we highlight a connection between volume regularity and lumpability of Markov chains. Moreover, we prove that if the stepwise eigenvectors are those associated to the first  $k$  eigenvalues and the gap between the  $k$ -th and the  $(k+1)$ -th eigenvalues is sufficiently large, the AVERAGING dynamics of Becchetti et al. recovers the underlying community structure of the graph in logarithmic time, with high probability.

**Keywords:** Distributed algorithms, Community detection, Markov chains, Spectral analysis

---

\*Partially supported by ERC Advanced Grant 788893 AMDROMA “Algorithmic and Mechanism Design Research in Online Markets” and MIUR PRIN project ALGADIMAR “Algorithms, Games, and Digital Markets”

†Partially supported by the University of “Tor Vergata” under research programme “Mission: Sustainability” project ISIDE (grant no. E81I18000110005)

# 1 Introduction

Clustering a graph in a way that reflects underlying community structure is a very important mining task [For10]. Informally speaking, in the classical setting, we are given a possibly weighted graph  $G$  and an integer  $k$ . Our goal is to partition the vertex set of  $G = (V, E)$  into  $k$  disjoint subsets, so that the  $k$  induced subgraphs have high inner and low outer expansion. Spectral techniques have proved amongst the most effective approaches to graph clustering [NJW02, SM00, VL07]. The general approach to spectral graph clustering [VL07] normally implies embedding the vertices of  $G$  into the  $k$ -dimensional subspace spanned by the main  $k$  eigenvectors of a matrix defined in terms of  $G$ 's adjacency matrix, typically its (normalized) Laplacian. Intuitively, one expects that, for a well-clustered graph with  $k$  communities, the profiles of the first  $k$  eigenvectors are correlated with the underlying community structure of  $G$ . Recent work has provided theoretical support to this approach. In particular, [LGT14] showed that, given the first  $k$  orthonormal eigenvectors of the normalized Laplacian, it is possible to produce a  $k$ -partition of the vertex set, corresponding to  $k$  suitably-defined indicator vectors, such that the associated values of the Rayleigh quotient are relatively small. More recently, [PSZ17] proved that, under suitable hypotheses on the spectral gap between the  $k$ -th and  $(k+1)$ -th eigenvalue of the normalized Laplacian of  $G$ , the span of the first  $k$  eigenvectors largely overlaps with the span of  $\{D^{\frac{1}{2}}\mathbf{g}_1, \dots, D^{\frac{1}{2}}\mathbf{g}_k\}$ , where  $D$  is the diagonal degree matrix of  $G$ , while the  $\mathbf{g}_i$ 's are indicator vectors describing a  $k$ -way partition  $\{S_i\}_{i=1}^k$  of  $V$  such that, for every  $i$ , the conductance of  $S_i$  is at most the  $k$ -way expansion constant  $\rho(k)$  [LGT14]. Note that, if  $\mathbf{v}$  is an eigenvector associated to the  $i$ -th smallest eigenvalue of the normalized Laplacian,  $D^{-\frac{1}{2}}\mathbf{v}$  is an eigenvector corresponding to the  $i$ -th largest eigenvalue of the random walk's transition matrix associated to  $G$ . Hence, when  $G$  is well-clustered, one might reasonably expect the first  $k$  eigenvectors of  $P$  to exhibit almost-“stepwise” profiles reflecting  $G$ 's underlying community structure. The aforementioned spectral approaches require explicit computation of the  $k$  main eigenvectors of a (generally symmetric) matrix.

In [BCN<sup>+</sup>17], the authors considered the case  $k = 2$  for which they proposed the following distributed algorithm (AVERAGING dynamics, Algorithm 1): “At the outset, every node picks an initial value, independently and uniformly at random in  $\{-1, 1\}$ ; then, in each synchronous round, every node updates its value to the average of those held by its neighbors. A node also tags itself **blue** if the last update increased its value, **red** otherwise” [BCN<sup>+</sup>17]. The authors showed that, under a variety of graph models exhibiting sparse balanced cuts, including the *stochastic block model* [HLL83], the process resulting from the above simple local rule converges, in logarithmic time, to a coloring that, depending on the model, exactly or approximately reflects the underlying cut. They further elaborated on how to extend the proposed approach to the case of multiple communities, providing an analysis for a strongly regular version of the stochastic block model with multiple communities. While results like those presented in [LGT14, PSZ17] provide further theoretical justification for spectral clustering, the approach proposed in [BCN<sup>+</sup>17] suggests that observing the temporal evolution of the power method applied to an initial random vector may, at least in some cases, provide equivalent information, without requiring explicit eigenvector computations.

## 1.1 Our contributions

The goal of this work is to take a further step in this direction by considering a more general class of graphs, even if still relatively “regular”, than the one considered in [BCN<sup>+</sup>17]. The analysis of the AVERAGING dynamics on this class is considerably harder, but it is likely to provide insights into the challenges of analyzing the general case, without all the intricacies of

the latter. Our contribution is as follows:

- We define the class of *k-volume-regular* graphs. This class of edge-weighted graphs includes those considered in [BCN<sup>+</sup>17] and it is the largest class of undirected, possibly weighted graphs that admit *k* “stepwise” eigenvectors (i.e., having constant values over the *k* steps that identify the hidden partition). This result uses a connection between volume regularity and lumpability of Markov chains [KS60, TK06].
- If the stepwise eigenvectors are those associated to the first *k* eigenvalues and the gap between the *k*-th and the (*k*+1)-th eigenvalues is sufficiently large, we show that running the AVERAGING dynamics for a suitable number of steps allows recovery of the underlying community structure of the graph, with high probability.<sup>1</sup> To prove this, we provide a family of mutually orthonormal vectors which, when the graph is volume-regular, span the eigenspace of the main *k* eigenvectors of the normalized adjacency matrix of the graph. It should be noted that the first and second of these vectors are respectively the main eigenvector and the Fiedler vector [Fie89] associated to the normalized adjacency matrix.
- While the results of [BCN<sup>+</sup>17] apply when the underlying communities are of the same size, our results do not require this assumption and they apply to weighted graphs. It should also be noted that volume regularity is a weaker notion than regularity of the graph.
- We further show that variants of the AVERAGING dynamics (and/or its labeling rule) can address different problems (e.g., identifying bipartiteness) and/or other graph classes.

We further note that the overall algorithm we consider can be viewed as a fully decentralized, synchronous algorithm that works in *anonymous* networks,<sup>2</sup> with a completely local clustering criterion, though it cannot be considered a *dynamics* in the sense of [BCN<sup>+</sup>17] since it requires a bound on the number of nodes in the underlying network.

Finally, this paper extends a preliminary version [BCPR19] in several ways. To begin, the main result presented in [BCPR19] was weaker, in the sense that the constraints imposed on the eigenvalues in [BCPR19, Theorem 9] polynomially depend on network parameters like the maximum degree and the number of vertices. In this respect, they are substantially stronger than those imposed to prove Theorem 4.1, where results (in particular, the time window in which recovery of the hidden partition is possible) are expressed in terms of the spectrum of the graph, while constraints imposed on the second eigenvalue only logarithmically depend on the aforementioned network parameters. In reframing these results, we also realized that the presence of a window in which recovery is possible is something that is hardly avoidable in general using the simple averaging heuristic of [BCN<sup>+</sup>17]. This is something we remark right after Theorem 4.1 (see Remark 1), while we also observe (see Remarks 3 and 4) that the analysis presented here also encompasses the class of regular graphs considered in [BCN<sup>+</sup>17] as a special case, something that was not obvious in [BCPR19]. Finally, the result given in [BCPR19] for bipartite graphs assumed volume regularity, an assumption that is not necessary as we show in Section 5.

## 1.2 Further related work

We briefly discuss further work that bears some relationship to this paper, either because it adopts simple and/or decentralized heuristics to uncover community structure, or because it relies on the use of spectral techniques.

---

<sup>1</sup>An event  $\mathcal{E}_n$  holds *with high probability (w.h.p.)* if  $\mathbf{P}(\mathcal{E}_n) = 1 - \mathcal{O}(n^{-\gamma})$ , for some constant  $\gamma > 0$ .

<sup>2</sup>Nodes do not possess distinguished identities.

**Decentralized heuristics for block reconstruction.** *Label propagation algorithms* [RAK07] are dynamics based on majority updating rules [AAE08] and have been applied for detecting communities in complex networks. Several papers present experimental results for such protocols on specific classes of clustered graphs [BC09, LM10, RAK07]. The only available rigorous analysis of a label propagation algorithm on planted partition graphs is the one presented in [KPS13], where the authors analyze a label propagation algorithm on  $\mathcal{G}_{2n,p,q}$  graphs in the case of dense topologies. In particular, their analysis considers the case where  $p = \Omega(1/n^{\frac{1}{4}-\varepsilon})$  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. A logarithmic bound for sparser topologies is conjectured in [KPS13].

Following [BCN<sup>+</sup>17], a number of recent papers analyze simple distributed algorithms for community detection that rely on elementary dynamics. In the AVERAGING dynamics considered in this paper, every node communicates in parallel with all its neighbors in each round. While this might be too expensive in scenarios characterized by dense topologies, it is simply infeasible in other settings (for instance, when links represent opportunistic meetings that occur asynchronously). Motivated by similar considerations, a first line of follow-up work considered “sparsified”, asynchronous variants of the AVERAGING dynamics [BCM<sup>+</sup>18, MMM18, SZ17].

Another interesting direction is the rigorous analysis of well-known (non-linear) dynamics based on *majority rules* on graphs that exhibit community structure. In [CNNS18], Cruciani et al. consider the *2-Choices* dynamics where, in 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 configuration where nodes in core and periphery have different states, the system either rapidly converges to the core’s state or reaches a metastable regime that reflects the underlying graph structure. Similar results have been also obtained for clustered regular graphs with dense communities in [CNS19], where the *2-Choices* dynamics is proposed as a distributed algorithm for community detection.

Although based on the AVERAGING dynamics and thus extremely simple and fully decentralized, the algorithm we consider in this paper is not itself a dynamics in the sense proposed in [BCN<sup>+</sup>17], since its clustering criterion is applied within a time window, which in turn requires (at least approximate) knowledge of the network size.

Because of their relevance for the reconstruction problem, we also briefly discuss the class of *belief propagation algorithms*, best known as message-passing algorithms for performing inference in graphical models [Mac03]. Though not a dynamics, belief propagation is still a simple approach. Moreover, there is non-rigorous, strong supporting evidence that some *belief propagation algorithms* might be optimal for the reconstruction problem [DKMZ11]. A rigorous analysis is a major challenge; in particular, convergence to the correct value of belief propagation is far from being fully-understood on graphs which are not trees [MK07, Wei00]. As we discuss in the next subsection, more complex algorithms inspired by belief propagation have been rigorously shown to perform reconstruction optimally.

**General algorithms for block reconstruction.** Several algorithms for community detection are *spectral*: They typically consider the eigenvector associated to the second largest eigenvalue of the adjacency matrix  $A$  of  $G$ , or the eigenvector corresponding to the largest eigenvalue of the matrix  $A - \frac{d}{n}J$  [Bop87, CO05, CO10, McS01],<sup>3</sup> since these are correlated with the hid-

---

<sup>3</sup> $A$  is the adjacency matrix of  $G$ ,  $J$  is the matrix having all entries equal to 1,  $d$  is the average degree, and  $n$  is the number of vertices.

den partition. More recently spectral algorithms have been proposed [AS15, BLM15, CO10, KMM<sup>+</sup>13, MNS13, PSZ17] that find a weak reconstruction even in the sparse, tight regime.

Interestingly, spectral algorithms turn out to be a feasible approach also in distributed settings. In particular, Kempe and McSherry [KM04] show that eigenvalue computations can be performed in a distributed fashion, yielding distributed algorithms for community detection under various models, including the stochastic block model. However, their algorithm does not match any simple decentralized computing model. In particular, the algorithm of Kempe and McSherry as well as any distributed version of the above mentioned centralized algorithms are neither dynamics, nor do they correspond to the notion of *light-weight* algorithm of Hassin and Peleg [HP01]. Moreover, 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 [Lan50] as a subroutine. This is not the case for the AVERAGING dynamics, since it removes the component of the state in the span of the main eigenvector.

In general, the reconstruction problem has been studied extensively using a multiplicity of techniques, which include combinatorial algorithms [DF89], belief propagation [DKMZ11] and variants of it [MNS16], spectral-based techniques [CO10, McS01], Metropolis approaches [JS98], and semidefinite programming [ABH14], among others.

### 1.3 Roadmap

The rest of this paper is organized as follows. In Section 2, we formally define the AVERAGING dynamics and briefly recall how it is connected with the transition matrix of a random walk on the underlying graph. We also define the notion of *community-sensitive algorithm* and the class of *clustered volume-regular graphs*. In Section 3 we show the relation between lumpability of Markov chains and volume-regular graphs. In Section 4 we state the main result of the paper (see Theorem 4.1) on the analysis of the AVERAGING for clustered volume-regular graphs: We give the two main technical lemmas and show how the main theorem derives from them. In Section 5, we show how slightly modified versions of the AVERAGING dynamics can be used to identify the hidden partition of other non-clustered volume-regular graphs, e.g., bipartite graphs. In Section 6 we briefly show how our approach can be extended to slightly more general graph classes than the ones considered in this paper. We finally highlight some open problems and directions for further research on the topic.

## 2 Preliminaries

**Notation.** Consider an undirected edge-weighted graph  $G = (V, E, w)$  with nonnegative weights. For each node  $u \in V$ , we denote by  $\delta(u)$  the *volume*, or *weighted degree*, of node  $u$ , namely  $\delta(u) = \sum_{v:(u,v) \in E} w(u, v)$ . Similarly, we denote the volume of a set of nodes  $T \subseteq V$  as  $\text{vol}(T) := \sum_{u \in T} \delta(u)$ .  $D$  denotes the diagonal matrix, such that  $D_{uu} = \delta(u)$  for each  $u \in V$ . Without loss of generality we assume  $\min_u \delta(u) = 1$ , since the behavior of the AVERAGING dynamics (and the corresponding analysis) is not affected by a normalization of the weights. We refer to the maximum volume of a node as  $\Delta := \max_u \delta(u)$ .

In the remainder,  $W$  denotes the *weighted adjacency matrix* of  $G$ , while  $P = D^{-1}W$  is the *transition matrix* of a random walk on  $G$ , in which a transition from node  $u$  to node  $v$  occurs with probability proportional to  $w(u, v)$ . We call  $\lambda_1, \dots, \lambda_n$  the eigenvalues of  $P$ , in non-increasing order, and  $\mathbf{v}_1, \dots, \mathbf{v}_n$  a family of eigenvectors of  $P$ , such that  $P\mathbf{v}_i = \lambda_i\mathbf{v}_i$ . We let  $N = D^{-\frac{1}{2}}WD^{-\frac{1}{2}} = D^{\frac{1}{2}}PD^{-\frac{1}{2}}$  denote the *normalized weighted adjacency matrix* of  $G$ . Note that  $N$  is real and symmetric (thus, the eigenvectors of  $N$  are orthogonal) and that its spectrum

is the same as that of  $P$ . We denote by  $\mathbf{w}_1, \dots, \mathbf{w}_n$  a family of eigenvectors of  $N$ , such that  $N\mathbf{w}_i = \lambda_i\mathbf{w}_i$ . It is important to note that  $\mathbf{w}_i$  is an eigenvector of  $N$  if and only if  $D^{-\frac{1}{2}}\mathbf{w}_i$  is an eigenvector of  $P$ .

We use the Bachmann–Landau asymptotic notation (i.e.,  $\omega, \Omega, \Theta, \mathcal{O}, o$ ) to describe the limiting behavior of functions depending on  $n$ . In this sense, our results only hold for large  $n$ . We say that an event  $\mathcal{E}_n$  holds *with high probability* (*w.h.p.*, in short) if  $\mathbf{P}(\mathcal{E}_n) = 1 - \mathcal{O}(n^{-\gamma})$ , for any positive constant  $\gamma$ .

## 2.1 Averaging dynamics

The simple algorithm we consider in this paper, named AVERAGING dynamics (Algorithm 1) after [BCN<sup>+</sup>17] in which the algorithm was first proposed, can be seen as an application of the power method, augmented with a Rademacher initialization and a suitable labeling scheme. In this form, it is best described as a distributed process, executed by the nodes of an underlying edge-weighted graph. The AVERAGING dynamics can be used as a building-block to achieve “community detection” in some classes of “regular” and “almost regular” graphs. Herein, we extend its use and analysis to broader graph classes and, in one case, to a different problem.

---

### Algorithm 1 AVERAGING dynamics

---

**Rademacher initialization:** At round  $t = 0$ , every node  $v \in V$  independently samples its value  $\mathbf{x}^{(0)}(v)$  from  $\{-1, +1\}$  uniformly at random.

**Update rule:** At each subsequent round  $t \geq 1$ , every node  $v \in V$ :

1. *Averaging:* updates its value  $\mathbf{x}^{(t)}(v)$  to the weighted average of the values of its neighbors at the end of the previous round.
  2. *Labeling:* if  $\mathbf{x}^{(t)}(v) \geq \mathbf{x}^{(t-1)}(v)$  then  $v$  sets  $\text{label}^{(t)}(v) = 1$ ; otherwise  $v$  sets  $\text{label}^{(t)}(v) = 0$ .
- 

**Spectral decomposition of the transition matrix.** Let  $\mathbf{x}^{(t)}$  denote the *state vector* at time  $t$ , i.e., the vector whose  $u$ -th entry is the value held by node  $u$  at time  $t$ . We let  $\mathbf{x}^{(0)} = \mathbf{x}$  denote the initial state vector. Globally, the *averaging* update rule of Algorithm 1 corresponds to one iteration of the power method, in this case an application of the transition matrix  $P$  to the current state vector, i.e.,  $\mathbf{x}^{(t)} = P\mathbf{x}^{(t-1)}$ . We can write

$$\mathbf{x}^{(t)} = P^t\mathbf{x} = D^{-\frac{1}{2}}N^tD^{\frac{1}{2}}\mathbf{x} \stackrel{(a)}{=} D^{-\frac{1}{2}}\sum_{i=1}^n\lambda_i^t\mathbf{w}_i\mathbf{w}_i^\top\sum_{i=1}^n\beta_i\mathbf{w}_i \stackrel{(b)}{=} \sum_{i=1}^n\lambda_i^t\beta_iD^{-\frac{1}{2}}\mathbf{w}_i,$$

where in (a) we spectrally decomposed the matrix  $N^t$  and expressed the vector  $D^{\frac{1}{2}}\mathbf{x}$  as a linear combination of the eigenvectors of  $N$ , i.e.,  $D^{\frac{1}{2}}\mathbf{x} = \sum_{i=1}^n\beta_i\mathbf{w}_i$ , with  $\beta_i = \langle D^{\frac{1}{2}}\mathbf{x}, \mathbf{w}_i \rangle$ ; in (b) we used that the eigenvectors of  $N$  are orthonormal, i.e., that  $\mathbf{w}_i^\top\mathbf{w}_i = 1$  for every  $i \in \{1, \dots, n\}$  and that  $\mathbf{w}_i^\top\mathbf{w}_j = 0$  for every  $i, j \in \{1, \dots, n\}$  and such that  $i \neq j$ . By explicitly writing the  $\beta_i$ s and by noting that  $\mathbf{w}_i = \frac{D^{\frac{1}{2}}\mathbf{v}_i}{\|D^{\frac{1}{2}}\mathbf{v}_i\|}$  we conclude that

$$\mathbf{x}^{(t)} = \sum_{i=1}^n\lambda_i^t\frac{\langle D^{\frac{1}{2}}\mathbf{x}, D^{\frac{1}{2}}\mathbf{v}_i \rangle}{\|D^{\frac{1}{2}}\mathbf{v}_i\|}D^{-\frac{1}{2}}\frac{D^{\frac{1}{2}}\mathbf{v}_i}{\|D^{\frac{1}{2}}\mathbf{v}_i\|} = \sum_{i=1}^n\lambda_i^t\alpha_i\mathbf{v}_i, \quad (1)$$

where  $\alpha_i := \frac{\langle D^{\frac{1}{2}}\mathbf{x}, D^{\frac{1}{2}}\mathbf{v}_i \rangle}{\|D^{\frac{1}{2}}\mathbf{v}_i\|^2} = \frac{\mathbf{x}^\top D\mathbf{v}_i}{\|D^{\frac{1}{2}}\mathbf{v}_i\|^2}$  is the length of the projection of  $D^{\frac{1}{2}}\mathbf{x}$  on  $D^{\frac{1}{2}}\mathbf{v}_i$ .

Note that  $\lambda_1 = 1$  and  $\mathbf{v}_1 = \mathbf{1}$ ,<sup>4</sup> since  $P$  is stochastic, and  $\lambda_i \in (-1, 1)$  for every  $i > 1$ , if  $G$  is connected and non bipartite. The long term behavior of the dynamics can be written as

$$\lim_{t \rightarrow \infty} \mathbf{x}^{(t)} = \lim_{t \rightarrow \infty} \sum_{i=1}^n \lambda_i^t \alpha_i \mathbf{v}_i = \alpha_1 \mathbf{1}, \quad \text{with } \alpha_1 = \frac{\sum_{u \in V} \delta(u) \mathbf{x}(u)}{\sum_{u \in V} \delta(u)} = \sum_{u \in V} \frac{\delta(u)}{\text{vol}(V)} \mathbf{x}(u),$$

i.e., each node converges to the initial global weighted average of the network.

## 2.2 Community-sensitive algorithms

We give the following definition of *community-sensitive algorithm*, that closely resembles that of locality-sensitive hashing (see, e.g., [LRU14]).

**Definition 2.1** (Community-sensitive algorithm). *Let  $\mathcal{A}$  be a randomized algorithm that takes in input a (possibly weighted) graph  $G = (V, E)$  with a hidden partition  $\mathcal{V} = \{V_1, \dots, V_k\}$  and assigns a Boolean value  $\mathcal{A}(G)[v] \in \{0, 1\}$  to each node  $v \in V$ . We say  $\mathcal{A}$  is an  $(\varepsilon, \delta)$ -community-sensitive algorithm, for some  $\varepsilon, \delta > 0$ , if the following two conditions hold:*

1. *For each set  $V_i$  of the partition and for each pair of nodes  $u, v \in V_i$  in that set, the probability that the algorithm assigns the same Boolean value to  $u$  and  $v$  is at least  $1 - \varepsilon$ :*

$$\forall i \in [k], \forall u, v \in V_i, \mathbf{P}(\mathcal{A}(G)[u] = \mathcal{A}(G)[v]) \geq 1 - \varepsilon.$$

2. *For each pair  $V_i, V_j$  of distinct sets of the partition and for each pair of nodes  $u \in V_i$  and  $v \in V_j$ , the probability that the algorithm assigns the same value to  $u$  and  $v$  is at most  $\delta$ :*

$$\forall i, j \in [k] \text{ with } i \neq j, \forall u \in V_i, \forall v \in V_j, \mathbf{P}(\mathcal{A}(G)[u] = \mathcal{A}(G)[v]) \leq \delta.$$

For example, for  $(\varepsilon, \delta) = (1/n, 1/2)$ , an algorithm that simply assigns the same value to all nodes would satisfy the first condition but not the second one, while an algorithm assigning 0 or 1 to each node with probability 1/2, independently of the other nodes, would satisfy the second condition but not the first one.

Note that Algorithm 1 is a distributed algorithm that, at each round  $t$ , assigns one out of two labels to each node of a graph. In the next section (see Theorem 4.1) we prove that a time window  $[T_1, T_2]$  exists, such that for all rounds  $t \in [T_1, T_2]$ , the assignment of the AVERAGING dynamics satisfies both conditions in Definition 2.1: The first condition with  $\varepsilon = \varepsilon(n) = \mathcal{O}(n^{-\frac{1}{2}})$ , the second with  $\delta = \delta(n) = 1 - \Omega(1)$ .

**Community-sensitive labeling.** We here generalize the concept of *community-sensitive labeling* (appeared in [BCM<sup>+</sup>18, Definition 3]), given only for the case of two communities, to the case of multiple communities. If we execute  $\ell = \Theta(\log n)$  independent runs of an  $(\varepsilon, \delta)$ -community-sensitive algorithm  $\mathcal{A}$ , each node is assigned a *signature* of  $\ell$  binary values, with pairwise Hamming distances probabilistically reflecting community membership of the nodes. More precisely, let  $\mathcal{A}$  be an  $(\varepsilon, \delta)$ -community-sensitive algorithm and let  $\mathcal{A}_1, \dots, \mathcal{A}_\ell$  be  $\ell = \Theta(\log n)$  independent runs of  $\mathcal{A}$ . For each node  $u \in V$ , let  $\mathbf{s}(u) = (s_1(u), \dots, s_\ell(u))$  denote the *signature* of node  $u$ , where  $s_i(u) = \mathcal{A}_i(G)[u]$ . For each pair nodes  $u, v$ , let  $h(u, v) = |\{i \in [\ell] : s_i(u) \neq s_i(v)\}|$  be the Hamming distance between  $\mathbf{s}(u)$  and  $\mathbf{s}(v)$ .

<sup>4</sup>Here and in the remainder,  $\mathbf{1}$  denotes the vector whose entries are 1.

**Lemma 2.2** (Community-sensitive labeling). *Let  $\mathcal{A}$  be an  $(\varepsilon, \delta)$ -community-sensitive algorithm with  $\varepsilon = \mathcal{O}(\frac{1}{n^\gamma})$  for any arbitrarily small positive constant  $\gamma$ , and  $\delta = 1 - \Omega(1)$ . Let  $\ell = \Theta(\log n)$ ,  $\alpha = \Omega(\frac{1}{n^{\gamma-c}})$  with  $c \in (0, \gamma]$ , and  $\beta = b(1 - \delta)$  for any constant  $b \in (0, 1)$  and such that  $0 \leq \alpha \leq \beta \leq 1$ . Then, for each pair of nodes  $u, v \in V$  it holds that:*

1. *If  $u$  and  $v$  belong to the same community then  $h(u, v) < \alpha\ell$ , w.h.p.*
2. *If  $u$  and  $v$  belong to different communities then  $h(u, v) \geq \beta\ell$ , w.h.p.*

*Proof.* From the definition of  $(\varepsilon, \delta)$ -community-sensitive algorithm we have that, if  $u$  and  $v$  belong to the same community, then  $\mathbf{E}[h(u, v)] = \sum_{i=1}^{\ell} \mathbf{P}(s_i(u) \neq s_i(v)) \leq \varepsilon\ell$ . Similarly, if they belong to different communities, then  $\mathbf{E}[h(u, v)] = \sum_{i=1}^{\ell} \mathbf{P}(s_i(u) \neq s_i(v)) \geq (1 - \delta)\ell$ . If  $u$  and  $v$  belong to the same community, we compute  $\mathbf{P}(h(u, v) > \alpha\ell)$  and by Markov inequality we get that

$$\mathbf{P}(h(u, v) \geq \alpha\ell) \leq \frac{\mathbf{E}[h(u, v)]}{\alpha\ell} \leq \frac{\varepsilon}{\alpha} = \mathcal{O}\left(\frac{1}{n^c}\right), \quad (2)$$

where in the last inequality we use the hypothesis  $\varepsilon = \mathcal{O}(\frac{1}{n^\gamma})$  and  $\alpha = \Omega(\frac{1}{n^{\gamma-c}})$ . On the other hand, if  $u$  and  $v$  belong to different communities, we apply Theorem A.1 to  $h(u, v)$  by using the lower bound on the expected value of  $h(u, v)$  and the hypothesis  $\ell = \Theta(\log n)$ . Thus,

$$\mathbf{P}(h(u, v) < b(1 - \delta)\ell) \leq \exp\left(-\frac{(1 - b)^2}{2}(1 - \delta)\ell\right) = \mathcal{O}\left(\frac{1}{n^d}\right), \quad (3)$$

where  $d$  is a positive constant. The thesis follows by combining Eqs. (2) and (3).  $\square$

### 2.3 Volume-regular graphs

Recall that, for an undirected edge-weighted graph  $G = (V, E, w)$ , we denote by  $\delta(u)$  the volume of a node  $u \in V$ , i.e.,  $\delta(u) = \sum_{v: (u, v) \in E} w(u, v)$ . Note that the transition matrix  $P$  of a random walk on  $G$  is such that  $P_{uv} = w(u, v) / \delta(u)$ . Given a partition  $\mathcal{V} = \{V_1, \dots, V_k\}$  of the set of nodes  $V$ , for a node  $u \in V$  and a partition index  $i \in [k]$ ,  $\delta_i(u)$  denotes the overall weight of edges connecting  $u$  to nodes in  $V_i$ ,  $\delta_i(u) = \sum_{v \in V_i: (u, v) \in E} w(u, v)$ . Hence,  $\delta(u) = \sum_{i=1}^k \delta_i(u)$ .

**Definition 2.3** (Volume-regular graph). *Let  $G = (V, E, w)$  be an undirected edge-weighted graph with  $|V| = n$  nodes and let  $\mathcal{V} = \{V_1, \dots, V_k\}$  be a  $k$ -partition of the nodes, for some  $k \in [n]$ . We say that  $G$  is volume regular with respect to  $\mathcal{V}$  if, for every pair of partition indexes  $i, j \in [k]$  and for every pair of nodes  $u, v \in V_i$ ,  $\frac{\delta_j(u)}{\delta(u)} = \frac{\delta_j(v)}{\delta(v)}$ . We say that  $G$  is  $k$ -volume regular if there exists a  $k$ -partition  $\mathcal{V}$  of the nodes such that  $G$  is volume regular with respect to  $\mathcal{V}$ .*

In other words,  $G$  is volume regular if there exists a partition of the nodes such that the fraction of a node's volume toward a set of the partition is constant across nodes of the same set. Note that all graphs with  $n$  nodes are trivially 1- and  $n$ -volume regular.

Let  $G = (V, E, w)$  be a  $k$ -volume regular graph and let  $P$  be the transition matrix of a random walk on  $G$ . In the next lemma we prove that the span of  $k$  linearly independent eigenvectors of  $P$  equals the span of the indicator vectors of the  $k$  communities of  $G$ . The proof makes use of the correspondence between random walks on volume regular graphs and ordinary lumpable Markov chains [KS60]; in particular the result follows from Lemma 3.2 and Lemma 3.3, that we prove in Section 3.

**Lemma 2.4.** *Let  $P$  be the transition matrix of a random walk on a  $k$ -volume regular graph  $G = (V, E, w)$  with  $k$ -partition  $\mathcal{V} = \{V_1, \dots, V_k\}$ . There exists a family  $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$  of linearly independent eigenvectors of  $P$  such that  $\text{Span}(\{\mathbf{v}_1, \dots, \mathbf{v}_k\}) = \text{Span}(\{\mathbf{1}_{V_1}, \dots, \mathbf{1}_{V_k}\})$ , with  $\mathbf{1}_{V_i}$  the indicator vector of the  $i$ -th set of the partition, for  $i \in [k]$ .*

In the rest of the paper we call “stepwise” the eigenvectors of  $P$  that can be written as linear combinations of the indicator vectors of the communities. In the next definition, we formalize the fact that a  $k$ -volume regular graph is *clustered* if the  $k$  linearly independent stepwise eigenvectors of  $P$ , whose existence is guaranteed by the above lemma, are associated to the  $k$  largest eigenvalues of  $P$ .

**Definition 2.5** (Clustered volume regular graph). *Let  $G = (V, E, w)$  be a  $k$ -volume regular graph and let  $P$  be the transition matrix of a random walk on  $G$ . We say that  $G$  is a clustered  $k$ -volume regular graph if the  $k$  stepwise eigenvectors of  $P$  are associated to the first  $k$  largest eigenvalues of  $P$ .*

### 3 Volume-regular graphs and lumpable Markov chains

The class of volume-regular graphs is deeply connected with the definition of *lumpability* [KS60] of Markov chains. We here first recall the definition of lumpable Markov chain and then show that a graph  $G$  is volume-regular if and only if the associated weighted random walk is a lumpable Markov chain.

**Definition 3.1** (Ordinary lumpability of Markov Chains). *Let  $\{X_t\}_t$  be a finite Markov chain with state space  $V$  and transition matrix  $P = (P_{uv})_{u,v \in V}$  and let  $\mathcal{V} = \{V_1, \dots, V_k\}$  be a partition of the state space. Markov chain  $\{X_t\}_t$  is ordinary lumpable with respect to  $\mathcal{V}$  if, for every pair of partition indexes  $i, j \in [k]$  and for every pair of nodes in the same set of the partition  $u, v \in V_i$ , it holds that*

$$\sum_{w \in V_i} P_{uw} = \sum_{w \in V_i} P_{vw}, \quad \forall u, v \in V_j. \quad (4)$$

We define the lumped matrix  $\hat{P}$  of the Markov Chain as the matrix such that  $\hat{P}_{ij} = \sum_{w \in V_i} P_{uw}$ , for any  $u \in V_j$ .

We first prove that random walks on Volume-regular graphs define exactly the subset of reversible and ordinary lumpable Markov chains.

**Lemma 3.2.** *A reversible Markov chain  $\{X_t\}_t$  is ordinary lumpable if and only if it is a random walk on a volume-regular graph.*

*Proof.* Assume first that  $\{X_t\}_t$  is ordinary lumpable and let  $P$  be the corresponding transition matrix. Consider the weighted graph  $G = (V, E, w)$  obtained from  $P$  as follows:  $V$  corresponds to the set of states in  $P$ , while  $w(u, v) = \pi(u)P_{uv}$ , for every  $u, v \in V$ , with  $\pi$  the stationary distribution of  $P$ . Note that  $G$  is an undirected graph, i.e.,  $w(u, v) = \pi(u)P_{uv} \stackrel{(a)}{=} \pi(v)P_{vu} = w(v, u)$ , where (a) holds because  $P$  is reversible. Moreover

$$\delta(u) = \sum_{z \in V} w(u, z) = \sum_{z \in V} \pi(u)P_{uz} = \pi(u) \sum_{z \in V} P_{uz} \stackrel{(a)}{=} \pi(u),$$

where (a) holds because  $P$  is stochastic. Thus  $G$  meets Definition 2.3 because, for any  $u, v \in V_i$ ,

$$\frac{\delta_j(u)}{\delta(u)} = \frac{1}{\pi(u)} \sum_{z \in V_j} w(u, z) = \sum_{z \in V_j} P_{uz} = \sum_{z \in V_j} P_{vz} = \frac{1}{\pi(v)} \sum_{z \in V_j} w(v, z) = \frac{\delta_j(v)}{\delta(v)}.$$

Next, assume  $G$  is  $k$ -volume-regular with respect to the partition  $\mathcal{V} = \{V_1, \dots, V_k\}$ . Let  $P$  be the transition matrix of the corresponding random walk. For every  $i, j \in [k]$  and for every

$u, v \in V_i$  we have:

$$\sum_{z \in V_j} P_{uz} = \sum_{z \in V_j} \frac{w(u, z)}{\delta(u)} = \frac{\delta_j(u)}{\delta(u)} \stackrel{(a)}{=} \frac{\delta_j(v)}{\delta(v)} = \sum_{z \in V_j} \frac{w(v, z)}{\delta(v)} = \sum_{z \in V_j} P_{vz},$$

where (a) follows from Definition 2.3. Moreover note that  $P$  is reversible with respect to distribution  $\pi$ , where  $\pi(u) = \frac{\delta(u)}{\text{vol}(G)}$ .  $\square$

Note that infinitely many  $k$ -volume-regular graphs have the same  $k$ -ordinary lumpable random walk chain.

We next show that a Markov chain is  $k$ -ordinary lumpable if and only if the corresponding transition matrix  $P$  has  $k$  stepwise, linearly independent eigenvectors.

**Lemma 3.3.** *Let  $P$  be the transition matrix of a Markov chain. Then  $P$  has  $k$  stepwise linearly independent eigenvectors if and only if  $P$  is ordinary lumpable.*

*Proof.* We divide the proof in two parts. First, we assume that  $P$  is ordinary lumpable and show that  $P$  has  $k$  stepwise linearly independent eigenvectors. Second, we assume that  $P$  has  $k$  stepwise linearly independent eigenvectors and show that  $P$  is ordinary lumpable.

1. Let  $P$  be ordinary lumpable and  $\widehat{P}$  its lumped matrix. Let  $\lambda_i, \mathbf{v}_i$  be the eigenvalues and eigenvectors of  $\widehat{P}$ , for each  $i \in [k]$ . Let  $\mathbf{w}_i \in \mathbb{R}^n$  be a stepwise vector defined as

$$\mathbf{w}_i = \underbrace{(\mathbf{v}_i(1), \dots, \mathbf{v}_i(1))}_{|V_1| \text{ times}}, \underbrace{(\mathbf{v}_i(2), \dots, \mathbf{v}_i(2))}_{|V_2| \text{ times}}, \dots, \underbrace{(\mathbf{v}_i(k), \dots, \mathbf{v}_i(k))}_{|V_k| \text{ times}}^\top,$$

where  $\mathbf{v}_i(j)$  indicates the  $j$ -th component of  $\mathbf{v}_i$ , and then the  $|V_j|$  components relative to  $V_j$  are all equal to  $\mathbf{v}_i(j)$ .

Since the eigenvectors  $\mathbf{v}_i$  of  $\widehat{P}$  are linearly independent, the vectors  $\mathbf{w}_i$  are also linearly independent. Moreover, it is easy to see that  $P\mathbf{w}_i = \lambda_i\mathbf{w}_i$  by just verifying the equation for every  $i \in [k]$ .

2. Assume  $P$  has  $k$  stepwise linearly independent eigenvectors  $\mathbf{w}_i$ , associated to  $k$  eigenvalues  $\lambda_i$ , for each  $i \in [k]$ . Let  $\mathbf{v}_i \in \mathbb{R}^k$  the vector that has as components the  $k$  constant values in the steps of  $\mathbf{w}_i$ . Since the  $\mathbf{w}_i$  are linearly independent, the  $\mathbf{v}_i$  also are.

For every eigenvector  $\mathbf{w}_i$  and for every two states  $x, y \in V_l$ , for every  $l \in [k]$ , we have that  $\lambda_i\mathbf{w}_i(x) = \lambda_i\mathbf{w}_i(y)$  since  $\mathbf{w}_i$  is stepwise. Then, since  $P\mathbf{w}_i = \lambda_i\mathbf{w}_i$ , we have that

$$\sum_{j=1}^k \sum_{z \in V_j} P_{xz} \mathbf{v}_i(j) = (P\mathbf{w}_i)(x) = (P\mathbf{w}_i)(y) = \sum_{j=1}^k \sum_{z \in V_j} P_{yz} \mathbf{v}_i(j).$$

Thus  $\sum_{j=1}^k \mathbf{v}_i(j) \sum_{z \in V_j} (P_{xz} - P_{yz}) = 0$  and then it follows that

$$\sum_{j=1}^k \mathbf{v}_i(j) \mathbf{u}_{xy}(j) = \langle \mathbf{u}_{xy}, \mathbf{v}_i \rangle = 0,$$

where  $\mathbf{u}_{xy}(j) := \sum_{z \in V_j} (P_{xz} - P_{yz})$ . Since the  $\mathbf{v}_i$  are  $k$  linearly independent vectors in a  $k$ -dimensional space,  $\mathbf{u}_{xy}$  cannot be orthogonal to all of them and then it has to be the null vector, i.e.,  $\mathbf{u}_{xy}(j) = 0$  for all  $j \in [k]$ . This implies that  $P$  is ordinary lumpable, i.e.,  $\sum_{z \in V_j} P_{xz} = \sum_{z \in V_j} P_{yz}$ . It is easy to verify that the eigenvalues and eigenvectors of  $\widehat{P}$  are exactly  $\lambda_i, \mathbf{v}_i$ , with  $i \in [k]$ .  $\square$

## 4 Averaging dynamics on clustered volume regular graphs

Let  $n_{\min} := \min_{i \in [k]} |V_i|$  and  $n_{\max} := \max_{i \in [k]} |V_i|$  be the maximum and minimum sizes of the communities of a volume-regular graph  $G = (V, E, w)$  with  $n$  nodes and  $k$ -partition  $\mathcal{V} = \{V_1, \dots, V_k\}$ . Recall also that  $\Delta$  is the maximum weighted degree of the nodes of  $G$  and  $\lambda_1, \dots, \lambda_n$  are the eigenvalues of the transition matrix of a random walk on  $G$  (see Section 2). In this section we prove the following result.

**Theorem 4.1.** *Let  $G = (V, E, w)$  be a connected clustered  $k$ -volume-regular graph with  $n$  nodes and  $k$ -partition  $\mathcal{V} = \{V_1, \dots, V_k\}$ , such that  $\Delta \leq \frac{\sqrt{n_{\min}}}{25}$  and  $2\Delta(n_{\max}/n_{\min}) < k \leq \sqrt{n}$ . Assume further that  $1 - \lambda_2 \leq \frac{\lambda_k \log(\lambda_k/\lambda_{k+1})}{7 \log(2\Delta n)}$  and  $\lambda_k \geq \frac{7\lambda_2 - 5}{2}$ . A non-empty time interval  $[T_1, T_2]$  exists, with  $T_1 = \mathcal{O}\left(\frac{\log n}{\log(\lambda_k/\lambda_{k+1})}\right)$  and  $T_2 = \Omega\left(\frac{\lambda_k}{1 - \lambda_2}\right)$ , such that for each  $t \in [T_1, T_2]$ , the AVERAGING dynamics truncated at round  $t$  is a  $(\mathcal{O}(n^{-1/2}), 1 - \Omega(1))$ -community-sensitive algorithm.*

**Remark 1** (The extent of the time-window). *Notice that the time window cannot be too long: by Cheeger's inequality  $1 - \lambda_2 \geq \frac{h_G^2}{2} \geq 1/(2\Delta^2 n^2)$ ,<sup>5</sup> thus  $T_2 = \mathcal{O}(\Delta^2 n^2)$ .*

**Remark 2** (The extent of non-regularity). *Notice that the condition  $k > 2\Delta(n_{\max}/n_{\min})$  implies*

$$\max_{i \in [k]} \{\text{vol}(V_i)\} \leq \Delta n_{\max} < \frac{k}{2} n_{\min} \leq \frac{k}{2} \min_{i \in [k]} \{\text{vol}(V_i)\}.$$

*In other words, the AVERAGING dynamics gives a good community-sensitive labeling when the communities are not too unbalanced in terms of their volumes. Moreover, the smaller the number of communities the more the volume-balance requirement is tight.*

In the remainder of this section, we first introduce further notation and then state the main technical lemmas (Lemmas 4.2 to 4.4), that will be used in the proof of Theorem 4.1, which concludes this section.

Let  $G = (V, E, w)$  be a clustered  $k$ -volume regular graph and, without loss of generality, let  $V_1, \dots, V_k$  be an arbitrary ordering of its communities. We introduce a family of stepwise vectors that generalize Fiedler vector [Fie89], namely

$$\left\{ \chi_i = \sqrt{\frac{\hat{m}_i}{m_i}} \mathbf{1}_{V_i} - \sqrt{\frac{m_i}{\hat{m}_i}} \mathbf{1}_{\hat{V}_i} : i \in [k-1] \right\}, \quad (5)$$

where  $\mathbf{1}_{V_i}$  is the indicator vector of the set  $V_i$  and, for convenience sake, we denoted by  $m_i$  the volume of the  $i$ -th community,  $\hat{V}_i$  the set of all nodes in communities  $i+1, \dots, k$ , and  $\hat{m}_i$  the volume of  $\hat{V}_i$ , i.e.,  $m_i := \sum_{u \in V_i} \delta(u)$ ,  $\hat{V}_i := \bigcup_{h=i+1}^k V_h$ , and  $\hat{m}_i := \sum_{h=i+1}^k m_h$ . Note that vectors  $\chi_i$ s are ‘‘stepwise’’ with respect to the communities of  $G$  (i.e., for every  $i \in [k-1]$ ,  $\chi_i(u) = \chi_i(v)$  whenever  $u$  and  $v$  belong to the same community).

Recall from Eq. (1) that the initial state vector can be written as  $\mathbf{x} = \sum_{i=1}^n \alpha_i \mathbf{v}_i$ . Let  $\mathbf{z} := \sum_{i=1}^k \alpha_i \mathbf{v}_i$  and note that  $\mathbf{z} = \alpha_1 \mathbf{1} + \sum_{i=1}^{k-1} \gamma_i \chi_i$  by applying Lemma 2.4 and because  $\text{Span}(\{\mathbf{1}, \chi_1, \dots, \chi_{k-1}\}) = \text{Span}(\{\mathbf{1}_{V_1}, \dots, \mathbf{1}_{V_k}\})$ . Let us now define the vector  $\mathbf{y} := \mathbf{z} - \alpha_1 \mathbf{1}$  or, equivalently,

$$\mathbf{y} = \sum_{i=1}^{k-1} \gamma_i \chi_i, \quad \text{where } \gamma_i = \frac{\mathbf{x}^\top D \chi_i}{\|D^{1/2} \chi_i\|^2}. \quad (6)$$

<sup>5</sup>This can be seen by observing that: *i)* the minimum volume of a cut must be at least half the minimum degree of the graph, which we normalize to 1, and *ii)* in computing  $h_G$ , we restrict to subsets of volume at most  $\text{vol}(G)$ , which is at most  $\Delta n$ .

Note that the coefficients  $\gamma_{is}$  are proportional to the length of the projection of the (inhomogeneously) contracted state vector on the (inhomogeneously) contracted  $D^{\frac{1}{2}}\chi_{is}$ ; the previous expression is valid since the vectors in  $\{D^{\frac{1}{2}}\mathbf{1}\} \cup \{D^{\frac{1}{2}}\chi_i : i \in [k-1]\}$  are mutually orthogonal.<sup>6</sup>

In Lemma 4.2 we show that every component of  $\mathbf{y}$ , i.e., the projection of the (inhomogeneously) contracted initial state vector  $D^{\frac{1}{2}}\mathbf{x}$  on the (inhomogeneously) contracted vectors  $D^{\frac{1}{2}}\chi_{is}$ , is not too small, w.h.p.

**Lemma 4.2** (Length of the projection of the state vector). *Let  $G = (V, E, w)$  be a connected clustered  $k$ -volume-regular graph with  $n$  nodes and  $k$ -partitions  $\mathcal{V} = \{V_1, \dots, V_k\}$ . Under the hypotheses of Theorem 4.1, for every  $u \in V$ ,*

$$\mathbf{P} \left( |\mathbf{y}(u)| \geq \frac{1}{\Delta n} \right) \geq 1 - \mathcal{O} \left( \frac{1}{\sqrt{n}} \right).$$

*Proof.* Without loss of generality, we assume  $u \in V_1$ , which possibly just amounts to a relabeling of the nodes. With this assumption, we have

$$\mathbf{y}(u) = \gamma_1 \chi_1(u) = \gamma_1 \sqrt{\frac{\hat{m}_1}{m_1}},$$

where the second equality follows from the definitions of the  $\chi_i$ 's (Eq. (5)) and the fact that  $u \in V_1$ . Next, observe that we have:

$$\|D^{1/2}\chi_1\|^2 = \frac{\hat{m}_1}{m_1} \sum_{v \in V_1} \delta(v) + \frac{m_1}{\hat{m}_1} \sum_{v \in \hat{V}_1} \delta(v) = \hat{m}_1 + m_1 = m,$$

where  $m := \text{vol}(V)$ . We now bound

$$|\gamma_1| = \frac{|\mathbf{x}^\top D\chi_1|}{\|D^{1/2}\chi_1\|^2} = \frac{|\mathbf{x}^\top D\chi_1|}{m}.$$

More precisely, we prove that it is at least  $1/m$  with probability  $1 - \mathcal{O} \left( \frac{1}{\sqrt{n}} \right)$ , where probability is computed over the randomness of  $\mathbf{x}$ .

Assume for the moment that  $\hat{m}_1 \geq m_1$ . From the definition of  $\chi_1$  we have:

$$\mathbf{x}^\top D\chi_1 = \mathbf{x}^\top D \left( \sqrt{\frac{\hat{m}_1}{m_1}} \mathbf{1}_{V_1} - \sqrt{\frac{m_1}{\hat{m}_1}} \mathbf{1}_{\hat{V}_1} \right) = \sqrt{\frac{m_1}{\hat{m}_1}} \mathbf{x}^\top D \left( \frac{\hat{m}_1}{m_1} \mathbf{1}_{V_1} - \mathbf{1}_{\hat{V}_1} \right).$$

Now, set  $\mathbf{w} = D \left( \frac{\hat{m}_1}{m_1} \mathbf{1}_{V_1} - \mathbf{1}_{\hat{V}_1} \right)$  and note that  $|\mathbf{w}(u)| \geq 1$  from the hypothesis that  $\hat{m}_1 \geq m_1$  and since  $\delta(v) \geq 1$ , for every  $v \in V$ . We can thus apply Theorem A.5 to  $\mathbf{w}$  with  $r = 0$ , so that we can write:

$$\mathbf{P} \left( |\mathbf{x}^\top D\chi_1| < \sqrt{\frac{m_1}{\hat{m}_1}} \right) = \mathbf{P} (|\mathbf{x}^\top \mathbf{w}| < 1) \leq \mathcal{O} \left( \frac{1}{\sqrt{n}} \right),$$

where the equality follows since  $\mathbf{x}^\top D\chi_1 = \sqrt{\frac{m_1}{\hat{m}_1}} \mathbf{x}^\top \mathbf{w}$ . Hence, with probability  $1 - \mathcal{O} \left( \frac{1}{\sqrt{n}} \right)$  we have  $|\gamma_1| \geq \sqrt{\frac{m_1}{\hat{m}_1}} \cdot \frac{1}{m}$  and thus, with the same probability:

$$|\mathbf{y}(u)| = |\gamma_1| \sqrt{\frac{\hat{m}_1}{m_1}} \geq \frac{1}{m} \geq \frac{1}{\Delta n}.$$

---

<sup>6</sup>The mutual orthogonality of the vectors, including  $D^{\frac{1}{2}}\mathbf{1}$ , is also one of the reasons why other ‘‘simpler’’ families of stepwise vectors, e.g., the indicator vectors of the communities, are not used instead.

Assume now that  $m_1 > \hat{m}_1$ . This time we write:

$$\mathbf{x}^\top D \boldsymbol{\chi}_1 = \sqrt{\frac{\hat{m}_1}{m_1}} \mathbf{x}^\top D \left( \mathbf{1}_{V_1} - \frac{m_1}{\hat{m}_1} \mathbf{1}_{\hat{V}_1} \right)$$

and we set  $\mathbf{w} = D \left( \mathbf{1}_{V_1} - \frac{m_1}{\hat{m}_1} \mathbf{1}_{\hat{V}_1} \right)$ . Note that, again,  $|\mathbf{w}(v)| \geq 1$  for every  $v \in V$ . Proceeding as in the previous case we obtain  $|\gamma_1| \geq \sqrt{\frac{\hat{m}_1}{m_1}} \cdot \frac{1}{m}$  with probability  $1 - \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$  and thus, with the same probability:

$$|\mathbf{y}(u)| = |\gamma_1| \sqrt{\frac{\hat{m}_1}{m_1}} \geq \frac{\hat{m}_1}{mm_1} = \frac{m - m_1}{mm_1} \stackrel{(a)}{\geq} \frac{1}{m} \geq \frac{1}{\Delta n},$$

where in (a) we used that  $m_i < \frac{m}{2}$  (see Remark 2). This concludes the proof.  $\square$

In Lemma 4.3 we show that given any ‘‘pair of steps’’ of the vector  $\mathbf{y}$  (defined in Eq. (6)), the two steps have different signs, with constant probability.

**Lemma 4.3** (Different communities, different signs). *Let  $G = (V, E, w)$  be a clustered  $k$ -volume regular graph with maximum weighted degree  $\Delta \leq \frac{\sqrt{n_{\min}}}{25}$  and with  $k > 2\Delta(n_{\max}/n_{\min})$ . For each pair of nodes  $u \in V_i$  and  $v \in V_j$ , with  $i \neq j$ , it holds that*

$$\mathbf{P}(\text{sgn}(\mathbf{y}(u)) \neq \text{sgn}(\mathbf{y}(v))) = \Omega(1).$$

*Proof.* Since the ordering of the communities (and consequent definition of the  $\boldsymbol{\chi}_i$ ’s, given in Eq. (5)) is completely arbitrary, we can assume  $i = 1$  and  $j = 2$ , without loss of generality. Let us define  $X(V_i) := \sum_{w \in V_i} \delta(w) \mathbf{x}(w)$ , where  $\mathbf{x} = \mathbf{x}^{(0)}$  is the initial state vector.

Note that  $\mathbf{y}(u) = \gamma_1 \boldsymbol{\chi}_1(u)$  and  $\mathbf{y}(v) = \gamma_1 \boldsymbol{\chi}_1(v) + \gamma_2 \boldsymbol{\chi}_2(v)$ , since the other terms of the  $\boldsymbol{\chi}_i$ s are equal to 0 on the components relative to  $u$  and  $v$ . Thus, with some algebra, we get

$$\begin{aligned} \mathbf{y}(u) &= \frac{1}{m} \left[ \frac{\hat{m}_1}{m_1} X(V_1) - X(V_2) - X(\hat{V}_2) \right], \\ \mathbf{y}(v) &= \frac{1}{m} \left[ \frac{m_1 m_2 + m \hat{m}_2}{\hat{m}_1 m_2} X(V_2) - X(V_1) - X(\hat{V}_2) \right], \end{aligned}$$

where  $\hat{V}_i := \bigcup_{h=i+1}^k V_h$ . Note that, by linearity of expectation,  $\mathbf{E}[X(V_i)] = 0$ . Moreover, since the terms  $\mathbf{x}(w)$ s are independent Rademacher random variables, we can write the standard deviation of  $X(V_i)$  as

$$\sigma(X(V_i)) = \sqrt{\sum_{w \in V_i} \sigma^2(\delta(w) \mathbf{x}(w))} = \sqrt{\sum_{w \in V_i} \left( \mathbf{E}[\delta(w)^2 \mathbf{x}(w)^2] - \mathbf{E}[\delta(w) \mathbf{x}(w)]^2 \right)} = \sqrt{\sum_{w \in V_i} \delta(w)^2}.$$

Then we can upper and lower bound the standard deviation  $\sigma(X(V_i))$  getting  $\frac{m_i}{\sqrt{n_i}} \leq \sigma(X(V_i)) \leq \Delta \sqrt{n_i}$ , where the lower bound follows from  $\|\mathbf{d}\|_2 \geq \|\mathbf{d}\|_1 / \sqrt{n_i}$ , where  $\mathbf{d}_i$  is the vector of weighted degrees of nodes in community  $V_i$ , and for the upper bound we used that  $\delta(w) \leq \Delta$ , for each  $w \in V$ .

Let us now define the following three events:

1.  $E_1: X(V_1) > \sigma(X(V_1)) \implies X(V_1) > \frac{m_1}{\sqrt{n_1}} \geq \frac{\min_i m_i}{\sqrt{n_{\max}}}$ ;
2.  $E_2: X(V_2) < -\sigma(X(V_2)) \implies X(V_2) < -\frac{m_2}{\sqrt{n_2}} \leq -\frac{\min_i m_i}{\sqrt{n_{\max}}}$ ;

$$3. E_3: 0 \leq X(\hat{V}_2) < (2/\sqrt{k}) \sigma(X(\hat{V}_2)) \implies 0 \leq X(\hat{V}_2) < 2\Delta \sqrt{(1/k) \sum_{i=3}^k n_i} \leq 2\Delta \sqrt{n_{\max}},$$

When  $E_1, E_2, E_3$  are true it follows directly that  $\mathbf{y}(v) < 0$ . As for  $\mathbf{y}(u) > 0$  we have

$$\begin{aligned} \frac{\hat{m}_1}{m_1} X(V_1) - X(V_2) - X(\hat{V}_2) &> \frac{\hat{m}_1}{m_1} \sigma(X(V_1)) + \sigma(X(V_2)) - (2/\sqrt{k}) \cdot \sigma(X(\hat{V}_2)) \\ &\geq \frac{kn_{\min}}{\sqrt{n_{\max}}} - 2\Delta \sqrt{n_{\max}} > 0, \end{aligned}$$

since, for the last inequality,  $k > 2\Delta(n_{\max}/n_{\min})$  by hypothesis.

Note that all three events  $E_1, E_2, E_3$  have probability at least constant and, being the events independent, also  $\mathbf{P}(E_1 \cap E_2 \cap E_3)$  is constant. Indeed, it is possible to prove the constant lower bounds on the probabilities by approximating the random variables with Gaussian ones using Berry-Esseen's theorem (Theorem A.4). Note that  $X(V_1), X(V_2), X(\hat{V}_2)$  all are of the form  $Z = \sum_{w \in T} Z_w$ , for some  $T \subseteq V$  and where  $Z_w = \delta(w)\mathbf{x}(w)$ . Recall that  $\mathbf{E}[Z_w] = 0$  and that  $\sigma^2(Z_w) = \delta(w)^2$ . Moreover, note that the third absolute moment of  $Z_w$  is  $\mathbf{E}[|Z_w|^3] = \delta(w)^3 \mathbf{E}[|\mathbf{x}(w)|^3] = \delta(w)^3$ . Therefore we can apply Theorem A.4 which claims that there exists a positive constant  $C \leq 1.88$  [Ber41] such that, for every  $z \in \mathbb{R}$ ,

$$|\mathbf{P}(Z \leq z \cdot \sigma(Z)) - \Phi(z)| \leq \frac{C}{\sigma(Z)} \max_{w \in T} \frac{\delta(w)^3}{\delta(w)^2} \leq \frac{C \cdot \Delta}{\sigma(Z)},$$

where  $\Phi$  is the cumulative distribution function of the standardized normal distribution. Thus

$$\mathbf{P}(Z > z \cdot \sigma(Z)) \geq 1 - \Phi(z) - \frac{C \cdot \Delta}{\sigma(Z)}. \quad (7)$$

Since  $\Delta \leq \frac{\sqrt{n_{\min}}}{25}$  by hypothesis and  $\sigma(Z) \geq \sqrt{|T|}$  for every  $T \subseteq V$ , taking  $z = 1$  it follows from Eq. (7) that

$$\mathbf{P}(E_1) = \mathbf{P}(X(V_1) > \sigma(X(V_1))) \geq 1 - \Phi(1) - \frac{C \cdot \Delta}{\sqrt{n_{\min}}} \geq 1 - \Phi(1) - \frac{C}{25} \geq \frac{1 - \Phi(1)}{2} \approx 0.08,$$

since  $\frac{1}{25} < \frac{1 - \Phi(1)}{2C} \approx 0.042$ . Since the distribution of  $Z$  is symmetric for every  $T \subseteq V$ , it holds that  $\mathbf{P}(E_2) \geq 1 - \Phi(1) - \frac{C}{25} \approx 0.08$ . Similarly, it also holds that

$$\mathbf{P}(E_3) = \frac{1}{2} - \mathbf{P}\left(X(\hat{V}_2) > (2/\sqrt{k}) \sigma(X(\hat{V}_2))\right) \geq \Phi(2/\sqrt{k}) + \frac{C}{25} - \frac{1}{2} > \frac{C}{25} \approx 0.075. \quad \square$$

Recall that the binary labeling of each node only depends on the difference of its state in two consecutive rounds (see Algorithm 1). In Lemma 4.4 we show that, under suitable assumptions on the transition matrix of a random walk on  $G$ , a large enough time window exists where, for each node  $u$ , the sign of the difference  $\mathbf{x}^{(t)}(u) - \mathbf{x}^{(t+1)}(u)$  of the state vector across two consecutive rounds equals the sign of  $\mathbf{y}(u)$ , w.h.p. Since  $\mathbf{y}$  (defined in Eq. (6)) is a stepwise vector, this implies that two nodes in the same community have the same label, w.h.p. For the sake of readability, in the proof of Lemma 4.4 we use two technical lemmas as black boxes, postponing their proofs to Subsection 4.1.

**Lemma 4.4** (Sign of the difference). *Let  $G = (V, E, w)$  be a clustered  $k$ -volume regular graph with maximum weighted degree  $\Delta \leq \frac{\sqrt{n_{\min}}}{25}$ . If  $\lambda_k \geq \frac{7\lambda_2 - 5}{2}$ ,  $1 - \lambda_2 \leq \frac{\lambda_k \log(\lambda_k/\lambda_{k+1})}{7 \log(2\Delta n)}$ ,  $|\mathbf{y}(u)| \geq \frac{1}{\Delta n}$  for every  $u \in V$ , then a non-empty time interval  $[T_1, T_2]$  exists, with  $T_1 = \mathcal{O}\left(\frac{\log n}{\log(\lambda_k/\lambda_{k+1})}\right)$  and  $T_2 = \Omega\left(\frac{\lambda_k}{1 - \lambda_2}\right)$ , such that, for every  $u \in V$  and every  $t \in [T_1, T_2]$  of the AVERAGING dynamics,*

$$\text{sgn}(\mathbf{x}^{(t)}(u) - \mathbf{x}^{(t+1)}(u)) = \text{sgn}(\mathbf{y}(u)).$$

*Proof.* Recall from Eq. (1) that the state vector at time  $t$ , i.e.,  $\mathbf{x}^{(t)}$ , can be written as the sum of the first  $k$  stepwise vectors of  $P$  and of the remaining ones, namely

$$\mathbf{x}^{(t)} = \alpha_1 \mathbf{1} + \sum_{i=2}^k \lambda_i^t \alpha_i \mathbf{v}_i + \sum_{i=k+1}^n \lambda_i^t \alpha_i \mathbf{v}_i = \alpha_1 \mathbf{1} + \mathbf{c}^{(t)} + \mathbf{e}^{(t)}.$$

In what follows we call  $\mathbf{c}^{(t)} := \sum_{i=2}^k \lambda_i^t \alpha_i \mathbf{v}_i$  the *core contribution* and  $\mathbf{e}^{(t)} := \sum_{i=k+1}^n \lambda_i^t \alpha_i \mathbf{v}_i$  the *error contribution*. If we look at the difference of the state vector in two consecutive rounds, the first term cancels out being constant over time, so that

$$\mathbf{x}^{(t)}(u) - \mathbf{x}^{(t+1)}(u) = \mathbf{c}^{(t)}(u) - \mathbf{c}^{(t+1)}(u) + \mathbf{e}^{(t)}(u) - \mathbf{e}^{(t+1)}(u)$$

for each node  $u \in V$ . Note that the sign of the difference between two consecutive rounds is determined by the difference of the core contributions,  $\mathbf{c}^{(t)}(u) - \mathbf{c}^{(t+1)}(u)$ , whenever

$$\left| \mathbf{c}^{(t)}(u) - \mathbf{c}^{(t+1)}(u) \right| > \left| \mathbf{e}^{(t)}(u) - \mathbf{e}^{(t+1)}(u) \right|. \quad (8)$$

To identify conditions on  $t$  for which Eq. (8) holds, we give suitable bounds on both hand sides of the inequality. In more detail:

1. In Lemma 4.5 we prove that  $|\mathbf{e}^{(t)}(u)| \leq \lambda_{k+1}^t \sqrt{\Delta n}$ , for every  $u \in V$ , so that

$$\left| \mathbf{e}^{(t)}(u) - \mathbf{e}^{(t+1)}(u) \right| \leq \left| \mathbf{e}^{(t)}(u) \right| + \left| \mathbf{e}^{(t+1)}(u) \right| \leq 2\lambda_{k+1}^t \sqrt{\Delta n}.$$

2. In Lemma 4.6 we prove that  $|\mathbf{c}^{(t)}(u) - \mathbf{c}^{(t+1)}(u)| > \lambda_k^t (1 - \lambda_2) |\mathbf{y}(u)|$  for every  $u \in V$  and for every time  $t < T_2$ , where  $T_2 \geq \frac{\lambda_k}{2(1-\lambda_2)}$ ; note that the hypotheses on  $1 - \lambda_2$  imply  $T_2 = \Omega\left(\frac{\log n}{\log(\lambda_k/\lambda_{k+1})}\right)$ . Moreover, the assumptions of Lemma 4.6 are satisfied, since  $\mathbf{y}(u) \neq 0$  and  $\lambda_k \geq \frac{7\lambda_2 - 5}{2}$ .

Combining Lemma 4.5 and Lemma 4.6, we see that Eq. (8) holds whenever

$$\lambda_k^t (1 - \lambda_2) |\mathbf{y}(u)| > 2\lambda_{k+1}^t \sqrt{\Delta n}, \quad (9)$$

An easy calculation shows that this happens for all  $t > T_1$ , where

$$T_1 := \frac{\log\left(\frac{2\sqrt{\Delta n}}{(1-\lambda_2)|\mathbf{y}(u)|}\right)}{\log(\lambda_k/\lambda_{k+1})}.$$

Note that  $T_1 = \mathcal{O}\left(\frac{\log n}{\log(\lambda_k/\lambda_{k+1})}\right)$  and, e.g.,  $T_1 = \mathcal{O}(\log n)$  whenever  $\frac{\lambda_k}{\lambda_{k+1}} = 1 + \Omega(1)$ .

We next show that, under the assumptions of the lemma, the window  $[T_1, T_2]$  is not empty and, actually, it has a width that depends on the magnitude of  $\lambda_2$  and the ratio  $\lambda_k/\lambda_{k+1}$ . To this purpose, we first observe that Cheeger's inequality for weighted graphs (Theorem A.3) implies  $1 - \lambda_2 \geq \frac{h_G^2}{2} \geq \frac{1}{2(\Delta n)^2}$  (recall the footnote in Remark 1). Moreover, recalling that we are assuming  $|\mathbf{y}(u)| \geq \frac{1}{\Delta n}$ ,<sup>7</sup> we have:

$$T_1 = \frac{\log\left(\frac{2\sqrt{\Delta n}}{(1-\lambda_2)|\mathbf{y}(u)|}\right)}{\log(\lambda_k/\lambda_{k+1})} \stackrel{(a)}{\leq} \frac{\log\left(4\Delta^3 n^3 \sqrt{\Delta n}\right)}{\log(\lambda_k/\lambda_{k+1})} < \frac{7 \log(2\Delta n)}{2 \log(\lambda_k/\lambda_{k+1})} \stackrel{(b)}{\leq} \frac{\lambda_k}{2(1-\lambda_2)} \stackrel{(c)}{\leq} T_2, \quad (10)$$

<sup>7</sup>It may be worth recalling that our hypothesis on  $|\mathbf{y}(u)|$  holds with high probability from Lemma 4.2.

where: in (a) we used Cheeger’s inequality (Theorem A.3) in the way described above and our assumptions on  $|\mathbf{y}(u)|$ , in (b) we used the hypothesis on  $1 - \lambda_2$ , which implies  $\log(\lambda_k/\lambda_{k+1}) \geq \frac{7(1-\lambda_2)\log(2\Delta n)}{\lambda_k}$ , and in (c) the lower bound on  $T_2$  given by Lemma 4.6.

From Lemma 4.6 we also know that  $\text{sgn}(\mathbf{c}^{(t)}(u) - \mathbf{c}^{(t+1)}(u)) = \text{sgn}(\mathbf{y}(u))$  for every time  $t < T_2$ ; therefore we conclude that

$$\text{sgn}(\mathbf{x}^{(t)}(u) - \mathbf{x}^{(t+1)}(u)) = \text{sgn}(\mathbf{c}^{(t)}(u) - \mathbf{c}^{(t+1)}(u)) = \text{sgn}(\mathbf{y}(u)),$$

for every node  $u \in V$  and for every round  $t \in [T_1, T_2]$  of the AVERAGING dynamics.  $\square$

*Proof of Theorem 4.1.* The binary labeling of the nodes of  $G$  produced by the AVERAGING dynamics during the time window  $[T_1, T_2]$  is such that the two conditions required by the definition of  $(\varepsilon, \delta)$ -community-sensitive algorithm (Definition 2.1) are met, with  $\varepsilon = \mathcal{O}(n^{-1/2})$  and  $\delta = 1 - \Omega(1)$ . Indeed, the first condition follows directly from Lemma 4.4 together with the fact that  $\mathbf{y}$  is a “stepwise” vector, while Lemma 4.2 implies that  $\varepsilon = \mathcal{O}(n^{-\frac{1}{2}})$ , since  $\mathbf{y}(u)$  is not too small with probability at least  $1 - \mathcal{O}(n^{-1/2})$ . The second condition, instead, follows directly from the combination of Lemmas 4.3 and 4.4.  $\square$

**Remark 3** (Equal-sized communities). *If  $\lambda_k = \lambda_2$ , then an alternative version of Lemma 4.6 would tell us that, for every node  $u \in V$ ,  $\mathbf{c}^{(t)}(u) - \mathbf{c}^{(t+1)}(u) = \lambda_k^t(1 - \lambda_2)\mathbf{y}(u)$  and thus  $\text{sgn}(\mathbf{c}^{(t)}(u) - \mathbf{c}^{(t+1)}(u)) = \text{sgn}(\mathbf{y}(u))$ , in every round (with no need of  $T_2$ ); this would imply an infinite time window starting at the first round  $t > T_1$  (where the “error contribution” becomes small). In this sense our result also covers the case of multiple communities analyzed in [BCN<sup>+</sup>17], with  $k$  equal-sized communities in an unweighted graph and then  $\lambda_k = \lambda_2$ .*

**Remark 4** (Two communities). *Our result also generalizes that of [BCN<sup>+</sup>17] in the simpler case of two communities. In fact we don’t require the graph to be regular, but only volume-regular, thus taking into account communities that are potentially unbalanced. Indeed, for  $k = 2$ , the AVERAGING dynamics truncated at round  $t$  is a  $(\mathcal{O}(n^{-\frac{1}{2}}), \mathcal{O}(n^{-\frac{1}{2}}))$ -community-sensitive algorithm for every round  $t > T_1$ , with  $T_1 = \mathcal{O}\left(\frac{\log n}{\log(\lambda_2/\lambda_3)}\right)$ . Therefore, a single run of the dynamics highlights the community structure, i.e., the sign of the difference  $\mathbf{x}^{(t)} - \mathbf{x}^{(t)}$  is equal for nodes in the same communities and different for nodes in different communities, w.h.p.*

#### 4.1 Proofs for Lemma 4.4

In this section we prove the two lemmas used in the proof of Lemma 4.4: the upper bound on the “error contribution” and the lower bound on the “core contribution.”

**Lemma 4.5** (Upper bound on the error contribution). *Let  $\mathbf{e}^{(t)} := \sum_{i=k+1}^n \lambda_i^t \alpha_i \mathbf{v}_i$ . For every  $u \in V$ , it holds that*

$$|\mathbf{e}^{(t)}(u)| \leq \lambda_{k+1}^t \sqrt{\Delta n}.$$

*Proof.* To bound all components of vector  $\mathbf{e}^{(t)}$  we use its  $\ell^\infty$  norm, defined for any vector  $\mathbf{x}$  as  $\|\mathbf{x}\|_\infty := \sup_i |\mathbf{x}(i)|$ . In particular

$$\|\mathbf{e}^{(t)}\|_\infty^2 \leq \|\mathbf{e}^{(t)}\|^2 = \left\| \sum_{i=k+1}^n \lambda_i^t \alpha_i \mathbf{v}_i \right\|^2 = \left\| \sum_{i=k+1}^n \lambda_i^t \beta_i D^{-\frac{1}{2}} \mathbf{w}_i \right\|^2.$$

By using Cauchy-Schwarz inequality (Theorem A.2) and applying the definition of spectral norm of an operator, i.e.,  $\|A\| := \sup_{\mathbf{x}: \|\mathbf{x}\|=1} \|A\mathbf{x}\|$ , we get that

$$\left\| \sum_{i=k+1}^n \lambda_i^t \beta_i D^{-\frac{1}{2}} \mathbf{w}_i \right\|^2 \leq \left\| D^{-\frac{1}{2}} \right\|^2 \left\| \sum_{i=k+1}^n \lambda_i^t \beta_i \mathbf{w}_i \right\|^2 = \left\| D^{-\frac{1}{2}} \right\|^2 \sum_{i=k+1}^n \lambda_i^{2t} \beta_i^2,$$

since the  $\mathbf{w}_i$ s are orthonormal. With some additional simple bounds it follows that

$$\begin{aligned} \left\| D^{-\frac{1}{2}} \right\|^2 \sum_{i=k+1}^n \lambda_i^{2t} \beta_i^2 &\leq \left\| D^{-\frac{1}{2}} \right\|^2 \lambda_{k+1}^{2t} \sum_{i=k+1}^n \beta_i^2 \leq \left\| D^{-\frac{1}{2}} \right\|^2 \lambda_{k+1}^{2t} \sum_{i=1}^n \beta_i^2 \\ &= \left\| D^{-\frac{1}{2}} \right\|^2 \lambda_{k+1}^{2t} \left\| D^{\frac{1}{2}} \mathbf{x} \right\|^2 \leq \left\| D^{-\frac{1}{2}} \right\|^2 \lambda_{k+1}^{2t} \left\| D^{\frac{1}{2}} \right\|^2 \|\mathbf{x}\|^2. \end{aligned}$$

By using the fact that the spectral norm of a diagonal matrix is equal to its maximum value, we conclude that

$$\left\| D^{-\frac{1}{2}} \right\|^2 \lambda_{k+1}^{2t} \left\| D^{\frac{1}{2}} \right\|^2 \|\mathbf{x}\|^2 = \frac{\max_u \delta(u)}{\min_u \delta(u)} \lambda_{k+1}^{2t} \|\mathbf{x}\|^2 \leq \lambda_{k+1}^{2t} \Delta n.$$

Thus, for every  $u \in V$  it holds that  $|\mathbf{e}^{(t)}(u)| \leq \sqrt{\|\mathbf{e}^{(t)}\|_\infty^2} \leq \lambda_{k+1}^t \sqrt{\Delta n}$ .  $\square$

In Lemma 4.6 we show that the difference of the core contribution in consecutive rounds can be approximated, for our purposes in Lemma 4.4, with  $\mathbf{y}$ .

**Lemma 4.6** (Lower bound on the core contribution). *Let  $\mathbf{c}^{(t)} := \sum_{i=2}^k \lambda_i^t \alpha_i \mathbf{v}_i$  and let  $\mathbf{y}(u) \neq 0$  for every  $u \in V$ . If  $\lambda_k \geq \frac{7\lambda_2 - 5}{2}$ , then, for every  $u \in V$  and for every  $t \leq T_2$ , with  $T_2 \geq \frac{\lambda_k}{2(1-\lambda_2)}$ , the following holds:*

- $\text{sgn}(\mathbf{c}^{(t)}(u) - \mathbf{c}^{(t+1)}(u)) = \text{sgn}(\mathbf{y}(u))$ ;
- $|\mathbf{c}^{(t)}(u) - \mathbf{c}^{(t+1)}(u)| > \lambda_k^t (1 - \lambda_2) |\mathbf{y}(u)|$ .

*Proof.* Let us define  $d_{i,j} := \lambda_i - \lambda_j$ . Note that

$$\begin{aligned} \mathbf{c}^{(t)} - \mathbf{c}^{(t+1)} &= \sum_{i=2}^k \lambda_i^t (1 - \lambda_i) \alpha_i \mathbf{v}_i = \sum_{i=2}^k (\lambda_k + d_{i,k})^t (1 - \lambda_2 + d_{2,i}) \alpha_i \mathbf{v}_i \\ &= \sum_{i=2}^k [\lambda_k^t + (\lambda_k + d_{i,k})^t - \lambda_k^t] (1 - \lambda_2 + d_{2,i}) \alpha_i \mathbf{v}_i = \sum_{i=2}^k [\lambda_k^t (1 - \lambda_2) + c_i] \alpha_i \mathbf{v}_i \\ &= \lambda_k^t (1 - \lambda_2) \sum_{i=2}^k \alpha_i \mathbf{v}_i + \sum_{i=2}^k c_i \alpha_i \mathbf{v}_i = \lambda_k^t (1 - \lambda_2) \mathbf{y} + \sum_{i=2}^k c_i \alpha_i \mathbf{v}_i, \end{aligned}$$

where in the last equality we applied Lemma 2.4 to get  $\sum_{i=2}^k \alpha_i \mathbf{v}_i = \mathbf{y}$ , and where we defined  $c_i := \lambda_k^t d_{2,i} + [(\lambda_k + d_{i,k})^t - \lambda_k^t] (1 - \lambda_2 + d_{2,i})$ . Using the definition of  $d_{i,j}$ , we get

$$\begin{aligned} c_i &= \lambda_k^t d_{2,i} + [(\lambda_k + d_{i,k})^t - \lambda_k^t] (1 - \lambda_2 + d_{2,i}) \\ &= \lambda_k^t (\lambda_2 - \lambda_i) + [(\lambda_k + (\lambda_i - \lambda_k))^t - \lambda_k^t] (1 - \lambda_2 + (\lambda_2 - \lambda_i)) \\ &= \lambda_k^t (\lambda_2 - \lambda_i) + (\lambda_i^t - \lambda_k^t) (1 - \lambda_i) \\ &= \lambda_i^t (1 - \lambda_i) - \lambda_k^t (1 - \lambda_2). \end{aligned}$$

Note that  $\min_i[\lambda_i^t(1-\lambda_i)] - \lambda_k^t(1-\lambda_2) \leq c_i \leq \max_i[\lambda_i^t(1-\lambda_i)] - \lambda_k^t(1-\lambda_2)$ , for every  $i \in \{2, \dots, k\}$ . Since the minimum and the maximum are obtained for  $i = k$  and  $i = 2$  respectively, we have  $\lambda_k^t(\lambda_2 - \lambda_k) \leq c_i \leq (\lambda_2^t - \lambda_k^t)(1 - \lambda_2)$ . Let us call the positive and negative terms of  $\mathbf{y}(u)$  as

$$\mathbf{y}^+(u) := \sum_{i:\alpha_i \mathbf{v}_i(u) > 0} \alpha_i \mathbf{v}_i(u) \quad \text{and} \quad \mathbf{y}^-(u) := - \sum_{i:\alpha_i \mathbf{v}_i(u) < 0} \alpha_i \mathbf{v}_i(u).$$

Therefore, for each  $u \in V$ , it holds that

$$\mathbf{c}^{(t)}(u) - \mathbf{c}^{(t+1)}(u) \geq \lambda_k^t(1 - \lambda_2)\mathbf{y}(u) + \lambda_k^t(\lambda_2 - \lambda_k)\mathbf{y}^+(u) - (\lambda_2^t - \lambda_k^t)(1 - \lambda_2)\mathbf{y}^-(u), \quad (11)$$

$$\mathbf{c}^{(t)}(u) - \mathbf{c}^{(t+1)}(u) \leq \lambda_k^t(1 - \lambda_2)\mathbf{y}(u) + (\lambda_2^t - \lambda_k^t)(1 - \lambda_2)\mathbf{y}^+(u) - \lambda_k^t(\lambda_2 - \lambda_k)\mathbf{y}^-(u). \quad (12)$$

In the following we look for a time  $T_2$  such that, for every  $t \leq T_2$  it holds that

$$\left| \mathbf{c}^{(t)}(u) - \mathbf{c}^{(t+1)}(u) \right| > \lambda_k^t(1 - \lambda_2) |\mathbf{y}(u)|.$$

Note that  $\mathbf{y}(u) = \mathbf{y}^+(u) - \mathbf{y}^-(u)$ . We consider two cases:  $\mathbf{y}(u) > 0$  and  $\mathbf{y}(u) < 0$ .

Case  $\mathbf{y}(u) > 0$ : We look for a time  $T_2$  such that, for every time  $t \leq T_2$ , it holds that

$$\lambda_k^t(\lambda_2 - \lambda_k)\mathbf{y}^+(u) > (\lambda_2^t - \lambda_k^t)(1 - \lambda_2)\mathbf{y}^-(u). \quad (13)$$

Indeed, since  $\mathbf{y}(u) > 0$ , we can use  $\mathbf{y}^+(u) > \mathbf{y}^-(u)$  to upper bound the right hand side of the previous equation, so that Eq. (13) holds for every  $t$  that satisfies:

$$\lambda_k^t(\lambda_2 - \lambda_k) > (\lambda_2^t - \lambda_k^t)(1 - \lambda_2) \iff \left( \frac{\lambda_2}{\lambda_k} \right)^t < \frac{1 - \lambda_k}{1 - \lambda_2},$$

i.e., for every  $t \leq T_2$ , where

$$T_2 := \frac{\ln\left(\frac{1-\lambda_k}{1-\lambda_2}\right)}{\ln\left(\frac{\lambda_2}{\lambda_k}\right)} = \frac{\ln\left(1 + \frac{\lambda_2 - \lambda_k}{1 - \lambda_2}\right)}{\ln\left(1 + \frac{\lambda_2 - \lambda_k}{\lambda_k}\right)}.$$

Next, note that  $\frac{\lambda_2 - \lambda_k}{1 - \lambda_2} \leq \frac{5}{2}$  whenever  $\lambda_k \geq \frac{7\lambda_2 - 5}{2}$ . Hence, under the hypotheses of the lemma, we can use that  $1 + x \geq e^{\frac{x}{2}}$  for every  $x \in [0, \frac{5}{2}]$  and  $1 + x \leq e^x$  for every  $x$ . Thus:

$$T_2 = \frac{\ln\left(1 + \frac{\lambda_2 - \lambda_k}{1 - \lambda_2}\right)}{\ln\left(1 + \frac{\lambda_2 - \lambda_k}{\lambda_k}\right)} \geq \frac{\left(\frac{\lambda_2 - \lambda_k}{2(1 - \lambda_2)}\right)}{\left(\frac{\lambda_2 - \lambda_k}{\lambda_k}\right)} = \frac{\lambda_k}{2(1 - \lambda_2)}.$$

Plugging Eq. (13) into Eq. (11) we finally get

$$\mathbf{c}^{(t)}(u) - \mathbf{c}^{(t+1)}(u) > \lambda_k^t(1 - \lambda_2)\mathbf{y}(u) > 0. \quad (14)$$

Case  $\mathbf{y}(u) < 0$ : Proceeding along the same lines we obtain:

$$(\lambda_2^t - \lambda_k^t)(1 - \lambda_2)\mathbf{y}^+(u) < \lambda_k^t(\lambda_2 - \lambda_k)\mathbf{y}^-(u) \quad (15)$$

for every  $t \leq T_2$ . Therefore, by combining Eq. (15) and Eq. (12) we obtain

$$\mathbf{c}^{(t)}(u) - \mathbf{c}^{(t+1)}(u) < \lambda_k^t(1 - \lambda_2)\mathbf{y}(u) < 0. \quad (16)$$

Finally, by combining Eq. (14) and Eq. (16):

- $\text{sgn}(\mathbf{c}^{(t)}(u) - \mathbf{c}^{(t+1)}(u)) = \text{sgn}(\mathbf{y}(u));$
- $|\mathbf{c}^{(t)}(u) - \mathbf{c}^{(t+1)}(u)| > \lambda_k^t(1 - \lambda_2) |\mathbf{y}(u)|.$  □

## 5 Bipartite graphs

Assume  $G = (V, E, w)$  is an edge-weighted bipartite graph with  $V = V_1 \cup V_2$  and  $E \subseteq V_1 \times V_2$ , i.e. a graph with hidden partition identified by the bipartition. In this case, basic properties of random walks imply that the AVERAGING dynamics does not converge to the global (weighted) average of the values, but it periodically oscillates. In fact, in this case the transition matrix  $P$  has an eigenvector  $\chi = \mathbf{1}_{V_1} - \mathbf{1}_{V_2}$  with eigenvalue  $\lambda_n = -1$  (as implied by Lemma 5.1). Thus, the state vector is mainly affected by the eigenvectors associated to the two eigenvalues of absolute value 1 (i.e.,  $\lambda_1$  and  $\lambda_n$ ). After a number of rounds of the dynamics that depends on  $1/\lambda_2$ , we have that, in even rounds, all nodes in  $V_i$  ( $i = 1, 2$ ) have a state that is close to some local average  $\mu_i$ ; in odd rounds, these values are swapped (as shown in Eq. (17)).

If one were observing the process in even rounds,<sup>8</sup> however, the states of nodes in  $V_1$  would converge to  $\mu_1$  and those of nodes in  $V_2$  would converge to  $\mu_2$ . Unfortunately, convergence to the local average for nodes belonging to the same community does not eventually become monotone (i.e., increasing or decreasing). This follows since the eigenvector associated to  $\lambda_2$  is no longer stepwise in general. However, we can easily modify the labeling scheme of the AVERAGING dynamics to perform *bipartiteness detection* as follows: Nodes apply the labeling rule every two time steps and they do it between the states of two consecutive rounds, i.e., each node  $v \in V$  sets  $\text{label}^{(2t)}(v) = 1$  if  $\mathbf{x}^{(2t)}(v) \geq \mathbf{x}^{(2t-1)}(v)$  and  $\text{label}^{(2t)}(v) = 0$  otherwise. We call this new protocol AVERAGING BIPARTITE dynamics.

We now show how AVERAGING BIPARTITE dynamics can perform *bipartiteness detection*. Recall that we denote with  $W \in \mathbb{R}^{n \times n}$  the weighted adjacency matrix of  $G$ . Since  $G$  is undirected and bipartite, the matrix  $W$  can be written as

$$W = \begin{pmatrix} 0 & W_1 \\ W_2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & W_1 \\ W_1^\top & 0 \end{pmatrix}.$$

Thus, the transition matrix of a random walk on  $G$ , i.e.,  $P = D^{-1}W$  where  $D^{-1}$  is a diagonal matrix and  $D_{ii} = \frac{1}{\delta(i)}$ , has the form

$$P = \begin{pmatrix} 0 & P_1 \\ P_1^\top & 0 \end{pmatrix}.$$

Lemma 5.1 shows that the spectrum of  $P$  is symmetric and it gives a relation between the eigenvectors of symmetric eigenvalues.

**Lemma 5.1.** *Let  $G = (V_1 \cup V_2, E, w)$  be an edge-weighted undirected bipartite graph with bipartition  $(V_1, V_2)$  and such that  $|V_i| = n_i$ . If  $\mathbf{v} = (\mathbf{v}_1, \mathbf{v}_2)^\top$ , with  $\mathbf{v}_i \in \mathbb{R}^{n_i}$ , is an eigenvector of  $P$  with eigenvalue  $\lambda$ , then  $\mathbf{v}' = (\mathbf{v}_1, -\mathbf{v}_2)^\top$  is an eigenvector of  $P$  with eigenvalue  $-\lambda$ .*

*Proof.* If  $P\mathbf{v} = \lambda\mathbf{v}$  then we have that  $P_1\mathbf{v}_2 = \lambda\mathbf{v}_1$  and  $P_1^\top\mathbf{v}_2 = \lambda\mathbf{v}_2$ . Using these two equalities we get that  $P\mathbf{v}' = -\lambda\mathbf{v}'$ . Indeed,

$$P\mathbf{v}' = \begin{pmatrix} 0 & P_1 \\ P_1^\top & 0 \end{pmatrix} \begin{pmatrix} \mathbf{v}_1 \\ -\mathbf{v}_2 \end{pmatrix} = \begin{pmatrix} -P_1\mathbf{v}_2 \\ P_1^\top\mathbf{v}_1 \end{pmatrix} = -\lambda \begin{pmatrix} \mathbf{v}_1 \\ -\mathbf{v}_2 \end{pmatrix}. \quad \square$$

The transition matrix  $P$  is stochastic, thus the vector  $\mathbf{1}$  (i.e., the vector of all ones) is an eigenvector associated to  $\lambda_1 = 1$ , that is the first largest eigenvalue of  $P$ . Lemma 5.1 implies that  $\chi = \mathbf{1}_{V_1} - \mathbf{1}_{V_2}$  is an eigenvector of  $P$  with eigenvalue  $\lambda_n = -1$ .

As in Section 2, we write the state vector at time  $t$  using the spectral decomposition of  $P$ . Let  $1 = \lambda_1 > \lambda_2 \geq \dots > \lambda_n = -1$  be the eigenvalues of  $P$ . We denote by  $\mathbf{1} = \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n = \chi$

---

<sup>8</sup>Or, equivalently, in odd rounds.

a family of  $n$  linearly independent eigenvectors of  $P$ , where each  $\mathbf{v}_i$  is the eigenvector associated to  $\lambda_i$ . Thus, we have that

$$\mathbf{x}^{(t)} = P^t \mathbf{x} = \sum_{i=1}^n \lambda_i^t \alpha_i \mathbf{v}_i = \alpha_1 \mathbf{1} + (-1)^t \alpha_n \boldsymbol{\chi} + \sum_{i=2}^{n-1} \lambda_i^t \alpha_i \mathbf{v}_i \quad (17)$$

where  $\alpha_i = \frac{\langle D^{\frac{1}{2}} \mathbf{x}, D^{\frac{1}{2}} \mathbf{v}_i \rangle}{\|D^{\frac{1}{2}} \mathbf{v}_i\|^2}$ . The last equation implies that  $\mathbf{x}^{(t)} = P^t \mathbf{x}$  does not converge to some value as  $t$  tends to infinity, but oscillates. In particular, nodes in  $V_1$  on even rounds and nodes in  $V_2$  on odd rounds, converge to  $\alpha_1 + \alpha_n$ . Instead in the symmetric case, i.e., odd rounds for nodes in  $V_1$  and even rounds for nodes in  $V_2$ , the process converges to  $\alpha_1 - \alpha_n$ . These quantities are proportional to the weighted average of the initial values in the first and in the second partition, respectively.

Theorem 5.2, whose proof follows, shows that AVERAGING BIPARTITE dynamics performs bipartiteness detection in  $\mathcal{O}(\log n / \log(1/\lambda_2))$  rounds. Note that, as in the case of volume-regular graphs with two communities (see Remark 4), one single run of the dynamics identifies the bipartition. Moreover, if  $\log(1/\lambda_2) = \Omega(1)$ , then the AVERAGING BIPARTITE dynamics takes logarithmic time to find the bipartition.

**Theorem 5.2.** *Let  $G = (V, E, w)$  be an edge-weighted bipartite graph with bipartition  $(V_1, V_2)$  and maximum weighted degree  $\Delta = \mathcal{O}(n^K)$ , for any arbitrary positive constant  $K$ . Then for every time  $t > T$ , with  $T = \mathcal{O}(\log n / \log(1/\lambda_2))$ , the AVERAGING BIPARTITE dynamics truncated at round  $t$  is a  $(\mathcal{O}(n^{-\frac{1}{2}}), \mathcal{O}(n^{-\frac{1}{2}}))$ -community-sensitive algorithm.*

*Proof.* We assume that the labeling rule is applied between every even and every odd round (conversely, the signs of the nodes in the analysis are swapped). Recall the definition of the error contribution, namely  $\mathbf{e}^{(t)}(u) = \sum_{i=2}^{n-1} \lambda_i^t \alpha_i \mathbf{v}_i(u)$ . We compute the difference between the state vectors of two consecutive steps by using Eq. (17), namely

$$\begin{aligned} \mathbf{x}^{(2t)} - \mathbf{x}^{(2t+1)} &= \alpha_1 \mathbf{1} + (-1)^{2t} \alpha_n \boldsymbol{\chi} + \mathbf{e}^{(2t)} - \alpha_1 \mathbf{1} - (-1)^{2t+1} \alpha_n \boldsymbol{\chi} - \mathbf{e}^{(2t+1)} \\ &= 2\alpha_n \boldsymbol{\chi} + \mathbf{e}^{(2t)} - \mathbf{e}^{(2t+1)}. \end{aligned}$$

We want to find a time  $T$  such that for every  $t > T$  the sign of a node  $u \in V$  depends only on  $\boldsymbol{\chi}(u)$ , i.e.,  $\text{sgn}(\mathbf{x}^{(2t)}(u) - \mathbf{x}^{(2t+1)}(u)) = \text{sgn}(\alpha_n \boldsymbol{\chi}(u))$ . Since  $|\boldsymbol{\chi}(u)| = 1$ , the last equation holds whenever

$$2|\alpha_n| > |\mathbf{e}^{(2t)}(u) - \mathbf{e}^{(2t+1)}(u)|. \quad (18)$$

We upper bound  $|\mathbf{e}^{(2t)}(u) - \mathbf{e}^{(2t+1)}(u)|$  by using Lemma 4.5, getting that  $|\mathbf{e}^{(2t)}(u) - \mathbf{e}^{(2t+1)}(u)| \leq 2\lambda_2^{2t} \sqrt{\Delta n}$ . Therefore, with some algebra we get that Eq. (18) holds in every round  $t > T$ , where  $T$  is defined as

$$T := \frac{\log(\sqrt{\Delta n}/|\alpha_n|)}{2 \log(1/\lambda_2)}.$$

To conclude the proof, we provide a lower bound on  $|\alpha_n|$  showing that it is not too small, w.h.p.

Recall that  $\alpha_i = \frac{\langle D^{\frac{1}{2}} \mathbf{x}, D^{\frac{1}{2}} \mathbf{v}_i \rangle}{\|D^{\frac{1}{2}} \mathbf{v}_i\|^2}$  and thus

$$\alpha_n = \frac{\langle D^{\frac{1}{2}} \mathbf{x}, D^{\frac{1}{2}} \boldsymbol{\chi} \rangle}{\|D^{\frac{1}{2}} \boldsymbol{\chi}\|^2} = \frac{1}{\text{vol}(V)} \sum_{v \in V} \delta(v) \mathbf{x}(v) \boldsymbol{\chi}(v). \quad (19)$$

The lower bound then follows, with high probability. Indeed,

$$\mathbf{P} \left( |\alpha_n| \leq \frac{1}{\Delta n} \right) \leq \mathbf{P} \left( |\alpha_n| \leq \frac{1}{\text{vol}(V)} \right) \stackrel{(a)}{=} \mathbf{P} \left( \left| \sum_{v \in V} \delta(v) \mathbf{x}(v) \chi(v) \right| \leq 1 \right) \stackrel{(b)}{=} \mathcal{O} \left( \frac{1}{\sqrt{n}} \right),$$

where in (a) we used Eq. (19) and in (b) we applied Theorem A.5. The thesis then follows from the above bound on  $|\alpha_n|$  and from the hypothesis on  $\Delta = \mathcal{O}(n^K)$ , for any arbitrary positive constant  $K$ .  $\square$

## 6 Discussion and Outlook

The focus of this work is on heuristics that implicitly perform spectral graph clustering, without explicitly computing the main eigenvectors of a matrix describing connectivity properties of the underlying network (typically, its Laplacian or a related matrix). In this perspective, we extended the work of Becchetti et al. [BCN<sup>+</sup>17] in several ways. In particular, for  $k$  communities, [BCN<sup>+</sup>17] considered an extremely regular case, in which the second eigenvalue of the (normalized) Laplacian has algebraic and geometric multiplicities  $k - 1$  and the corresponding eigenspace is spanned by a basis of indicator vectors. We considered a more general case in which the first  $k$  eigenvalues are in general different, but the span of the corresponding eigenvectors again admits a base of indicator vectors. We also made a connection between this stepwise property and lumpability properties of the underlying random walk, which results in a class of volume-regular graphs, that may not have constant degree, nor exhibit balanced communities. We further showed that our approach naturally lends itself to addressing related, yet different problems, such as identifying bipartiteness. Finally, in the paragraphs that follow we discuss extensions to slightly more general classes than the ones considered in this work.

**Other graph classes.** Consider  $k$ -volume regular graphs whose  $k$  stepwise eigenvectors are associated to the  $k$  largest eigenvalues, in absolute value. These graphs include many  $k$ -partite graphs (e.g., regular ones), graphs that are “close” to being  $k$ -partite (i.e., ones that would become  $k$ -partite upon removal of a few edges). Differently from the clustered case (Theorem 4.1) some of the  $k$  eigenvalues can in general be negative.

Consider the following variant of the labeling scheme of the AVERAGING dynamics, in which nodes apply their labeling rule only on even rounds, comparing their value with the one they held at the end of the last even round, i.e., each node  $v \in V$  sets  $\text{label}^{(2t)}(v) = 1$  if  $\mathbf{x}^{(2t)}(v) \geq \mathbf{x}^{(2t-2)}(v)$  and  $\text{label}^{(2t)}(v) = 0$  otherwise. Since the above protocol amounts to only taking even powers of eigenvalues, the analysis of this modified protocol proceeds along the same lines as the clustered case, while the results of Theorem 4.1 seamlessly extend to this class of graphs.

**Outlook.** Though far from conclusive, we believe our results point to potentially interesting directions for future research. In general, our analysis sheds further light on the connections between temporal evolution of the power method and spectral-related clustering properties of the underlying network. At the same time, we showed that variants of the AVERAGING dynamics (and/or its labeling rule) might be useful in addressing different problems and/or other graph classes, as the examples given in Section 5 suggest. On the other hand, identifying  $k$  hidden partitions using the algorithm presented in [BCN<sup>+</sup>17] requires relatively strong assumptions on the  $k$  main eigenvalues and knowledge of an upper bound to the graph size,<sup>9</sup> while the analysis

---

<sup>9</sup>As anecdotal experimental evidence suggests, the presence of a time window to perform labeling is not an artifact of our analysis.

becomes considerably more intricate than the perfectly regular and completely balanced case addressed in [BCN<sup>+</sup>17]. Some aspects of our analysis (e.g., the aforementioned presence of a size-dependent time window in which the labeling rule has to be applied) suggest that more sophisticated variants of the AVERAGING dynamics might be needed to express the full power of a spectral method that explicitly computes the  $k$  main eigenvectors of a graph-related matrix. While we believe this goal can be achieved, designing and analyzing such an algorithm might prove a challenging task.

## Appendix

### A Useful inequalities

**Theorem A.1** (Extension of Chernoff Bounds [DP09]). *Let  $X = \sum_{i=1}^n X_i$  where  $X_i$  are independent distributed random variables taking values in  $\{0, 1\}$  and let  $\mu = \mathbf{E}[X]$ . Suppose that  $\mu_L \leq \mu \leq \mu_H$ . Then, for  $0 < \delta < 1$ ,*

$$\mathbf{P}(X > (1 + \delta)\mu_H) \leq \exp\left(-\frac{\delta^2}{3}\mu_H\right),$$

$$\mathbf{P}(X < (1 - \delta)\mu_L) \leq \exp\left(-\frac{\delta^2}{2}\mu_L\right).$$

**Theorem A.2** (Cauchy-Schwarz's inequality). *For all vectors  $\mathbf{u}, \mathbf{v}$  of an inner product space it holds that  $|\langle \mathbf{u}, \mathbf{v} \rangle|^2 \leq \langle \mathbf{u}, \mathbf{u} \rangle \cdot \langle \mathbf{v}, \mathbf{v} \rangle$ , where  $\langle \cdot, \cdot \rangle$  is the inner product.*

**Theorem A.3** (Cheeger's inequality [Chu96]). *Let  $P$  be the transition matrix of a connected edge-weighted graph  $G = (V, E, w)$  and let  $\lambda_2$  be its second largest eigenvalue. Let  $|E(S, V \setminus S)| = \sum_{u \in S, v \in V \setminus S} w(u, v)$  and  $h_G = \min_{S: \text{vol}(S) \leq \frac{\text{vol}(V)}{2}} \frac{|E(S, V \setminus S)|}{\text{vol}(S)}$ . Then*

$$\frac{1 - \lambda_2}{2} \leq h_G \leq \sqrt{2(1 - \lambda_2)}.$$

**Theorem A.4** (Berry-Esseen's theorem [Ber41]). *Let  $X_1, \dots, X_n$  be independent random variables with mean  $\mu_i = 0$ , variance  $\sigma_i^2 > 0$ , and third absolute moment  $\rho_i < \infty$ , for every  $i = 1, \dots, n$ . Let  $S_n = \sum_{i=1}^n X_i$  and let  $\sigma = \sqrt{\sum_{i=1}^n \sigma_i^2}$  be the standard deviation of  $S_n$ ; let  $F_n$  be the cumulative distribution function of  $\frac{S_n}{\sigma}$ ; let  $\Phi$  the cumulative distribution function of the standard normal distribution. Then, there exists a positive constant  $C$  such that, for all  $x$  and for all  $n$ ,*

$$|F_n(x) - \Phi(x)| \leq \frac{C\psi}{\sigma},$$

where  $\psi := \max_{i \in \{1, \dots, n\}} \frac{\rho_i}{\sigma_i^3}$ .

**Theorem A.5** (Littlewood-Offord's small ball [Erd45]). *Let  $a_1, \dots, a_n \in \mathbb{R}$  be real numbers with  $|a_i| \geq 1$  for every  $i = 1, \dots, n$  and let  $r \in \mathbb{R}$  be any real number. Let  $\{X_i : i = 1, \dots, n\}$  be a family of independent Rademacher random variables (taking values  $\pm 1$  with probability  $1/2$ ) and let  $X$  be their sum weighted with the  $a_i$ s, i.e.,  $X = \sum_{i=1}^n a_i X_i$ , then*

$$\mathbf{P}(|X - r| < 1) = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right).$$

## References

- [AAE08] Dana Angluin, James Aspnes, and David Eisenstat. A Simple Population Protocol for Fast Robust Approximate Majority. *Distributed Computing*, 21(2):87–102, 2008. (Preliminary version in DISC’07).
- [ABH14] Emmanuel Abbe, Afonso S. Bandeira, and Georgina Hall. Exact recovery in the stochastic block model. *IEEE Trans. on Information Theory*, 62(1):471–487, 2014.
- [AS15] Emmanuel Abbe and Colin Sandon. Detection in the stochastic block model with multiple clusters: proof of the achievability conjectures, acyclic bp, and the information-computation gap. *arXiv preprint arXiv:1512.09080*, 2015.
- [BC09] Michael J. Barber and John W. Clark. Detecting network communities by propagating labels under constraints. *Physical Review E*, 80(2):026129, 2009.
- [BCM<sup>+</sup>18] Luca Becchetti, Andrea E.F. Clementi, Pasin Manurangsi, Emanuele Natale, Francesco Pasquale, Prasad Raghavendra, and Luca Trevisan. Average whenever you meet: Opportunistic protocols for community detection. In *26th Annual European Symposium on Algorithms, ESA 2018, August 20-22, 2018, Helsinki, Finland*, pages 7:1–7:13, 2018.
- [BCN<sup>+</sup>17] Luca Becchetti, Andrea E.F. Clementi, Emanuele Natale, Francesco Pasquale, and Luca Trevisan. Find your place: Simple distributed algorithms for community detection. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017, Barcelona, Spain, Hotel Porta Fira, January 16-19*, pages 940–959, 2017.
- [BCPR19] Luca Becchetti, Emilio Cruciani, Francesco Pasquale, and Sara Rizzo. Step-by-step community detection in volume-regular graphs. In *30th International Symposium on Algorithms and Computation (ISAAC 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.
- [Ber41] Andrew C Berry. The accuracy of the gaussian approximation to the sum of independent variates. *Transactions of the american mathematical society*, 49(1):122–136, 1941.
- [BLM15] Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Non-backtracking spectrum of random graphs: community detection and non-regular Ramanujan graphs. In *Proc. of 56rd Ann. IEEE Symp. on Foundations of Computer Science (FOCS’15)*, pages 1347–1357. IEEE, 2015.
- [Bop87] Ravi B. Boppana. Eigenvalues and graph bisection: An average-case analysis. In *Proc. of the 28th Ann. IEEE Symp. on Foundations of Computer Science (FOCS’87)*, pages 280–285. IEEE, 1987.
- [Chu96] Fan R.K. Chung. Laplacians of graphs and Cheeger’s inequalities. In *Combinatorics, Paul Erdős is eighty*, volume 2, pages 157–172. János Bolyai Math. Soc., 1996.
- [CNNS18] Emilio Cruciani, Emanuele Natale, André Nusser, and Giacomo Scornavacca. Phase transition of the 2-Choices dynamics on core-periphery networks. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS 2018, Stockholm, Sweden, July 10-15, 2018*, pages 777–785, 2018.

- [CNS19] Emilio Cruciani, Emanuele Natale, and Giacomo Scornavacca. Distributed community detection via metastability of the 2-Choices dynamics. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence, AAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, 2019.
- [CO05] Amin Coja-Oghlan. *Spectral techniques, semidefinite programs, and random graphs*. Habilitation thesis, Humboldt University Berlin, 2005.
- [CO10] Amin Coja-Oghlan. Graph partitioning via adaptive spectral techniques. *Combinatorics, Probability and Computing*, 19(02):227–284, 2010.
- [DF89] Martin E. Dyer and Alan M. Frieze. The solution of some random NP-hard problems in polynomial expected time. *Journal of Algorithms*, 10(4):451–489, 1989.
- [DKMZ11] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84(6):066106, 2011.
- [DP09] Devdatt P Dubhashi and Alessandro Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, 2009.
- [Erd45] Paul Erdős. On a lemma of littlewood and offord. *Bulletin of the American Mathematical Society*, 51(12):898–902, 1945.
- [Fie89] Miroslav Fiedler. Laplacian of graphs and algebraic connectivity. *Banach Center Publications*, 25(1):57–70, 1989.
- [For10] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75–174, 2010.
- [HLL83] Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.
- [HP01] Yehuda Hassin and David Peleg. Distributed probabilistic polling and applications to proportionate agreement. *Information and Computation*, 171(2):248–268, 2001.
- [JS98] Mark Jerrum and Gregory B. Sorkin. The metropolis algorithm for graph bisection. *Discrete Applied Mathematics*, 82(1):155–175, 1998.
- [KM04] David Kempe and Frank McSherry. A decentralized algorithm for spectral analysis. In *Proc. of the 36th Ann. ACM Symp. on Theory of Computing (STOC’04)*, pages 561–568, 2004.
- [KMM<sup>+</sup>13] Florent Krzakala, Cristopher Moore, Elchanan Mossel, Joe Neeman, Allan Sly, Lenka Zdeborová, and Pan Zhang. Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences*, 110(52):20935–20940, 2013.
- [KPS13] Kishore Kothapalli, Sriram V. Pemmaraju, and Vivek Sardeshmukh. On the analysis of a label propagation algorithm for community detection. In *Proc. of the 14th Int. Conf. on Distributed Computing and Networking (ICDCN’13)*, pages 255–269, 2013.
- [KS60] John G. Kemeny and J. Laurie Snell. *Finite Markov chains*. D. van Nostrand Company, inc., Princeton, N.J., 1960.

- [Lan50] Cornelius 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.
- [LGT14] James R Lee, Shayan Oveis Gharan, and Luca Trevisan. Multiway spectral partitioning and higher-order cheeger inequalities. *Journal of the ACM (JACM)*, 61(6):37, 2014.
- [LM10] Xin Liu and Tsuyoshi Murata. Advanced modularity-specialized label propagation algorithm for detecting communities in networks. *Physica A: Statistical Mechanics and its Applications*, 389(7):1493–1500, 2010.
- [LRU14] Jure Leskovec, Anand Rajaraman, and Jeffrey David Ullman. *Mining of massive datasets*. Cambridge university press, 2014.
- [Mac03] David J.C. MacKay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.
- [McS01] Frank McSherry. Spectral partitioning of random graphs. In *Proc. of the 42nd IEEE Symp. on Foundations of Computer Science (FOCS'01)*, pages 529–537, 2001.
- [MK07] Joris M. Mooij and Hilbert J. Kappen. Sufficient conditions for convergence of the sum–product algorithm. *IEEE Transactions on Information Theory*, 53(12):4422–4437, 2007.
- [MMM18] Frederik Mallmann-Trenn, Cameron Musco, and Christopher Musco. Eigenvector computation and community detection in asynchronous gossip models. In *45th International Colloquium on Automata, Languages, and Programming, ICALP 2018, July 9-13, 2018, Prague, Czech Republic*, pages 159:1–159:14, 2018.
- [MNS13] Elchanan Mossel, Joe Neeman, and Allan Sly. A proof of the block model threshold conjecture. *arXiv preprint arXiv:1311.4115*, 2013.
- [MNS16] Elchanan Mossel, Joe Neeman, and Allan Sly. Belief propagation, robust reconstruction and optimal recovery of block models. *Ann. Appl. Probab.*, 26(4):2211–2256, 08 2016.
- [NJW02] Andrew Y Ng, Michael I Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in neural information processing systems*, pages 849–856, 2002.
- [PSZ17] Richard Peng, He Sun, and Luca Zanetti. Partitioning well-clustered graphs: Spectral clustering works! *SIAM Journal on Computing*, 46(2):710–743, 2017.
- [RAK07] Usha Nandini Raghavan, Réka Albert, and Soundar Kumara. Near linear time algorithm to detect community structures in large-scale networks. *Physical Review E*, 76(3):036106, 2007.
- [SM00] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *Departmental Papers (CIS)*, page 107, 2000.
- [SZ17] He Sun and Luca Zanetti. Distributed graph clustering and sparsification. *CoRR*, abs/1711.01262, 2017.

- [TK06] Jianjun Paul Tian and D Kannan. Lumpability and commutativity of markov processes. *Stochastic analysis and Applications*, 24(3):685–702, 2006.
- [VL07] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.
- [Wei00] Yair Weiss. Correctness of local probability propagation in graphical models with loops. *Neural computation*, 12(1):1–41, 2000.