

# Eigenvector Computation and Community Detection in Asynchronous Gossip Models

Frederik Mallmann-Trenn<sup>1</sup>, Cameron Musco<sup>1</sup>, and Christopher Musco<sup>1</sup>

<sup>1</sup>MIT CSAIL, Cambridge MA, US, {mallmann, cnmusco, cpmusco}@mit.edu

May 1, 2018

## Abstract

We give a simple distributed algorithm for computing adjacency matrix eigenvectors for the communication graph in an asynchronous gossip model. We show how to use this algorithm to give state-of-the-art asynchronous community detection algorithms when the communication graph is drawn from the well-studied *stochastic block model*. Our methods also apply to a natural alternative model of randomized communication, where nodes within a community communicate more frequently than nodes in different communities.

Our analysis simplifies and generalizes prior work by forging a connection between asynchronous eigenvector computation and Oja’s algorithm for streaming principal component analysis. We hope that our work serves as a starting point for building further connections between the analysis of stochastic iterative methods, like Oja’s algorithm, and work on asynchronous and gossip-type algorithms for distributed computation.

**keywords**— block model, community detection, distributed clustering, eigenvector computation, gossip algorithms, population protocols

**funding**— This work was supported in part by NSF Award Numbers BIO-1455983, CCF-1461559, CCF-0939370, and CCF-1565235. Cameron Musco was partially supported by an NSF graduate student fellowship.

# 1 Introduction

Motivated by the desire to process and analyze increasingly large networks—in particular social networks—considerable research has focused on finding efficient distributed protocols for problems like triangle counting, community detection, PageRank computation, and node centrality estimation. Many of the most popular systems for massive-scale graph processing, including Google’s Pregel [23] and Apache Giraph [33] (used by Facebook), employ programming models based on the simulation of distributed message passing algorithms, in which each node is viewed as a processor that can send messages to its neighbors.

Apart from computational benefits, distributed graph processing can also be required when privacy constraints apply: for example, EU regulations restrict the personal data sent to countries outside of the EU [9]. Distributed algorithms avoid possibly problematic aggregation of network information, allowing each node to compute a local output based on their own neighborhood and messages received from their neighbors.

One of the main problems of interest in network analysis is the computation of the eigenvectors of a networks’ adjacency matrix (or related incidence matrices, such as the graph Laplacian). The extremal eigenvectors have many important applications—from graph partitioning and community detection [15, 26], to embedding in graph-based machine learning [5, 29], to measuring node centrality and computing importance scores like PageRank [6].

Due to their importance, there has been significant work on distributed eigenvector approximation. In *synchronous* message passing systems, it is possible to simulate the well-known power method for iterative eigenvector approximation [21]. However, this algorithm requires that each node communicates synchronously with all of its neighbors in each round.

In an attempt to relax this requirement, models in which a subset of neighbors are sampled in each communication round [22] have been studied. However, the computation of graph eigenvectors in fully asynchronous and gossip-based message passing systems, in which nodes communicate with a single neighbor at a time in an asynchronous fashion, is not well-understood. While a number of algorithms have been proposed, which give convergence to the true eigenvectors as the number of iterations goes to infinity, strong finite iteration approximation bounds are not known [16, 27].

**Our contributions** In this work, we give state-of-the-art algorithms for graph eigenvector computation in asynchronous systems with randomized schedulers, including the classic gossip model [8, 14] and population protocol model [2]. We show that in these models, communication graph eigenvectors can be computed via a very simple adaption of Oja’s classic iterative algorithm for principal components analysis [30]. Our analysis leverages recent work studying Oja’s algorithm for streaming covariance matrix eigenvector estimation [1, 20].

By making an explicit connection between work on streaming eigenvector estimation and asynchronous computation, we hope to generally expand the toolkit of techniques that can be applied to analyzing graph algorithms in asynchronous systems.

As a motivating application, we use our results to give state-of-the-art distributed community detection protocols, significantly improving upon prior work for the well-studied stochastic-block model and related models where nodes communicate more frequently within their community than outside of it. We summarize our results below.

**Asynchronous eigenvector computation.** First, we provide an algorithm (Algorithm 2) that approximates the  $k$  largest eigenvectors  $\mathbf{v}_1, \dots, \mathbf{v}_k$  for an arbitrary communication matrix (essentially a normalized adjacency matrix, defined formally in Definition 2.1).

For an  $n$ -node network, the algorithm ensures, with good probability, that each node  $u \in [n]$  computes the  $u^{th}$  entries of vectors  $\tilde{\mathbf{v}}_1, \dots, \tilde{\mathbf{v}}_k$  such that for all  $i \in [k]$ ,  $\|\tilde{\mathbf{v}}_i - \mathbf{v}_i\|_2^2 \leq \epsilon$ . Each message

sent by the algorithm requires communicating just  $O(k)$  numbers, and the global time complexity is  $\tilde{O}(\frac{\Lambda k^3}{\text{gap} \cdot \min(\text{gap}, \gamma_{mix}) \epsilon^3})$  local rounds, where  $\text{gap}$  is the minimal gap between the  $k$  largest eigenvalues,  $\gamma_{mix}$  is roughly speaking the spectral gap, i.e., the difference between the largest and second-largest eigenvalue, and  $\Lambda$  is the sum of the  $k$  largest eigenvalues. We note that we use  $\tilde{O}(\cdot)$  to suppress logarithmic terms, and in particular, factors of  $\text{poly} \log n$ . See [Theorem 3.4](#) for a more precise statement.

For illustration, consider a communication graph generated via the stochastic block model –  $G(n, p, q)$ , which has  $n$  nodes, partitioned into two equal-sized clusters. Each intracluster edge added independently with probability  $p$  and each intercluster edge is added with probability  $q < p$ . If, for example,  $p = \Omega(\frac{\log n}{n})$  and  $q = p/2$ , and  $k = 2$ , we can bound with high probability  $\Lambda = \Theta(1/n)$ ,  $\text{gap} = \Theta(1/n)$ , and  $\gamma_{mix} = \Theta(1/n)$ , which yields an eigenvector approximation algorithm running in  $\tilde{O}(\frac{n}{\epsilon^3})$  global rounds, or  $\tilde{O}(\frac{1}{\epsilon^3})$  local rounds.

**Approximate community detection.** Second, we harness our eigenvector approximation routine for community detection in the stochastic block model with connection probabilities  $p, q$  (we give two natural definitions of this model in an asynchronous distributed system with a random scheduler; see [Definitions 2.4](#) and [2.5](#)). After executing our protocol ([Algorithm 5](#)), with good probability, all but an  $\epsilon$  fraction of the nodes output a correct community label in  $\tilde{O}(1/\epsilon^3 \rho^2)$  local rounds, where  $\rho = \min(\frac{q}{p+q}, \frac{p-q}{p+q})$ . For example, when  $q = p/2$ , this complexity is  $\tilde{O}(1/\epsilon^3)$ . See [Theorem 4.2](#) and [Theorem 4.3](#) for precise bounds.

**Exact community detection.** Finally, we show how to produce an exact community labeling, via a simple gossip-based error correction scheme. For ease of presentation, here we just state our results in the case when  $q = p/2$  and we refer to [Section 5](#) ([Theorems 5.1](#) and [5.2](#)) for general results. Starting from an approximate labeling in which only a small constant fraction of the nodes are incorrectly labeled, we show that, with high probability, after  $O(\log n)$  local rounds, all nodes are labeled correctly.

**Related work** Community detection via graph eigenvector computation and other spectral methods has received ample attention in centralized setting [[25](#), [10](#), [38](#)]. Such methods are known to recover communities in the stochastic block model close to the information theoretic limit. Interestingly, many state-of-the-art community detection algorithms in this model, which improve upon spectral techniques, are based on message passing (belief propagation) algorithms [[12](#), [28](#)]. However, these algorithms are not known to work in asynchronous contexts.

Community detection in asynchronous distributed systems has received less attention. It has recently been tackled in a beautiful paper by Becchetti et al. [[3](#)]. The algorithm studied in this paper is a very simple averaging protocol, originally considered by the authors in a synchronous setting [[4](#)]. Each node starts with a random value chosen uniformly in  $\{-1, 1\}$ . Each time two nodes communicate, they update their values to the average of their previous values. After each round of communication, a node's estimated community is given by the *sign* of the change of its value due to the averaging update in that round.

Becchetti et al. analyze their algorithm for *regular* clustered graphs, including regular stochastic block model graphs, where all nodes have exactly  $a$  edges to (randomly selected) nodes in their cluster and exactly  $b < a$  edges to nodes outside their cluster. As discussed in [[3](#)], for regular graphs their protocol can be viewed as estimating the sign of entries in the second largest adjacency matrix eigenvector. Thus, it has close connections with our protocols, which explicitly estimate this eigenvector and label communities using the signs of its entries.

The results of Becchetti et al. apply with  $O(\text{polylog } n)$  local rounds of communication when either  $\frac{a}{b} = \Omega(\log^2 n)$ , or when  $a-b = \Omega(\sqrt{a+b})$ . In contrast, our results for the (non-regular) stochastic block model give  $O(\text{polylog } n)$  local runtime when  $\frac{p}{q} = \Omega(1)$  or  $n(p-q) = \Omega(\sqrt{n(p+q)\log n})$ . Here we assume that  $q$  is not too small – see [Theorem 4.3](#) for details. Note that  $n \cdot p$  and  $n \cdot q$  can be compared to  $a$  and  $b$ , since they are the expected number of intra- and inter-cluster edges respectively. Thus, our results give comparable bounds, tightening those of Becchetti et al. in some regimes and holding in the most commonly studied family of stochastic block model graphs, without any assumption of regularity<sup>1</sup>.

Outside of community detection, our approach to asynchronous eigenvector approximation is related to work on asynchronous distributed stochastic optimization [\[37, 11, 31\]](#). Often, it is assumed that many processors update some decision variable in parallel. If these updates are sufficiently sparse, overwrites are rare and the algorithm converges as if it were run in a synchronous manner. Our implementation of Oja’s algorithm falls under this paradigm. Each update to our eigenvector estimates is sparse – requiring a modification just by the two nodes that communicate at a given time. In this way, we can fully parallelize the algorithm, even in an asynchronous system.

## 2 Preliminaries

### 2.1 Notation

For integer  $n > 0$ , let  $[n] \stackrel{\text{def}}{=} \{1, \dots, n\}$ . Let  $\mathbf{1}_{n,m}$  be an  $n \times m$  all-ones matrix and  $\mathbf{I}_{n \times n}$  be an  $n \times n$  identity. Let  $\mathbf{e}_i$  be the  $i^{\text{th}}$  standard basis vector, with length apparent from context. Let  $V$  denote a set of nodes with cardinality  $|V| = n$ . Let  $\mathcal{P}$  be the set of all unordered node pairs  $(u, v)$  with  $u \neq v$ .  $|\mathcal{P}| = \binom{n}{2}$ .

For vector  $\mathbf{x} \in \mathbb{R}^n$ ,  $\|\mathbf{x}\|_2$  is the Euclidean norm. For matrix  $\mathbf{M} \in \mathbb{R}^{n \times m}$ ,  $\|\mathbf{M}\|_2 = \max_{\mathbf{x}} \frac{\|\mathbf{M}\mathbf{x}\|_2}{\|\mathbf{x}\|_2}$  is the spectral norm.  $\|\mathbf{M}\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^m \mathbf{M}_{i,j}^2}$  denotes the Frobenius norm.  $\mathbf{M}^T$  is the matrix transpose of  $\mathbf{M}$ . When  $\mathbf{M} \in \mathbb{R}^{n \times n}$  is symmetric we let  $\lambda_1(\mathbf{M}) \geq \lambda_2(\mathbf{M}) \geq \dots \geq \lambda_n(\mathbf{M})$  denote its eigenvalues.  $\mathbf{M}$  is positive semidefinite (PSD) if  $\lambda_i(\mathbf{M}) \geq 0$  for all  $i$ . For symmetric  $\mathbf{M}, \mathbf{N} \in \mathbb{R}^{n \times n}$  we use  $\mathbf{M} \preceq \mathbf{N}$  to indicate that  $\mathbf{N} - \mathbf{M}$  is PSD.

### 2.2 Computational model

We define an asynchronous distributed computation model that encompasses both the well-studied population protocol [\[2\]](#) and asynchronous gossip models [\[8\]](#). Computation proceeds in rounds and a random scheduler chooses a single pair of nodes to communicate in each round. The choice is independent across rounds, but may be nonuniform across node pairs.

**Definition 2.1** (Asynchronous communication model). *Let  $V$  be a set of nodes with  $|V| = n$ . Computation proceeds in rounds, with every node  $v \in V$  having some state  $s(v, t)$  in round  $t$ .*

*Recall that  $\mathcal{P}$  denotes all unordered pairs of nodes in  $V$ . Let  $w : \mathcal{P} \rightarrow \mathbb{R}^+$  be a nonnegative weight function. In each round, a random scheduler chooses exactly one  $(u, v) \in \mathcal{P}$  with probability  $w(u, v) / \left[ \sum_{(i,j) \in \mathcal{P}} w(i, j) \right]$  and  $u, v$  both update their states according to some common (possibly randomized) transition function  $\sigma$ . Specifically, they set  $s(v, t+1) = \sigma(s(v, t), s(u, t))$  and  $s(u, t+1) = \sigma(s(u, t), s(v, t))$ .*

---

<sup>1</sup> We note that the analysis of Becchetti et al. seems likely to extend to our alternative communication model ([Definition 2.4](#)), where the communication graph is weighted and regular

Note that in our analysis we often identify the weight function  $w$  with a symmetric weight matrix  $\mathbf{W} \in \mathbb{R}^{n \times n}$  where  $\mathbf{W}_{u,u} = 0$  and  $\mathbf{W}_{u,v} = \mathbf{W}_{v,u} = w(u,v) / \left[ \sum_{(i,j) \in \mathcal{P}} w(i,j) \right]$ . Let  $\mathbf{D}$  be a diagonal matrix with  $\mathbf{D}_{u,u} = \sum_{v \in V} \mathbf{W}_{u,v}$ .  $\mathbf{D}_{u,u}$  is the probability that node  $u$  communicates in any given round. Since two nodes are chosen in each round,  $\sum_u \mathbf{D}_{u,u} = 2$ . We will refer to  $\mathbf{D} + \mathbf{W}$  as the *communication matrix* of the communication model.

**Remark 2.2** (Asynchronous algorithms). *Since the transition function  $\sigma$  in Definition 2.1 is universal, nodes can be seen as identical processes, with no knowledge of  $w$  or unique ids. We do assume that nodes can initiate and terminate a protocol synchronously. That is, nodes interact from round 0 up to some round  $T$ , after which they cease to interact, or begin a new protocol. This assumption is satisfied if each node has knowledge of the global round number but, in general, is much weaker. For example, in the asynchronous gossip model discussed below, it is sufficient for nodes to have access to a synchronized clock.*

We use algorithm to refer to a sequence of transition functions, each corresponding to a subroutine run for specified number of rounds. Subroutines are run sequentially. The first has input nodes with identical starting states (as prescribed by Definition 2.1) but later subroutines start once nodes have updated their states and thus have distinguished inputs.

**Remark 2.3** (Simulation of existing models). *The standard population protocol model [2] is recovered from Definition 2.1 by setting  $w(u,v) = 1$  for all  $(u,v)$  – i.e., pairs of nodes communicate uniformly at random. A similar model over a fixed communication graph  $G = (E, V)$  is recovered by setting  $w(u,v) = 1$  for all  $(u,v) \in E$  and  $w(u,v) = 0$  for  $(u,v) \notin E$ .*

Definition 2.1 also encompasses the asynchronous gossip model [8, 14], where each node holds an independent Poisson clock and contacts a random neighbor when the clock ticks. If we identify rounds with clock ticks, let  $\lambda_u$  be the rate of node  $u$ 's clock, and let  $p(u,v)$  be the probability that  $u$  contacts  $v$  when its clock ticks. Then the probability that nodes  $u$  and  $v$  interact in a given round is  $\frac{1}{2} \left[ \frac{\lambda_u}{\sum_{z \in V} \lambda_z} \cdot p(u,v) + \frac{\lambda_v}{\sum_{z \in V} \lambda_z} \cdot p(v,u) \right]$ . With  $w(u,v)$  set to this value, Definition 2.1 corresponds exactly to the asynchronous gossip model.

## 2.3 Distributed community detection problem

This paper studies the very general problem of computing communication matrix eigenvectors with asynchronous protocols run by the nodes in  $V$ . One primary application of computing eigenvectors is to detect community structure in  $G$ . Below we formalize this application as the *distributed community detection problem* and introduce two specific cases of interest.

In the distributed community detection problem, the weight function  $w$  and corresponding weight matrix  $\mathbf{W}$  of Definition 2.1 are clustered: nodes in the same cluster are more likely to communicate than nodes in different clusters. The goal is for each node to independently identify what cluster it belongs to (up to a permutation of the cluster labels).

We consider two models of clustering. In the first  $(n, p, q)$ -weighted communication model, the weight function directly reflects the increased likelihood of intracommunity communication. In the second,  $G(n, p, q)$ -communication model, weights are uniform on a graph sampled from the well-studied planted-partition or *stochastic block model* [19]. For simplicity, we focus on the setting in which there are two equal sized clusters, but believe that our techniques can be extended to handle a larger number of clusters, potentially with unbalanced sizes.

**Definition 2.4** ( $(n, p, q)$ -weighted communication model). An asynchronous model (Definition 2.1), where node set  $V$  is partitioned into disjoint sets  $V_1, V_2$  with  $|V_1| = |V_2| = n/2$ . For values  $q < p$ ,  $w(u, v) = p$  if  $u, v \in V_i$  for some  $i$  and  $w(u, v) = q$  if  $u \in V_i$  and  $v \in V_j$  for  $i \neq j$ .

**Definition 2.5** ( $G(n, p, q)$ -communication model). An asynchronous model (Definition 2.1), where node set  $V$  is partitioned into disjoint sets  $V_1, V_2$  with  $|V_1| = |V_2| = n/2$ . The weight matrix  $\mathbf{W}$  is a normalized adjacency matrix of a random graph  $G(V, E)$  generated as follows: for each pair of nodes  $u, v \in V$ , add edge  $(u, v)$  to edge set  $E$  with probability  $p$  if  $u$  and  $v$  are in the same partition  $V_i$  and probability  $q < p$  if  $u$  and  $v$  are in different partitions.

Analysis of community detection in the  $(n, p, q)$ -weighted communication model is more elegant, and will form the basis of our analysis for the  $G(n, p, q)$ -communication model, which more closely matches models considered in prior work on in both distributed and centralized settings. Formally, we define the distributed community detection problem as follows:

**Definition 2.6** (Distributed community detection problem). An algorithm executing in the communication models of Definition 2.4 and Definition 2.5 solves community detection in  $T$  rounds if for every  $t \geq T$ , all nodes in  $V_1$  hold some integer state  $s_1 \in \{-1, 1\}$ , while all nodes in  $V_2$  hold state  $s_2 = -s_1$ . An algorithm solves the community detection problem in  $L$  local rounds if every node's state remains fixed after  $L$  local interactions with other nodes.

### 3 Asynchronous Oja's algorithm

Our main contribution is a distributed algorithm for computing eigenvectors of the communication matrix  $\mathbf{D} + \mathbf{W}$ . These eigenvectors can be used to solve the distributed community detection problem or in other applications. Our main algorithm is a distributed, asynchronous adaptation of Oja's classic iterative eigenvector algorithm [30], described below:

---

#### Algorithm 1 OJA'S METHOD (CENTRALIZED)

---

**Input:**  $\mathbf{x}_0, \dots, \mathbf{x}_{T-1} \in \mathbb{R}^n$  drawn i.i.d. from some distribution  $\mathcal{D}$  such that for some constant  $C$ ,  $\mathbb{P}_{\mathbf{x} \sim \mathcal{D}}[\|\mathbf{x}\|_2^2 \leq C] = 1$  and  $\mathbb{E}_{\mathbf{x} \sim \mathcal{D}}[\mathbf{x}\mathbf{x}^T] = \mathbf{M}$ . Rank parameter  $k$  and step size  $\eta$ .

**Output:** Orthonormal  $\tilde{\mathbf{V}} \in \mathbb{R}^{n \times k}$  whose columns approximate  $\mathbf{M}$ 's  $k$  top eigenvectors.

1: Choose  $\mathbf{Q}_0$  with entries drawn i.i.d. from the standard normal distribution  $\mathcal{N}(0, 1)$ .

2: **for**  $t = 0, \dots, T - 1$  **do**

3:      $\mathbf{Q}_{t+1} := (\mathbf{I} + \eta \mathbf{x}_t \mathbf{x}_t^T) \mathbf{Q}_t$ .

4: **end for**

5: **return**  $\tilde{\mathbf{V}}_T := \text{orth}(\mathbf{Q}_T)$ .

▷ Orthonormalizes the columns of  $\mathbf{Q}_T$ .

---

#### 3.1 Approximation bounds for Oja's method

A number of recent papers have provided strong convergence bounds for the *centralized* version of Oja's method [1, 20]. We will rely on the following theorem, which we prove in Appendix A using a straightforward application of the arguments in [1].

**Theorem 3.1.** Let  $\mathbf{M} \in \mathbb{R}^{n \times n}$  be a PSD matrix with  $\frac{\sum_{i=1}^k \lambda_i(\mathbf{M})}{C} \leq \Lambda$  and  $\frac{\lambda_k(\mathbf{M}) - \lambda_{k+1}(\mathbf{M})}{C} \geq \text{gap}$  for some values  $\Lambda, \text{gap}$ . For any  $\epsilon, \delta \in (0, 1)$ , let  $\xi = \frac{n}{\delta \epsilon \cdot \text{gap}}$ ,  $\eta = \frac{c_1 \epsilon^2 \cdot \text{gap} \cdot \delta^2}{C \Lambda k \log^3 \xi}$  for some sufficiently

small constant  $c_1$ , and  $T = \frac{c_2 \cdot (\log \xi + 1/\epsilon)}{C \cdot \text{gap} \cdot \eta}$  for sufficiently large  $c_2$ . Then with probability  $\geq 1 - \delta$ , [Algorithm 1](#) run with step size  $\eta$  returns  $\tilde{\mathbf{V}}_T$  satisfying,

$$\|\mathbf{Z}^T \tilde{\mathbf{V}}_T\|_F^2 \leq \epsilon.$$

where  $\mathbf{Z}$  is an orthonormal basis for the bottom  $n - k$  eigenvectors of  $\mathbf{M}$ .

If  $\tilde{\mathbf{V}}_T$  exactly spanned  $\mathbf{M}$ 's top  $k$  eigenvectors,  $\|\mathbf{Z}^T \tilde{\mathbf{V}}_T\|_F^2$  would equal 0. To obtain an approximation of  $\epsilon$ , the number of iterations required by Oja's method naturally depends inversely on  $\epsilon$ , the failure probability  $\delta$ , and the gap between eigenvalues  $\lambda_k(\mathbf{M})$  and  $\lambda_{k+1}(\mathbf{M})$ .

### 3.2 Distributed Oja's method via random edge sampling

Oja's method can be implemented in the asynchronous communication model ([Definition 2.1](#)) to compute top eigenvectors of the communication matrix  $\mathbf{D} + \mathbf{W}$ , defined in [Section 2.2](#).

For any pair of nodes  $(u, v)$ , let  $\mathbf{e}_{u,v} = \mathbf{e}_u + \mathbf{e}_v$  be the vector with all zero entries except 1's in its  $u^{\text{th}}$  and  $v^{\text{th}}$  positions. Given weight function  $w$  and associated matrix  $\mathbf{W}$ , let  $\mathcal{D}_{\mathbf{W}}$  be the distribution in which each  $\mathbf{e}_{u,v}$  is selected with probability  $\mathbf{W}_{u,v}$ . That is, the same distribution by which edges are selected to be active by the scheduler in [Definition 2.1](#). Noting that  $\mathbf{e}_{u,v} \mathbf{e}_{u,v}^T$  is all zero except at its  $(u, u)$ ,  $(v, v)$ ,  $(u, v)$ , and  $(v, u)$  entries, we can see that

$$\mathbb{E}_{\mathbf{e}_{u,v} \sim \mathcal{D}_{\mathbf{W}}} [\mathbf{e}_{u,v} \mathbf{e}_{u,v}^T] = \sum_{(u,v) \in \mathcal{P}} \mathbf{W}_{u,v} \cdot \mathbf{e}_{u,v} \mathbf{e}_{u,v}^T = \mathbf{D} + \mathbf{W}, \quad (1)$$

where  $\mathcal{P}$  denotes the set of unordered node pairs  $(u, v)$  with  $u \neq v$ . So if we run Oja's algorithm with  $\mathbf{e}_{u,v}$  sampled according to  $\mathcal{D}_{\mathbf{W}}$ , we will obtain an approximation to the top eigenvectors of  $\mathbf{D} + \mathbf{W}$ . Note that this matrix is PSD, by the fact that each  $\mathbf{e}_{u,v} \mathbf{e}_{u,v}^T$  is PSD.

Furthermore, the algorithm can be implemented in our communication model as an *extremely simple averaging protocol*. Each iteration of [Algorithm 1](#) requires computing  $\mathbf{Q}_{t+1} = (\mathbf{I} + \eta \mathbf{x}_t \mathbf{x}_t^T) \mathbf{Q}_t$ . If  $\mathbf{x}_t = \mathbf{e}_{u,v}$  for  $\mathbf{e}_{u,v} \sim \mathcal{D}_{\mathbf{W}}$ , we can see that computing  $\mathbf{Q}_{t+1}$  just requires updating the  $u^{\text{th}}$  and  $v^{\text{th}}$  rows of  $\mathbf{Q}_t$ . Thus, if the  $n$  rows of  $\mathbf{Q}_t$  are distributed across  $n$  nodes, this update can be done locally by nodes  $u$  and  $v$  when they are chosen to interact by the randomized scheduler. Specifically, letting  $[q_u^{(1)}, \dots, q_u^{(k)}]$  be the  $u^{\text{th}}$  row of  $\mathbf{Q}_t$ , stored as the state at node  $u$ , applying  $(\mathbf{I} + \eta \mathbf{e}_{u,v} \mathbf{e}_{u,v}^T)$  just requires setting for all  $i \in [k]$ :

$$q_u^{(i)} := (1 + \eta) q_u^{(i)} + \eta q_v^{(i)}. \quad (2)$$

Node  $v$  makes a symmetric update, and all other entries of  $\mathbf{Q}_t$  remain fixed.

We give the pseudocode for this protocol in [Algorithm 2](#). Along with the main iteration based on the simple update in (2), the nodes need to implement Step 5 of [Algorithm 1](#), where  $\mathbf{Q}_T$  is orthogonalized. This can be done with a gossip-based protocol, which we abstract as the routine `AsynchOrth`. We give an implementation of `AsynchOrth` in [Section 3.3](#).

**Remark 3.2** (Choice of communication matrix). *While, as we will show, the eigenvectors of  $\mathbf{D} + \mathbf{W}$  are naturally useful in our applications to community detection, the above techniques easily extend to computing eigenvectors of other matrices. For example, if we set  $\mathbf{e}_{u,v} = \mathbf{e}_u - \mathbf{e}_v$ ,  $\mathbb{E}_{\mathbf{e}_{u,v} \sim \mathcal{D}_{\mathbf{W}}} [\mathbf{e}_{u,v} \mathbf{e}_{u,v}^T] = \mathbf{D} - \mathbf{W} = \mathbf{L}$ , a scaled Laplacian of the communication graph.*

---

**Algorithm 2** ASYNCHRONOUS OJA'S ( $\text{AsynchOja}(T, T', \eta)$ )

---

**Input:** Time bounds  $T, T'$ , step size  $\eta$ .

**Initialization:**  $\forall u$ , chose  $[q_u^{(1)}, \dots, q_u^{(k)}]$  independently from standard Gaussian  $\mathcal{N}(0, 1)$ .

- 1: **if**  $t < T$  **then**
  - 2:      $(u, v)$  is chosen by the randomized scheduler.
  - 3:     For all  $i \in [k]$ ,  $q_u^{(i)} := (1 + \eta)q_u^{(i)} + \eta q_v^{(i)}$ .  $\triangleright$  Computes of  $(\mathbf{I} + \eta \mathbf{e}_{u,v} \mathbf{e}_{u,v}^T) \mathbf{Q}_t$ .
  - 4: **else**
  - 5:      $[\hat{v}_u^{(1)}, \dots, \hat{v}_u^{(k)}] = \text{AsynchOrth}([q_u^{(1)}, \dots, q_u^{(k)}], T')$ .  $\triangleright$  Implements of  $\tilde{\mathbf{V}}_T = \text{orth}(\mathbf{Q}_T)$ .
  - 6: **end if**
- 

Note that in the pseudocode above, when nodes  $u, v$  interact in the asynchronous model, they only need to share their respective values of  $q_u^{(i)}$  and  $q_v^{(i)}$  for  $i \in [k]$ .

Up to the orthogonalization step, we see that [Algorithm 2](#) *exactly simulates* [Algorithm 1](#) on input  $\mathbf{M} = \mathbf{D} + \mathbf{W}$ . Thus, assuming that  $\text{AsynchOrth}([q_u^{(1)}, \dots, q_u^{(k)}])$  exactly computes  $\tilde{\mathbf{V}}_T = \text{orth}(\mathbf{Q}_T)$  as in Step 5 of [Algorithm 1](#), the error bound of [Theorem 3.1](#) applies directly. Specifically, if we let the local states,  $[q_1^{(1)}, \dots, q_n^{(1)}], \dots, [q_1^{(k)}, \dots, q_n^{(k)}]$  correspond to the  $k$  length- $n$  vectors in  $\tilde{\mathbf{V}}_T$ , [Theorem 3.1](#) shows that  $\|\mathbf{Z}^T \tilde{\mathbf{V}}_T\|_F^2 \leq \epsilon$ . In [Section 3.3](#) we show that this bound still holds when  $\text{AsynchOrth}$  computes an approximate orthogonalization.

### 3.3 Distributed orthogonalization and eigenvector guarantees

In fact, a specific orthogonalization strategy yields a stronger bound, which is desirable in many applications, including community detection: [Algorithm 2](#) can actually well approximate *each* of  $\mathbf{D} + \mathbf{W}$ 's top  $k$  eigenvectors, instead of just the subspace they span.

Specifically, let  $\tilde{\mathbf{v}}_i$  denote the  $i^{\text{th}}$  column of  $\tilde{\mathbf{V}}_T$  and  $\mathbf{v}_i$  denote the  $i^{\text{th}}$  eigenvector of  $\mathbf{D} + \mathbf{W}$ . We want  $(\tilde{\mathbf{v}}_i^T \mathbf{v}_i)^2 \geq 1 - \epsilon$  for all  $i$ . Such a guarantee requires sufficiently large gaps between the top  $k$  eigenvalues, so that their corresponding eigenvectors are identifiable. If these gaps exist, the guarantee can be using the following orthogonalization procedure:

---

**Algorithm 3** ORTHOGONALIZATION VIA CHOLESKY FACTORIZATION (CENTRALIZED)

---

**Input:**  $\mathbf{Q} \in \mathbb{R}^{n \times k}$  with full column rank. **Output:** Orthonormal span for  $\mathbf{Q}$ ,  $\tilde{\mathbf{V}} \in \mathbb{R}^{n \times k}$ .

- 1:  $\mathbf{L} := \text{chol}(\mathbf{Q}^T \mathbf{Q})$   $\triangleright$  Cholesky decomp. returns lower triangular  $\mathbf{L}$  with  $\mathbf{L}\mathbf{L}^T = \mathbf{Q}^T \mathbf{Q}$ .
  - 2: **return**  $\tilde{\mathbf{V}} := \mathbf{Q}(\mathbf{L}^T)^{-1}$   $\triangleright$  Orthonormalize  $\mathbf{Q}_T$ 's columns using the Cholesky factor.
- 

**Remark 3.3.** [Algorithm 3](#) requires an input that is full-rank, which always includes  $\mathbf{Q}_T$  in [Algorithms 1](#) and [2](#):  $\mathbf{Q}_0$ 's entries are random Gaussians so it is full-rank with probability 1 and each  $(\mathbf{I} + \eta \mathbf{x}_t^T \mathbf{x}_t)$  is full-rank since  $\eta < \|\mathbf{x}_t\|$ . Thus,  $\mathbf{Q}_T = \prod_{t=0}^{T-1} (\mathbf{I} + \eta \mathbf{x}_t^T \mathbf{x}_t) \mathbf{Q}_0$  is too.

Ultimately, our  $\text{AsynchOrth}$  is an asynchronous distributed implementation of [Algorithm 3](#). We first prove an eigenvector approximation bound under the assumption that this implementation is exact ([Corollary C.1](#) in [Appendix C](#)) and then adapt that result to account for the fact that  $\text{AsynchOrth}$  only outputs an approximate solution.

Pseudocode for  $\text{AsynchOrth}$  is included below. Each node first computes a (scaled) approximation to every entry of  $\mathbf{Q}^T \mathbf{Q}$  using a simple averaging technique. Nodes then locally compute

$\mathbf{L} = \text{chol}(\mathbf{Q}^T \mathbf{Q})$  and the  $u^{\text{th}}$  row of  $\tilde{\mathbf{V}}_T = \mathbf{Q}(\mathbf{L}^T)^{-1}$ . In [Appendix C](#) we argue that, due to numerical stability of Cholesky decomposition, each node's output is close to the  $u^{\text{th}}$  row of an exactly computed  $\tilde{\mathbf{V}}_T$ , despite the error in constructing  $\mathbf{Q}^T \mathbf{Q}$ .

---

**Algorithm 4** ASYNCHRONOUS CHOLESKY ORTHOGONALIZATION (`AsynchOrth`( $T$ ))

---

**Input:** Time bound  $T$ .

**Initialization:** Each node holds  $[q_u^{(1)}, \dots, q_u^{(k)}]$ . For all  $i, j \in [k]$ , let  $r_u^{(i,j)} := q_u^{(i)} \cdot q_u^{(j)}$ .

- 1: **if**  $t < T$  **then**
  - 2:    $(u, v)$  is chosen by the randomized scheduler.
  - 3:   for all  $i, j \in [k]$ ,  $r_u^{(i,j)} := \frac{r_u^{(i,j)} + r_v^{(i,j)}}{2}$ . ▷ Estimation of  $\frac{1}{n} \mathbf{q}_i^T \mathbf{q}_j$  via averaging.
  - 4: **else**
  - 5:   Form  $\mathbf{R}_u \in \mathbb{R}^{k \times k}$  with  $(\mathbf{R}_u)_{i,j} = (\mathbf{R}_u)_{j,i} := n \cdot r_u^{(i,j)}$ . ▷ Approximation of  $\mathbf{Q}^T \mathbf{Q}$ .
  - 6:    $\mathbf{L}_u := \text{chol}(\mathbf{R}_u)$ .
  - 7:    $[\hat{v}_u^{(1)}, \dots, \hat{v}_u^{(k)}] := [q_u^{(1)}, \dots, q_u^{(k)}] \cdot (\mathbf{L}_u^T)^{-1}$ . ▷ Approximation of  $u^{\text{th}}$  row of  $\mathbf{Q}(\mathbf{L}_u^T)^{-1}$ .
  - 8: **end if**
- 

In [Appendix C](#) we prove the following result when [Algorithm 4](#) is used to implement `AsynchOrth` as a subroutine for [Algorithm 2](#), `AsynchOja`( $T, T', \eta$ ):

**Theorem 3.4** (Asynchronous eigenvector approximation). *Let  $\mathbf{v}_1, \dots, \mathbf{v}_k$  be the top  $k$  eigenvectors of the communication matrix  $\mathbf{D} + \mathbf{W}$  in an asynchronous communication model, and let  $\Lambda, \overline{\text{gap}}, \gamma_{\text{mix}}$  be bounds satisfying:  $\Lambda \geq \sum_{j=1}^k \lambda_j(\mathbf{D} + \mathbf{W})$ ,  $\overline{\text{gap}} \leq \min_{j \in [k]} [\lambda_j(\mathbf{D} + \mathbf{W}) - \lambda_{j+1}(\mathbf{D} + \mathbf{W})]$ , and  $\gamma_{\text{mix}} \leq \min [\frac{1}{n}, \log(\lambda_2^{-1}(\mathbf{I} - \frac{1}{2}\mathbf{D} + \frac{1}{2}\mathbf{W}))]$ .*

*For any  $\epsilon, \delta \in (0, 1)$ , let  $\xi = \frac{n}{\delta \epsilon \overline{\text{gap}}}$ . Let  $\eta = \frac{c_1 \epsilon^2 \overline{\text{gap}} \cdot \delta^2}{\Lambda k^3 \log^3 \xi}$  for sufficiently small  $c_1$ , and  $T = \frac{c_2 \cdot (\log \xi + 1/\epsilon)}{\overline{\text{gap}} \cdot \eta}$ ,  $T' = \frac{c_3 (\log \xi + 1/\epsilon) \cdot \lambda_1(\mathbf{D} + \mathbf{W})}{\overline{\text{gap}} \cdot \gamma_{\text{mix}}}$  for sufficiently large  $c_2, c_3$ . For all  $u \in [n], i \in [k]$ , let  $\hat{v}_u^{(j)}$  be the local state computed by [Algorithm 2](#). If  $\hat{\mathbf{V}} \in \mathbb{R}^{n \times k}$  is given by  $(\hat{\mathbf{V}})_{u,j} = \hat{v}_u^{(j)}$  and  $\hat{\mathbf{v}}_i$  is the  $i^{\text{th}}$  column of  $\hat{\mathbf{V}}$ , then with probability  $\geq 1 - \delta - e^{-\Theta(n)}$ , for all  $i \in [k]$ :*

$$|\hat{\mathbf{v}}_i^T \mathbf{v}_i| \geq 1 - \epsilon \quad \text{and} \quad \|\hat{\mathbf{v}}_i\|_2 \leq 1 + \epsilon.$$

## 4 Distributed community detection

From the results of [Section 3](#), we obtain a simple population protocol for distributed community detection that works for many clustered communication models, including the  $(n, p, q)$ -weighted communication and  $G(n, p, q)$ -communication models of [Definitions 2.4](#) and [2.5](#).

In particular, we show that if each node  $u \in V$  can locally compute the  $u^{\text{th}}$  entry of an approximation  $\hat{\mathbf{v}}_2$  to the second eigenvector of the communication matrix  $\mathbf{D} + \mathbf{W}$ , then it can solve the community detection problem locally:  $u$  just sets its state to the sign of this entry.

---

**Algorithm 5** ASYNCHRONOUS COMMUNITY DETECTION ( $\text{AsynchCD}(T, T', \eta)$ )

---

**Input:** Time bounds  $T, T'$ , step size  $\eta$ .

- 1: Run  $\text{AsynchOja}(T, T', \eta)$  (Algorithm 2) with  $k = 2$ .
  - 2: Set  $\hat{\chi}_u := \text{sign}(\hat{v}_u^{(2)})$ .
- 

Here  $\hat{\chi}_u \in \{-1, 1\}$  is the final state of node  $u$ . We will claim that this state solves the community detection problem of Definition 2.6. We use the notation  $\hat{\chi}_u$  because we will use  $\chi$  to denote the true *cluster indicator vector* for communities  $V_1$  and  $V_2$  in a given communication model:  $\chi_u = 1$  for  $u \in V_1$  and  $\chi_u = -1$  for  $u \in V_2$ .

In particular, we will show that if  $\eta$  is set so that  $\text{AsynchOja}$  outputs eigenvectors with accuracy  $\epsilon$ , then a  $1 - O(\epsilon)$  fraction of nodes will correctly identify their clusters. In Section 5 we show how to implement a ‘cleanup phase’ where, starting with  $\epsilon$  set to a small constant (e.g.  $\epsilon = .1$ ), the nodes can converge to a state with *all* cluster labels correct with high probability.

#### 4.1 Community detection in the $(n, p, q)$ -weighted communication model

We start with an analysis for the  $(n, p, q)$ -weighted communication model. Recall that in this model the nodes are partitioned into two sets,  $V_1$  and  $V_2$ , each with  $n/2$  elements. Without loss of generality we can identify the nodes with integer labels such that  $1, \dots, n/2 \in V_1$  and  $n/2 + 1, \dots, n \in V_2$ . We define the weighted cluster indicator matrix,  $\mathbf{C}^{(p,q)} \in \mathbb{R}^{n \times n}$ :

$$\mathbf{C}^{(p,q)} \stackrel{\text{def}}{=} \begin{bmatrix} p \cdot \mathbf{1}_{\frac{n}{2} \times \frac{n}{2}} & q \cdot \mathbf{1}_{\frac{n}{2} \times \frac{n}{2}} \\ q \cdot \mathbf{1}_{\frac{n}{2} \times \frac{n}{2}} & p \cdot \mathbf{1}_{\frac{n}{2} \times \frac{n}{2}} \end{bmatrix}. \quad (3)$$

$p$  and  $q$  can be arbitrary, but we will always take  $p > q > 0$ . It is easy to check that  $\mathbf{C}^{(p,q)}$  is a rank two matrix with eigendecomposition:

$$\mathbf{C}^{(p,q)} = \frac{n}{2} \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} \begin{bmatrix} p+q & 0 \\ 0 & p-q \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \end{bmatrix} \quad \text{where} \quad \mathbf{v}_1 = \frac{\mathbf{1}_{n \times 1}}{\sqrt{n}}, \quad \mathbf{v}_2 = \frac{\chi}{\sqrt{n}}. \quad (4)$$

So, if all nodes could compute their corresponding entry in the *second eigenvector* of  $\mathbf{C}^{(p,q)}$ , then by simply returning the sign of this entry, they would solve the distributed community detection problem (Definition 2.6). If they compute this eigenvector approximately, then we can still show that a large fraction of them correctly solve community detection. Specifically:

**Lemma 4.1.** *Let  $\mathbf{v}_2$  be the second eigenvector of  $\mathbf{C}^{(p,q)}$  for any  $p > q > 0$ . If  $\tilde{\mathbf{v}}_2$  satisfies:*

$$|\tilde{\mathbf{v}}_2^T \mathbf{v}_2| \geq 1 - \epsilon \quad \text{and} \quad \|\tilde{\mathbf{v}}_2\|_2 \leq 1 + \epsilon. \quad (5)$$

*for  $\epsilon \leq 1$ , then  $\text{sign}(\tilde{\mathbf{v}}_2)$  gives a labeling such that, after ignoring at most  $5\epsilon n$  nodes, all remaining nodes in  $V_1$  have the same labeling, and all in  $V_2$  have the opposite.*

*Proof.* We follow the argument from [34]. Let  $\|\mathbf{y}\|_0$  denote the number of non-zeros in a vector  $\mathbf{y}$ . Since  $p > q > 0$ ,  $\mathbf{v}_2 = \frac{\chi}{\sqrt{n}}$  by (4), so the number of nodes misclassified by  $\text{sign}(\tilde{\mathbf{v}}_2)$  is:

$$\min_{s \in \{-1, 1\}} \|\text{sign}(s \cdot \tilde{\mathbf{v}}_2) - \text{sign}(\mathbf{v}_2)\|_0.$$

For  $s \in \{-1, 1\}$ , if  $\text{sign}(s \cdot \tilde{\mathbf{v}}_2)$  and  $\text{sign}(\mathbf{v}_2)$  differ on a specific coordinate, then since each entry of  $\mathbf{v}_2$  has value  $\pm \frac{1}{\sqrt{n}}$ ,  $s \cdot \tilde{\mathbf{v}}_2$  and  $\mathbf{v}_2$  must differ by at least  $\frac{1}{\sqrt{n}}$  on that coordinate. It follows that  $\|\tilde{\mathbf{v}}_2 - s \cdot \mathbf{v}_2\|_2^2 \geq \frac{1}{n} \|\text{sign}(s \cdot \tilde{\mathbf{v}}_2) - \text{sign}(\mathbf{v}_2)\|_0$  and hence;

$$\min_{s \in \{-1, 1\}} \|\text{sign}(s \cdot \tilde{\mathbf{v}}_2) - \text{sign}(\mathbf{v}_2)\|_0 \leq n \cdot \min_{s \in \{-1, 1\}} \|s \cdot \tilde{\mathbf{v}}_2 - \mathbf{v}_2\|_2^2. \quad (6)$$

We can bound the righthand side of (6) using (5). Specifically, assuming  $\epsilon \leq 1$ :

$$\begin{aligned} n \cdot \min_{s \in \{-1, 1\}} \|s \cdot \tilde{\mathbf{v}}_2 - \mathbf{v}_2\|_2^2 &= n \cdot \min_{s \in \{-1, 1\}} (\|\mathbf{v}_2\|_2^2 + \|\tilde{\mathbf{v}}\|_2^2 - 2(\tilde{\mathbf{v}}_i^T \mathbf{v}_2)) \\ &\leq n \cdot (1 + (1 + \epsilon)^2 - 2(1 - \epsilon)) \leq 5n\epsilon \end{aligned}$$

Plugging back into (6), we have  $\min_{s \in \{-1, 1\}} \|\text{sign}(s \cdot \tilde{\mathbf{v}}_2) - \text{sign}(\mathbf{v}_2)\|_0 \leq 5n\epsilon$ , so  $\text{sign}(\tilde{\mathbf{v}}_2)$  only classifies at most a  $5\epsilon$  fraction of nodes incorrectly, giving the lemma.  $\square$

With Lemma 4.1 in place, we can then apply Theorem 3.4 to prove the correctness of **AsynchCD** (Algorithm 5) for the  $(n, p, q)$ -weighted communication model

**Theorem 4.2** ( $\epsilon$ -approximate community detection:  $(n, p, q)$ -weighted communication model). *Consider Algorithm 5 in the  $(n, p, q)$ -weighted communication model. Let  $\rho = \min\left(\frac{q}{p+q}, \frac{p-q}{p+q}\right)$ . For sufficiently small constant  $c_1$  and sufficiently large  $c_2$  and  $c_3$ , let*

$$\eta = \frac{c_1 \epsilon^2 \delta^2 \rho}{\log^3\left(\frac{n}{\epsilon \delta \rho}\right)}, \quad T = \frac{c_2 n \left(\log^3\left(\frac{n}{\epsilon \delta \rho}\right) + \frac{\log\left(\frac{n}{\epsilon \delta \rho}\right)}{\epsilon}\right)}{\epsilon^2 \delta^2 \rho^2}, \quad T' = \frac{c_3 n \left(\log\left(\frac{n}{\epsilon \delta \rho}\right) + \frac{1}{\epsilon}\right)}{\rho^2}.$$

*With probability  $1 - \delta$ , after ignoring  $\epsilon n$  nodes, all remaining nodes in  $V_1$  terminate in some state  $s_1 \in \{-1, 1\}$ , and all nodes in  $V_2$  terminate in state  $s_2 = -s_1$ . Suppressing polylogarithmic factors in the parameters, the total number of global rounds and local rounds required are:  $T + T' = \tilde{O}\left(\frac{n}{\epsilon^3 \delta^2 \rho^2}\right)$  and  $L = \tilde{O}\left(\frac{1}{\epsilon^3 \delta^3 \rho^2}\right)$ .*

*Proof.* In the  $(n, p, q)$ -weighted communication model the weight and degree matrices are:

$$\mathbf{W} = \frac{4}{n^2(p+q) - 2np} \cdot (\mathbf{C}^{(p,q)} - p \cdot \mathbf{I}_{n \times n}) \quad \text{and} \quad \mathbf{D} = \frac{2}{n} \cdot \mathbf{I}_{n \times n}.$$

Thus, referring to the eigendecomposition of  $\mathbf{C}^{(p,q)}$  shown in (4), the top eigenvector of  $\mathbf{D} + \mathbf{W}$  is  $\mathbf{v}_1 = \mathbf{1}_{n \times 1} / \sqrt{n}$  with corresponding eigenvalue:  $\lambda_1 = \frac{4}{n^2(p+q) - 2np} \cdot \left(\frac{n(p+q)}{2} - p\right) + \frac{2}{n} = \frac{4}{n}$ . The second eigenvector is the scaled cluster indicator vector  $\mathbf{v}_2 = \boldsymbol{\chi} / \sqrt{n}$  with eigenvalue

$$\lambda_2 = \frac{4}{n^2(p+q) - 2np} \cdot \left(\frac{n(p-q)}{2} - p\right) + \frac{2}{n} = \frac{4}{n} \cdot \frac{p}{p + \frac{n}{n-2} \cdot q}.$$

Finally, for all remaining eigenvalues of  $\mathbf{D} + \mathbf{W}$ ,  $\{\lambda_3, \dots, \lambda_n\}$ ,  $\lambda_i = \frac{2}{n} - \frac{4p}{n^2(p+q)-2np}$ . We can bound the eigenvalue gaps:

$$\lambda_1 - \lambda_2 \geq \frac{4}{n} - \frac{4}{n} \cdot \frac{p}{p+q} = \frac{4q}{n(p+q)} \quad \lambda_2 - \lambda_3 = \frac{2(p-q)}{n(p+q)-2p} \geq \frac{2(p-q)}{n(p+q)}$$

Let  $\rho = \min\left(\frac{q}{p+q}, \frac{p-q}{p+q}\right)$ . We bound the mixing time of  $\mathbf{W} + \mathbf{D}$  by noting that  $\lambda_2(\mathbf{I} - 1/2\mathbf{D} + 1/2\mathbf{W}) \leq 1 - \frac{2q}{n(p+q)}$ . Then using that  $\log(1/x) \geq 1 - x$  for all  $x \in (0, 1]$ ,  $\log(\lambda_2^{-1}(\mathbf{I} - 1/2\mathbf{D} + 1/2\mathbf{W})) \geq \frac{2q}{n(p+q)} \geq \frac{2\rho}{n}$ . We then apply [Theorem 3.4](#) with  $k = 2$ ,  $\Lambda = \frac{4}{n} + \frac{4}{n} \frac{p}{p + \frac{n}{n-2}q} \leq \frac{8}{n}$ ,  $\overline{\text{gap}} = \frac{4}{n} \cdot \min\left(\frac{q}{p+q}, \frac{p-q}{2(p+q)}\right) \geq \frac{2\rho}{n}$ , and  $\gamma_{mix} = \frac{2\rho}{n}$ . With these parameters we set, for sufficiently small  $c_1$  and large  $c_2, c_3$ ,

$$\eta = \frac{c_1 \epsilon^2 \delta^2 \cdot \rho}{\log^3\left(\frac{n}{\epsilon \delta \rho}\right)}, \quad T = \frac{c_2 \cdot n \cdot \left(\log^3\left(\frac{n}{\epsilon \delta \rho}\right) + \frac{\log\left(\frac{n}{\epsilon \delta \rho}\right)}{\epsilon}\right)}{\epsilon^2 \delta^2 \rho^2}, \quad T' = \frac{c_3 \cdot n \cdot \left(\log\left(\frac{n}{\epsilon \delta \rho}\right) + \frac{1}{\epsilon}\right)}{\rho^2}$$

where to bound  $T'$  we use that  $\frac{\lambda_1(\mathbf{D} + \mathbf{W})}{\overline{\text{gap}}} \leq \frac{2}{\rho}$ . Let  $\hat{\mathbf{V}} \in \mathbb{R}^{n \times k}$  be given by  $(\hat{\mathbf{V}})_{u,j} = \hat{v}_u^{(j)}$  where  $\hat{v}_u^{(j)}$  are the states of `Asynch0ja`( $T, T', \eta$ ) and let  $\hat{\mathbf{v}}_2$  be the second column of  $\hat{\mathbf{V}}$ . With these parameters, [Theorem 3.4](#) gives with probability  $\geq 1 - \delta$  that  $|\hat{\mathbf{v}}_2^T \mathbf{v}_2| \geq 1 - \epsilon$  and  $\|\hat{\mathbf{v}}_2\|_2 \leq 1 + \epsilon$ .

Applying [Lemma 4.1](#) then gives the theorem if we adjust  $\epsilon$  by a factor of  $1/5$ . Recall that the second eigenvector of  $\mathbf{D} + \mathbf{W}$  is identical to that of  $\mathbf{C}^{(p,q)}$ . Additionally, in expectation, each node is involved in  $L = \frac{2(T+T')}{n}$  interactions. This bound holds for all nodes within a factor 2 with probability  $1 - \delta$  by a Chernoff bound, since  $L = \Omega(\log(n/\delta))$ . We can union bound over our two failure probabilities and adjust  $\delta$  by  $1/2$  to obtain overall failure probability  $\leq \delta$ .  $\square$

## 4.2 Community Detection in the $G(n, p, q)$ -communication model

In the  $G(n, p, q)$ -communication model, nodes communicate using a random graph which is equal to the communication graph in the  $(n, p, q)$ -weighted communication model *in expectation*. Using an approach similar to [\[34\]](#), which simplifies the perturbation method used in [\[24\]](#), we can prove that in the  $G(n, p, q)$ -communication model  $\mathbf{W}$  is a small perturbation of  $\mathbf{C}^{(p,q)}$  and so the second eigenvector of  $\mathbf{D} + \mathbf{W}$  approximates that of  $\mathbf{C}^{(p,q)}$  – i.e., the cluster indicator vector  $\chi$ . We defer this analysis to [Appendix D](#), stating the main result here:

**Theorem 4.3** ( $\epsilon$ -approximate community detection:  $G(n, p, q)$ -communication model). *Consider [Algorithm 5](#) in the  $G(n, p, q)$ -communication model. Let  $\rho = \min\left(\frac{q}{p+q}, \frac{p-q}{p+q}\right)$ . For sufficiently small constant  $c_1$  and sufficiently large  $c_2$  and  $c_3$  let*

$$\eta = \frac{c_1 \epsilon^2 \delta^2 \rho}{\log^3\left(\frac{n}{\epsilon \delta \rho}\right)}, \quad T = \frac{c_2 n \left(\log^3\left(\frac{n}{\epsilon \delta \rho}\right) + \frac{\log\left(\frac{n}{\epsilon \delta \rho}\right)}{\epsilon}\right)}{\epsilon^2 \delta^2 \rho^2}, \quad T' = \frac{c_3 n \left(\log\left(\frac{n}{\epsilon \delta \rho}\right) + \frac{1}{\epsilon}\right)}{\rho^2}.$$

If  $\frac{\min[q, p-q]}{\sqrt{p+q}} \geq \frac{c_4 \sqrt{\log(n/\delta)}}{\epsilon \sqrt{n}}$  for large enough constant  $c_4$ , then, with probability  $1 - \delta$ , after ignoring  $\epsilon n$  nodes, all remaining nodes in  $V_1$  terminate in some state  $s_1 \in \{-1, 1\}$ , and all nodes in  $V_2$  terminate in state  $s_2 = -s_1$ . Suppressing polylogarithmic factors, the total number of global rounds and local rounds required are:  $T + T' = \tilde{O}\left(\frac{n}{\epsilon^3 \delta^2 \rho^2}\right)$  and  $L = \tilde{O}\left(\frac{1}{\epsilon^3 \delta^3 \rho^2}\right)$ .

If for example,  $p, q = \Theta(1)$  and thus the  $G(n, p, q)$  graph is dense, we can recover the communities with probability  $1 - \delta$  up to  $O(1)$  error as long as  $q \leq p - c\sqrt{\log(n/\delta)/n}$  for sufficiently large constant  $c$ . Alternatively, if  $p, q = \Theta(\log(n/\delta)/n)$ , so the  $G(n, p, q)$  graph is sparse, we require  $q \leq cp$  for sufficiently small  $c$ .

## 5 Cleanup Phase

After we apply [Theorem 4.3](#) (respectively, [Theorem 4.2](#)) an  $\epsilon$ -fraction of nodes are incorrectly clustered. The goal of this section is to provide a simple algorithm that improves this clustering so that *all nodes* are labeled correctly after a small number of rounds.

For the  $(n, p, q)$ -weighted communication model, doing so is straightforward. After running [Algorithm 2](#) and selecting a label, each time a node communicates in the future it records the chosen label of the node it communicates with. Ultimately, it changes its label to the majority of labels encountered. If  $\epsilon$  is small enough so  $p(1 - \epsilon) > q + \epsilon p$ , this majority tends towards the node's correct label. The number of required rounds for the majority to be correct, with good probability for all nodes, is a simple a function of  $p, q$ , and  $\epsilon$ .

The  $G(n, p, q)$ -communication model is more difficult. [Theorem 4.3](#) does not guarantee how incorrectly labeled nodes are distributed: it is possible that a majority of a node's neighbors fall into the set of  $\epsilon n$  “bad nodes”. In that case, even after infinitely many rounds of communication, the majority label encountered will not tend towards the node's correct identity.

As a remedy, we introduce a phased algorithm ([Algorithm 6](#)) where each node updates its label to the majority of labels seen during a phase. We show that in each phase the fraction of incorrectly labeled nodes decreases by a constant factor. Our analysis establishes a graph theoretic bound on the external edge density of most subsets of nodes. Specifically, for all subsets  $S$  below a certain size, we show that, with high probability, there are at most  $|S|/3$  nodes which have enough connections to  $S$  so that if an adversary gave all nodes in  $S$  incorrect labels, it could cause these nodes to have an incorrect majority label. This bound guarantees that at most  $|S|/3$  bad labels ‘propagate’ to the next phase of the algorithm.

We analyze [Algorithm 6](#) in the  $(n, p, q)$ -weighted communication model in [Appendix E](#) (in this case we just set  $k = 1$ ) and  $G(n, p, q)$ -communication model in [Appendix F](#).

---

**Algorithm 6** CLEANUP PHASE (pseudocode for node  $u$ )

---

**Input:** Number of phases  $k$  and number of rounds per phase  $r$ .

**Output:** Label  $\hat{\chi}_u \in \{-1, 1\}$

---

```

1: for Phase 1 to  $k$  do
2:   for Round  $i = 1$  to  $r$  do
3:      $S_i := \hat{\chi}_v$ , where  $\hat{\chi}_v$  denotes the  $i^{\text{th}}$  sample of node  $u$ .
4:   end for
5:    $\hat{\chi}_u := 1$  if  $\sum_i^r S_i \geq 0$ ,  $\hat{\chi}_u := -1$  otherwise.
6: end for

```

---

**Theorem 5.1.** Consider the  $(n, p, q)$ -weighted communication model. Assume that a fraction of at most  $\epsilon \leq 1/64$  of the nodes are incorrectly clustered after [Algorithm 2](#). As long as  $p' = (1 - \epsilon)p$  and  $q' = q + \epsilon p$  satisfy  $p' > q'$ , [Algorithm 6](#) ensures that all nodes are correctly labeled with high probability after  $O(\frac{p \ln n}{(\sqrt{p'} - \sqrt{q'})^2})$  local rounds. In particular, for  $q \leq p/2$  and  $\epsilon < 1/8$ , the number of local rounds required is  $O(\log n)$ .

**Theorem 5.2.** Consider the  $G(n, p, q)$ -communication model. Let  $\Delta = \frac{p}{2} - \frac{q}{2} - \sqrt{12p \ln n/n} - \sqrt{12q \ln n/n}$ . Assume that  $\Delta = \Omega(\ln n/n)$  and at most  $\epsilon \leq \Delta/24p$  nodes are incorrectly clustered after [Algorithm 2](#). As long as  $p'' = \frac{p}{2} - \sqrt{\frac{6p \ln n}{n}} - \frac{\Delta}{12}$  and  $q'' = \frac{q}{2} + \sqrt{\frac{6q \ln n}{n}} + \frac{\Delta}{12}$  satisfy  $p'' > q''$ , [Algorithm 6](#) ensures that all nodes are correctly labeled with high probability after  $O(\frac{p \ln^2 n}{(\sqrt{p''} - \sqrt{q''})^2})$  local rounds. In particular, for  $q \leq p/2$  the number of local rounds required is  $O(\log^2 n)$ .

Note that if  $p - q = \Omega(\sqrt{\log n/n})$ , then  $\Delta$  simplifies to  $\Delta = \Theta(p - q)$ . Incidentally,  $p - q = \Omega(\sqrt{\log n/n})$  is sometimes tight because, in this regime, clustering correctly can be infeasible: some nodes will simply have more neighbors in the opposite cluster. Consider for example when  $p = 1/2 + \sqrt{\ln n/(10n)}$  and  $q = 1/2$ .

## References

- [1] Zeyuan Allen-Zhu and Yuanzhi Li. First efficient convergence for streaming k-pca: a global, gap-free, and near-optimal rate. pages 487–492, 2017.
- [2] James Aspnes and Eric Ruppert. An introduction to population protocols. In *Middleware for Network Eccentric and Mobile Applications*, pages 97–120. Springer, 2009.
- [3] Luca Becchetti, Andrea Clementi, , Pasin Manurangsi, Emanuele Natale, Francesco Pasquale, Prasad Raghavendra, and Luca Trevisan. Average whenever you meet: Opportunistic protocols for community detection. [arXiv:1703.05045](#), 2017.
- [4] Luca Becchetti, Andrea Clementi, Emanuele Natale, Francesco Pasquale, and Luca Trevisan. Find your place: Simple distributed algorithms for community detection. In *Proceedings of the 28th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 940–959, 2017.

- [5] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in neural information processing systems*, pages 585–591, 2002.
- [6] Phillip Bonacich. Some unique properties of eigenvector centrality. *Social networks*, 29(4):555–564, 2007.
- [7] Stephen Boyd, Arpita Ghosh, Balaji Prabhakar, and Devavrat Shah. Gossip algorithms: Design, analysis and applications. In *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, volume 3, pages 1653–1664. IEEE, 2005.
- [8] Stephen Boyd, Arpita Ghosh, Balaji Prabhakar, and Devavrat Shah. Randomized gossip algorithms. *IEEE/ACM Transactions on Networking (TON)*, 14(SI):2508–2530, 2006.
- [9] Shakila Bu-Pasha. Cross-border issues under eu data protection law with regards to personal data protection. 26:1–16, 05 2017.
- [10] Sam Cole, Shmuel Friedland, and Lev Reyzin. A simple spectral algorithm for recovering planted partitions. *Special Matrices*, 5:139–157, 2017.
- [11] Christopher M De Sa, Ce Zhang, Kunle Olukotun, and Christopher Ré. Taming the wild: A unified analysis of hogwild-style algorithms. In *Advances in neural information processing systems*, pages 2674–2682, 2015.
- [12] 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.
- [13] James W Demmel. *Applied numerical linear algebra*, volume 56. Siam, 1997.
- [14] Alexandros G Dimakis, Soumya Kar, José MF Moura, Michael G Rabbat, and Anna Scaglione. Gossip algorithms for distributed signal processing. *Proceedings of the IEEE*, 98(11):1847–1864, 2010.
- [15] Chris HQ Ding, Xiaofeng He, Hongyuan Zha, Ming Gu, and Horst D Simon. A min-max cut algorithm for graph partitioning and data clustering. In *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, pages 107–114. IEEE, 2001.
- [16] Nisrine Ghadban, Paul Honeine, Farah Mourad-Chehade, Joumana Farah, and Clovis Francis. Gossip algorithms for principal component analysis in networks. In *Signal Processing Conference (EUSIPCO), 2015 23rd European*, pages 2366–2370. IEEE, 2015.
- [17] Torben Hagerup and Christine Rüb. A guided tour of chernoff bounds. *Information Processing Letters*, 33(6):305–308, 1990.
- [18] Nicholas J Higham. *Accuracy and stability of numerical algorithms*. SIAM, 2002.
- [19] Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

- [20] Prateek Jain, Chi Jin, Sham M. Kakade, Praneeth Netrapalli, and Aaron Sidford. Streaming pca: Matching matrix bernstein and near-optimal finite sample guarantees for oja’s algorithm. In *Proceedings of the 29th Annual Conference on Computational Learning Theory (COLT)*, 2016.
- [21] David Kempe and Frank McSherry. A decentralized algorithm for spectral analysis. *Journal of Computer and System Sciences*, 74(1):70–83, 2008.
- [22] Satish Babu Korada, Andrea Montanari, and Sewoong Oh. Gossip pca. In *Proceedings of the ACM SIGMETRICS joint international conference on Measurement and modeling of computer systems*, pages 209–220. ACM, 2011.
- [23] Grzegorz Malewicz, Matthew H Austern, Aart JC Bik, James C Dehnert, Ilan Horn, Naty Leiser, and Grzegorz Czajkowski. Pregel: a system for large-scale graph processing. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, pages 135–146. ACM, 2010.
- [24] F. McSherry. Spectral partitioning of random graphs. In *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 529–537, 2001.
- [25] Frank McSherry. Spectral partitioning of random graphs. In *42nd Annual Symposium on Foundations of Computer Science, FOCS 2001, 14-17 October 2001, Las Vegas, Nevada, USA*, pages 529–537, 2001.
- [26] Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005.
- [27] Gemma Morral, Pascal Bianchi, and Jérémie Jakubowicz. Asynchronous distributed principal component analysis using stochastic approximation. In *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pages 1398–1403. IEEE, 2012.
- [28] Elchanan Mossel, Joe Neeman, and Allan Sly. Belief propagation, robust reconstruction and optimal recovery of block models. In *Conference on Learning Theory*, pages 356–370, 2014.
- [29] 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.
- [30] Erkki Oja. Simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15(3):267–273, 1982.
- [31] Benjamin Recht, Christopher Re, Stephen Wright, and Feng Niu. Hogwild: A lock-free approach to parallelizing stochastic gradient descent. In *Advances in neural information processing systems*, pages 693–701, 2011.
- [32] Mark Rudelson and Roman Vershynin. Smallest singular value of a random rectangular matrix. *Communications on Pure and Applied Mathematics*, 62(12):1707–1739, 2009.
- [33] Semih Salihoglu, Jaeho Shin, Vikesh Khanna, Ba Quan Truong, and Jennifer Widom. Graft: A debugging tool for apache giraph. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pages 1403–1408. ACM, 2015.

- [34] Daniel Spielman. Spectral graph theory: Spectral partitioning in a stochastic block model. University Lecture, 2015.
- [35] Ji-guang Sun. Rounding-error and perturbation bounds for the cholesky and ldlt factorizations. *Linear algebra and its applications*, 173:77–97, 1992.
- [36] Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434, 2012.
- [37] John Tsitsiklis, Dimitri Bertsekas, and Michael Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE transactions on automatic control*, 31(9):803–812, 1986.
- [38] Van Vu. A simple SVD algorithm for finding hidden partitions. *Combinatorics, Probability and Computing*, 27(1):124–140, 2018.

## A Oja’s Error Bound

In this section we give a full proof of [Theorem 3.1](#), restated below:

**Theorem 3.1.** *Let  $\mathbf{M} \in \mathbb{R}^{n \times n}$  be a PSD matrix satisfying:  $\frac{\sum_{i=1}^k \lambda_i(\mathbf{M})}{C} \leq \Lambda$ , and  $\frac{\lambda_k(\mathbf{M}) - \lambda_{k+1}(\mathbf{M})}{C} \geq \text{gap}$  for some bounds  $\Lambda, \text{gap}$ . For any  $\epsilon, \delta \in (0, 1]$ , let  $\xi \stackrel{\text{def}}{=} \frac{n}{\delta \epsilon \cdot \text{gap}}$ ,  $\eta = \frac{c_1 \epsilon^2 \cdot \text{gap} \cdot \delta^2}{C \Lambda k \log^3 \xi}$  for sufficiently small  $c_1$ , and  $T = \frac{c_2 \cdot (\log \xi + 1/\epsilon)}{C \cdot \text{gap} \cdot \eta}$  for sufficiently large  $c_2$ . Then [Algorithm 1](#) run with step size  $\eta$  satisfies, with probability  $\geq 1 - \delta$ , for all  $t \in [T - \frac{c_2/\epsilon}{2C \cdot \text{gap} \cdot \eta}, T + \frac{c_2/\epsilon}{2C \cdot \text{gap} \cdot \eta}]$ , letting  $\tilde{\mathbf{V}}_t = \text{orth}(\mathbf{Q}_t)$ :*

$$\|\mathbf{Z}^T \tilde{\mathbf{V}}_t\|_F^2 \leq \epsilon.$$

where  $\mathbf{Z}$  is an orthonormal basis for the bottom  $n - k$  eigenvectors of  $\mathbf{M}$ .

*Proof.* We first note that Theorem 1 of [1] requires  $\mathbb{P}_{\mathbf{x} \sim \mathcal{D}}[\|\mathbf{x}\|_2^2 \leq 1] = 1$ , while [Theorem 3.1](#) allows vectors with norm up to some bound  $C$ . It is clear that this suffices since we scale the step size  $\eta$ , along with the gap and  $\Lambda$  parameters by a factor of  $\frac{1}{C}$  as compared to their definitions in [1]. This translates to applying Theorem 1 to  $\frac{1}{C}\mathbf{M}$  – which has identical eigenvectors of  $\mathbf{M}$ . Thus, for the remainder of the proof we relabel  $\mathbf{M} = \frac{1}{C}\mathbf{M}$  and  $\eta = C\eta$ .

We next note that Theorem 1 of [1] sets  $\Lambda = \sum_{i=1}^k \lambda_i(\mathbf{M})$ , and  $\text{gap} = \lambda_k(\mathbf{M}) - \lambda_{k+1}(\mathbf{M})$ , while in [Theorem 3.1](#) we just require  $\Lambda$  and  $\text{gap}$  to be bounds on the respective quantities. It is not hard to see that the theorem still holds with these bounds. Specifically, the proof of Theorem 1 follows from the proof of Theorem 2 with parameter  $\rho$  set to any value  $\text{gap} \leq \lambda_k(\mathbf{M}) - \lambda_{k+1}(\mathbf{M})$ .  $\Lambda$  is also only used as an upper bound on  $\sum_{i=1}^k \lambda_i(\mathbf{M})$  in the proof of Lemma iii.K.1 and the subsequent proofs of Lemma Main 4 and Lemma Main 5.

Finally, we must argue that Theorem 5 follows from the proof of Theorem 1, even though our version of Oja’s algorithm uses a fixed step size  $\eta$ , rather than a step size which changes by round.

Even in our fixed  $\eta$  setting we employ the analysis of [1], which considers three epochs of rounds. Let  $\xi \stackrel{\text{def}}{=} \frac{n}{\delta \epsilon \cdot \text{gap}}$ . Set  $\eta = \frac{c_1 \epsilon^2 \cdot \text{gap} \cdot \delta^2}{\Lambda k \log^3 \xi}$ ,  $T_0 = \frac{c_2 \log \xi}{\text{gap} \cdot \eta}$ ,  $T_1 = \frac{c_3}{\text{gap} \cdot \eta}$ , and  $T_2 = \frac{c_4 T_1}{\epsilon}$  for some constants  $c_1, c_2, c_3, c_4$ . Overall we have  $T = T_0 + T_1 + T_2 = O\left(\frac{\Lambda k \log^3 \xi \cdot (\log \xi + 1/\epsilon)}{\epsilon^2 \cdot \text{gap}^2 \cdot \delta^2}\right)$ .

Theorem 1 of [1] follows from Theorem 2. The proof of this theorem first invokes Lemma Main 4. For sufficiently small  $q = \text{poly}(\xi)$ ,  $\Xi_{\mathbf{Z}} = \Theta\left(\frac{nk}{\delta^2} \ln \frac{n}{\delta}\right)$  and  $\Xi_x = \Theta\left(\frac{\sqrt{k \ln \frac{T}{\delta}}}{\delta}\right)$ , letting  $\eta$  be a step size which is fixed over all rounds, it requires:

$$\frac{2q(\Xi_{\mathbf{Z}}^{3/2} + n\Xi_x^2)}{\Lambda} \leq \eta = O\left(\frac{\text{gap}}{\Lambda \Xi_x^2 \ln \frac{nT}{q}}\right) \quad (7)$$

$$\sum_{t=1}^T \Lambda \eta^2 \Xi_x^2 = O\left(\frac{1}{\ln \frac{nT}{q}}\right) \quad (8)$$

$$\exists T_0 \leq T : \sum_{t=1}^{T_0} \eta = \Omega\left(\frac{\ln \Xi_{\mathbf{Z}}}{\text{gap}}\right). \quad (9)$$

Using that  $\Xi_x = \Theta\left(\frac{\sqrt{k \ln \frac{T}{\delta}}}{\delta}\right)$ , the upper bound of (7) holds as long as  $\eta = O\left(\frac{\text{gap} \cdot \delta^2}{\Lambda k \log^2 \xi}\right)$ , which is satisfied by our setting of  $\eta$  if  $c_1$  is small enough. The lower bound holds easily as well since  $q = \text{poly}(\xi)$  for some sufficiently small polynomial.

(8) holds as long as  $\eta = O\left(\frac{\delta}{\sqrt{\Lambda T k \log \xi}}\right)$ , which holds by our setting of  $\eta$  and  $T$  as long as  $c_1$  is sufficiently small compared to  $c_2, c_3$ , and  $c_4$ . Finally, (9) holds if  $T_0 = \Omega\left(\frac{\log \frac{nk}{\delta}}{\text{gap} \cdot \eta}\right)$ , which again holds for our setting of parameters.

In the proof of Theorem 2, Lemma Main 6 is next invoked with  $\Xi_x = \Xi_{\mathbf{Z}} = 2$  and  $T_0$  identified as the first round of computation. This lemma requires the conditions of Lemma Main 4 (discussed above) along with those of Lemma Main 5, which require that there is some  $\Delta \leq 1/\sqrt{8}$  such that:

$$\frac{T_1}{\ln^2 T_1} = \Omega\left(\frac{\log \frac{n}{q}}{\Delta^2}\right) \quad (10)$$

$$\forall t \in [T_1 + 1, T] : 2\eta \text{gap} - 4\eta^2 = \Omega\left(\frac{1}{t}\right) \quad (11)$$

$$\eta = O\left(\frac{1}{\sqrt{\Lambda t \Delta}}\right). \quad (12)$$

In fact, for our result, we can invoke Lemma Main 5 directly. Since we use an eigengap assumption, this lemma bounds (in the notation of [1])  $\|\mathbf{Z}^T \mathbf{P}_t \mathbf{Q} (\mathbf{V}^T \mathbf{P}_t \mathbf{Q})^{-1}\|_F^2 \leq \frac{5T_1 / \ln^2(T_1)}{(t-T_0) / \ln^2(t-T_0)}$ , for all  $t \in [T_0 + T_1, T]$ . This upper bounds  $\|\mathbf{Z}^T \tilde{\mathbf{V}}_t\|_F^2$  by Lemma 2.2. Since we set  $T_2 = \frac{c_4 T_1}{\epsilon}$  and  $T_1 = \frac{c_3}{\text{gap} \cdot \eta} \geq \frac{1}{\epsilon}$ , setting  $c_4$  sufficiently large gives  $\|\mathbf{Z}^T \tilde{\mathbf{V}}_T\|_F^2 \leq \epsilon$  and in fact,  $\|\mathbf{Z}^T \tilde{\mathbf{V}}_t\|_F^2 \leq \epsilon$  for all  $t \in [T - (1 - c_5)T_2, T]$  for any constant  $c_5 \in (0, 1]$ . Our final error bound follows by driving  $c_5$  sufficiently small and noting that we can shift  $T$ , multiplying it by at most constant factor, so that our bound holds for all  $t \in [T - (1 - c_5)T_2, T + (1 - c_5)T_2]$ .

It just remains to verify the conditions of Lemma Main 5. We set  $\Delta = \frac{c_5 \text{gap} \epsilon}{\sqrt{\Lambda}}$  for sufficiently small  $c_5$ . (10) holds since  $T_1 = \frac{c_3}{\text{gap} \eta} = \Omega\left(\frac{\log \xi \cdot \Lambda}{\text{gap}^2 \cdot \epsilon^2}\right)$ . For (11), first note that if  $c_1$  is set small enough, we can lower bound  $2\eta \text{gap} - 4\eta^2 \geq \eta \text{gap}$ . Thus (11) holds since we have  $T_1 = \frac{c_3}{\text{gap} \cdot \eta}$  for sufficiently large  $c_3$ . Finally, (12) holds since we have  $\sqrt{\Lambda} T_2 \Delta = c_5 c_4 T_1 \text{gap} = \frac{c_3 c_4 c_5}{\eta}$ . Thus, all conditions of Lemma Main 5 hold with our fixed  $\eta$ , yielding the theorem.  $\square$

## B Asynchronous Averaging

In this section we introduce a simple asynchronous averaging algorithm, which is used in our orthogonalization routine `AsynchOrth` (Algorithm 4), analyzed in Appendix C.

Our bounds closely follow classic work on asynchronous gossip algorithms [7], however we include a full proof for completeness, since our setting is more general than typically considered.

---

### Algorithm 7 ASYNCHRONOUS AVERAGING

---

**Initialization:** Each  $u$  holds value  $x_u$ .  $y_u := x_u$ .

**Update:** If  $(u, v)$  is chosen by the randomized scheduler:

1:  $y_u := \frac{y_u + y_v}{2}$ .

---

**Lemma B.1.** *Consider a set of nodes executing Algorithm 7 in the asynchronous communication model (Definition 2.1) with weight matrix  $\mathbf{W}$  and degree matrix  $\mathbf{D}$ . Let  $x_{avg} = \frac{1}{n} \sum_{v \in V} x_v$ . With probability  $\geq 1 - \delta$  in all rounds  $t \geq \frac{\log(1/\epsilon\delta)}{\log(\lambda_2^{-1}(\mathbf{I} - \frac{1}{2}\mathbf{D} + \frac{1}{2}\mathbf{W}))}$ ,  $\sum_{v \in V} (y_v - x_{avg})^2 \leq \epsilon \cdot \sum_{v \in V} (x_v - x_{avg})^2$ .*

That is, the mean squared error of the estimates of  $x_{avg}$  converges linearly, with rate dependent on the second eigenvalue of  $\mathbf{I} - \frac{1}{2}\mathbf{D} + \frac{1}{2}\mathbf{W}$ . Note that  $\mathbf{D} - \mathbf{W}$  is the Laplacian matrix corresponding to the communication model and  $\log(\lambda_2^{-1}(\mathbf{I} - \frac{1}{2}\mathbf{D} + \frac{1}{2}\mathbf{W})) = \log\left(\frac{1}{1 - \frac{1}{2}\lambda_{n-1}(\mathbf{L})}\right) \approx \lambda_{n-1}(\mathbf{L})$  is roughly its smallest nonzero eigenvalue.

*Proof.* Our proof follows that of [7, 8]. We can write  $x_{avg} = \frac{1}{n} \sum_{v \in V} x_v = \frac{1}{n} \cdot \mathbf{1}^T \mathbf{x}$  where  $\mathbf{x} \in \mathbb{R}^n$  contains each node's value as its entries. Recall that for convenience we identify the vertex set  $V$  with  $[n] = \{1, \dots, n\}$ . So the  $u^{th}$  entry of  $\mathbf{x}$  contains  $x_u$ . Let  $\mathbf{y}^t$  denote the vector of estimated averages at round  $t$ . Initially  $\mathbf{y}^0 = \mathbf{x}$ .

In each step of Algorithm 7, two nodes  $u$  and  $v$ , selected with probability  $\mathbf{W}_{u,v}$ , average their two values. We can write this update as a matrix product with  $\mathbf{y}^t$ . Specifically, for any pair  $(u, v)$ , let  $\mathbf{e}_{u,v} = \mathbf{e}_u - \mathbf{e}_v$ . We have, for  $\mathbf{e}_{u,v}$  chosen with probability  $\mathbf{W}_{u,v}$ ,

$$\mathbf{y}^{t+1} = (\mathbf{I} - \frac{1}{2}\mathbf{e}_{u,v}\mathbf{e}_{u,v}^T)\mathbf{y}^t. \quad (13)$$

We first note that  $\mathbf{1}^T(\mathbf{e}_{u,v}\mathbf{e}_{u,v}^T) = \mathbf{0}$ . Thus, by (13), for every  $t$ , we have  $\frac{1}{n}\mathbf{1}^T\mathbf{y}^t = \frac{1}{n}\mathbf{1}^T\mathbf{y}^0 = \frac{1}{n}\mathbf{1}^T\mathbf{x} = x_{avg}$ . That is, the average value held at the nodes always equals the true average. We bound the error from this average:  $\mathbf{z}^t \stackrel{\text{def}}{=} \mathbf{y}^t - x_{avg} \cdot \mathbf{1} = (\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T)\mathbf{y}^t$ . By (13) we have:

$$\begin{aligned} \mathbf{z}^{t+1} &= (\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T)(\mathbf{I} - \frac{1}{2}\mathbf{e}_{u,v}\mathbf{e}_{u,v}^T)\mathbf{y}^t = (\mathbf{I} - \frac{1}{2}\mathbf{e}_{u,v}\mathbf{e}_{u,v}^T)(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T)\mathbf{y}^t \\ &= (\mathbf{I} - \frac{1}{2}\mathbf{e}_{u,v}\mathbf{e}_{u,v}^T)\mathbf{z}^t \end{aligned}$$

where the second step follows since  $\mathbf{e}_{u,v}\mathbf{e}_{u,v}^T\mathbf{1}\mathbf{1}^T = \mathbf{1}\mathbf{1}^T\mathbf{e}_{u,v}\mathbf{e}_{u,v}^T = \mathbf{0}$  and  $\mathbf{I}$  commutes with all matrices. We can thus compute the expected norm  $\|\mathbf{z}^{t+1}\|_2^2 = (\mathbf{z}^{t+1})^T\mathbf{z}^{t+1}$  as:

$$\mathbb{E}[(\mathbf{z}^{t+1})^T\mathbf{z}^{t+1}|\mathbf{z}^t] = \sum_{\mathcal{P}} \mathbf{W}_{u,v} \cdot (\mathbf{z}^t)^T \left( \mathbf{I} - \frac{1}{2}\mathbf{e}_{u,v}\mathbf{e}_{u,v}^T \right)^2 \mathbf{z}^t \quad (14)$$

recalling that  $\mathcal{P}$  is the set of unordered pairs  $(u, v)$  with  $u \neq v$  and  $\mathbf{W}_{u,v}$  is the probability that such a pair is chosen to interact in any round  $t$ .  $(\mathbf{I} - \frac{1}{2}\mathbf{e}_{u,v}\mathbf{e}_{u,v}^T)^2 = \mathbf{I} - \mathbf{e}_{u,v}\mathbf{e}_{u,v}^T + \frac{1}{4}(\mathbf{e}_{u,v}\mathbf{e}_{u,v}^T)^2 = \mathbf{I} - \frac{1}{2}\mathbf{e}_{u,v}\mathbf{e}_{u,v}^T$  since  $(\mathbf{e}_{u,v}\mathbf{e}_{u,v}^T)^2 = 2\mathbf{e}_{u,v}\mathbf{e}_{u,v}^T$ . That is,  $\mathbf{I} - \frac{1}{2}\mathbf{e}_{u,v}\mathbf{e}_{u,v}^T$  is a projection matrix. Plugging back into (14):

$$\begin{aligned}\mathbb{E}[(\mathbf{z}^{t+1})^T \mathbf{z}^{t+1} | \mathbf{z}^t] &= (\mathbf{z}^t)^T \left( \sum_{\mathcal{P}} \mathbf{W}_{u,v} \cdot \left( \mathbf{I} - \frac{1}{2}\mathbf{e}_{u,v}\mathbf{e}_{u,v}^T \right) \right) \mathbf{z}^t \\ &= (\mathbf{z}^t)^T \left( \mathbf{I} - \frac{1}{2}(\mathbf{D} - \mathbf{W}) \right) \mathbf{z}^t.\end{aligned}\tag{15}$$

Denote  $\mathbf{H} \stackrel{\text{def}}{=} (\mathbf{I} - \frac{1}{2}(\mathbf{D} - \mathbf{W}))$ .  $\mathbf{H}$  is a sum over PSD matrices  $(\mathbf{I} - \frac{1}{2}\mathbf{e}_{u,v}\mathbf{e}_{u,v}^T)$  and so is itself PSD. It has top eigenvalue  $\lambda_1(\mathbf{H}) = 1$ , and top eigenvector  $\mathbf{1}$  (we can see this since  $(\mathbf{D} - \mathbf{W})\mathbf{1} = \mathbf{0}$ ). Since  $\mathbf{1}^T \mathbf{z}^t = \mathbf{1}^T (\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T) \mathbf{y}^t = 0$ , we thus have from (15),  $\mathbb{E}[\|\mathbf{z}^{t+1}\|_2^2 | \|\mathbf{z}^t\|_2^2] \leq \lambda_2(\mathbf{H})\|\mathbf{z}^t\|_2^2$  and by iterating:

$$\mathbb{E}[\|\mathbf{z}^t\|_2^2] \leq \lambda_2(\mathbf{H})^t \|\mathbf{z}^0\|_2^2\tag{16}$$

If we set  $t \geq \frac{\log(1/\epsilon\delta)}{\log(\lambda_2(\mathbf{H})^{-1})}$  applying (16) gives  $\mathbb{E}[\|\mathbf{z}^t\|_2^2] \leq \epsilon\delta\|\mathbf{z}^0\|_2^2$  and thus by Markov's inequality,  $\mathbb{P}[\|\mathbf{z}^t\|_2^2 \leq \epsilon\|\mathbf{z}^0\|_2^2] \geq 1 - \delta$ . This gives the bound since  $\|\mathbf{z}^t\|_2^2 = \sum_{v \in V} (y_v^t - x_{avg})^2$ . It just remains to note that since  $\mathbf{z}^{t+1} = (\mathbf{I} - \frac{1}{2}\mathbf{e}_{u,v}\mathbf{e}_{u,v}^T) \mathbf{z}^t$  and  $\|\mathbf{I} - \frac{1}{2}\mathbf{e}_{u,v}\mathbf{e}_{u,v}^T\|_2 = 1$ , the error strictly decreases in each round, so once it is bounded in round  $t$ , it is bounded in all subsequent rounds.  $\square$

## C Distributed orthogonalization proofs

In this section we give a full analysis of the distributed orthogonalization routine `AsynchOrth` described in Algorithm 4. We are interested in the error bounds it gives for eigenvector approximation when used as a subroutine in Algorithm 2. We begin by analyzing the idealized case when Algorithm 4 is assumed to *exactly* implement the centralized Algorithm 3. We then account for the fact that the distributed implementation is approximate.

**Corollary C.1** (Distributed eigenvector approximation with exact Cholesky orthogonalization). *Let  $\mathbf{v}_1, \dots, \mathbf{v}_k$  be the top  $k$  eigenvectors of the communication matrix  $\mathbf{D} + \mathbf{W}$  in an asynchronous communication model, and let  $\Lambda, \overline{\text{gap}}$ , be bounds satisfying:  $\sum_{j=1}^k \lambda_j(\mathbf{D} + \mathbf{W}) \leq \Lambda$  and  $\min_{j \in [k]} [\lambda_j(\mathbf{D} + \mathbf{W}) - \lambda_{j+1}(\mathbf{D} + \mathbf{W})] \geq \overline{\text{gap}}$ .*

*For any  $\epsilon, \delta \in (0, 1)$ , let  $\xi = \frac{n}{\delta\epsilon\overline{\text{gap}}}$  and let  $\eta = \frac{c_1\epsilon^2\overline{\text{gap}}\cdot\delta^2}{\Lambda k^3 \log^3 \xi}$  for sufficiently small  $c_1$  and  $T = \frac{c_2 \cdot (\log \xi + 1/\epsilon)}{\overline{\text{gap}} \cdot \eta}$  for sufficiently large  $c_2$ . For all  $u \in [n], i \in [k]$ , let  $q_u^{(j)}$  be the local state computed by Algorithm 2 prior to Step 4. Then for all  $u \in [n], i \in [k]$ , let  $\tilde{v}_u^{(j)}$  be result of running `AsynchOrth` in Step 4 with an algorithm that exactly implements Algorithm 3. If  $\tilde{\mathbf{V}}$  is given by  $(\tilde{\mathbf{V}})_{u,j} = \tilde{v}_u^{(j)}$  and  $\tilde{\mathbf{v}}_i$  is the  $i^{\text{th}}$  column of  $\tilde{\mathbf{V}}$ , then with probability  $\geq 1 - \delta$ :*

$$|\tilde{\mathbf{v}}_i^T \mathbf{v}_i| \geq 1 - \epsilon \quad \text{and} \quad \|\tilde{\mathbf{v}}_i\|_2 = 1 \quad \text{for all } i \in [k].$$

We note that a similar eigenvector bound is given in [1] with better dependence on  $k$ . However, it has worse dependence on  $\Lambda$  and so is too weak for our applications.

*Proof.* Let  $\tilde{\mathbf{Q}}$  be given by  $(\mathbf{Q})_{u,j} = q_u^{(j)}$ . [Algorithm 3](#) first computes  $\mathbf{L} = \text{chol}(\mathbf{Q}^T \mathbf{Q})$  and then  $\tilde{\mathbf{V}} = \mathbf{Q}(\mathbf{L}^T)^{-1}$ . Letting  $\mathbf{Q}_i \in \mathbb{R}^{n \times i}$  denote the first  $i$  columns of  $\mathbf{Q}$ , it is well known that, letting  $\mathbf{L}_i$  denote the upper left  $i \times i$  submatrix of  $\mathbf{L}$ ,  $\mathbf{L}_i = \text{chol}(\mathbf{Q}_i^T \mathbf{Q}_i)$ . Furthermore, since  $\mathbf{L}$  is triangular,  $(\mathbf{L}_i^T)^{-1}$  is just the upper  $i \times i$  submatrix of  $(\mathbf{L}^T)^{-1}$ . Thus we can see that the first  $i$  columns of  $\mathbf{V}$  are *identical* to the output that would be obtained if the algorithm were run with the same step size  $\eta$  and step count  $T$ , but with just the first  $i$  vectors of  $\mathbf{Q}$  – i.e. with each node only keeping track of just  $q_u^{(1)}, \dots, q_u^{(i)}$  instead of  $q_u^{(1)}, \dots, q_u^{(k)}$ .

With this observation, we can prove the corollary by applying [Theorem 3.1](#) for each  $i \in [k]$ . We apply the theorem with rank  $i$ , error  $\epsilon/2$ , failure probability  $\delta/k$ ,  $\Lambda \geq \sum_{j=1}^k \lambda_j(\mathbf{D} + \mathbf{W}) \geq \sum_{j=1}^i \lambda_j(\mathbf{D} + \mathbf{W})$ , and  $\overline{\text{gap}} \leq \min_{j \in [k]} [\lambda_j(\mathbf{M}) - \lambda_{j+1}(\mathbf{M})] \leq \lambda_i(\mathbf{D} + \mathbf{W}) - \lambda_{i+1}(\mathbf{D} + \mathbf{W})$ .

Denote  $\mathbf{V}_i = [\mathbf{v}_1, \dots, \mathbf{v}_i]$ , and let  $\mathbf{V}_{-i}$  span the remaining eigenvectors of  $\mathbf{D} + \mathbf{W}$ . By our application of the theorem with rank  $i$ , with probability  $\geq 1 - \delta/k$ ,  $\|\mathbf{V}_{-i}^T \tilde{\mathbf{V}}_i\|_F^2 \leq \epsilon/2$  and:

$$\|\mathbf{V}_i^T \tilde{\mathbf{V}}_i\|_F^2 \geq i - \epsilon/2 \quad (17)$$

since  $\|\tilde{\mathbf{V}}_i\|_F^2 = i$  and by the Pythagorean theorem,  $\|\tilde{\mathbf{V}}_i\|_F^2 = \|\mathbf{V}_i^T \tilde{\mathbf{V}}_i\|_F^2 + \|\mathbf{V}_{-i}^T \tilde{\mathbf{V}}_i\|_F^2$ . This holds for all  $i$  simultaneously with probability  $\geq 1 - \delta$  after union bounding over  $k$  applications of the theorem. Since  $\|\mathbf{V}_i^T \tilde{\mathbf{V}}_{i-1}\|_F^2 \leq \|\tilde{\mathbf{V}}_{i-1}\|_F^2 = i - 1$ , (17) gives:

$$\|\mathbf{V}_i^T \tilde{\mathbf{v}}_i\|_2^2 = \|\mathbf{V}_i^T \tilde{\mathbf{V}}_i\|_F^2 - \|\mathbf{V}_i^T \tilde{\mathbf{V}}_{i-1}\|_F^2 \geq 1 - \epsilon/2. \quad (18)$$

For  $i = 1$ , this completes the proof since  $\|\mathbf{V}_1^T \tilde{\mathbf{v}}_1\|_2^2 = (\mathbf{v}_1^T \tilde{\mathbf{v}}_1)^2$ . For  $i > 1$ , we also have by our application of [Theorem 3.1](#) with rank  $i - 1$ ,  $\|\mathbf{V}_{i-1}^T \tilde{\mathbf{V}}_{i-1}\|_F^2 \geq i - 1 - \epsilon/2$  and so:

$$\|\mathbf{V}_{i-1}^T \tilde{\mathbf{v}}_i\|_2^2 \leq \|\tilde{\mathbf{V}}_{i-1}\|_F^2 - \|\mathbf{V}_{i-1}^T \tilde{\mathbf{V}}_{i-1}\|_F^2 \leq \epsilon/2. \quad (19)$$

Since  $\|\mathbf{V}_i^T \tilde{\mathbf{v}}_i\|_2^2 = \|\mathbf{V}_{i-1}^T \tilde{\mathbf{v}}_i\|_2^2 + (\mathbf{v}_i^T \tilde{\mathbf{v}}_i)^2$ , in combination (18) and (19) give  $(\mathbf{v}_i^T \tilde{\mathbf{v}}_i)^2 \geq 1 - \epsilon$  and hence  $|\mathbf{v}_i^T \tilde{\mathbf{v}}_i| \geq 1 - \epsilon$ .  $\|\tilde{\mathbf{v}}_i\|_2 = 1$  follows from:

$$\tilde{\mathbf{V}}^T \tilde{\mathbf{V}} = \mathbf{L}^{-1} \mathbf{Q}^T \mathbf{Q} (\mathbf{L}^T)^{-1} = \mathbf{L}^{-1} \mathbf{L} \mathbf{L}^T (\mathbf{L}^T)^{-1} = \mathbf{I}_{k \times k}.$$

□

With [Corollary C.1](#) in place, we next focus on the additional error introduced by the fact that [Algorithm 4](#) only implements [Algorithm 3](#) approximately. We first bound how well [Algorithm 4](#) approximates  $\mathbf{Q}^T \mathbf{Q}$  via averaging.

**Lemma C.2.** *Consider a set of nodes executing [AsynchOrth\(T\)](#) ([Algorithm 4](#)) in the asynchronous communication model with weight matrix  $\mathbf{W}$  and degree matrix  $\mathbf{D}$  satisfying:*

$$\gamma_{\text{mix}} \leq \min \left[ \frac{1}{n}, \log \left( \lambda_2^{-1} \left( \mathbf{I} - \frac{1}{2} \mathbf{D} + \frac{1}{2} \mathbf{W} \right) \right) \right]. \text{ Let } \mathbf{Q} \in \mathbb{R}^{n \times k} \text{ be given by } (\mathbf{Q})_{u,j} = q_u^{(j)}.$$

*For any  $\epsilon, \delta \in (0, 1)$ , if  $T \geq \frac{c \log \left( \frac{n \|\mathbf{Q}\|_2}{\epsilon \delta} \right)}{\gamma_{\text{mix}}}$  for sufficiently large constant  $c$ , with probability  $\geq 1 - \delta$ , for all  $u$ , and all  $i, j \in [k]$ ,*

$$|(\mathbf{R}_u)_{i,j} - (\mathbf{Q}^T \mathbf{Q})_{i,j}| \leq \epsilon.$$

*Proof.* We apply [Lemma B.1](#) of [Appendix B](#) to bound the accuracy of the averaging protocol in computing  $(\mathbf{Q}^T \mathbf{Q})_{i,j}$ . Specifically, with probability  $\geq 1 - \delta$ , for every round  $t' \geq T$ , letting  $\epsilon' =$

$\left(\frac{\epsilon}{n^{3/2}\|\mathbf{Q}\|_2}\right)^2$  the theorem gives:

$$\left[\left(\mathbf{R}_u\right)_{i,j} - \frac{1}{n}(\mathbf{Q}^T \mathbf{Q})_{i,j}\right]^2 \leq n^2 \cdot \sum_{u \in V} \left[r_u^{(i,j)} \frac{1}{n}(\mathbf{Q}^T \mathbf{Q})_{i,j}\right]^2 \leq n^2 \epsilon' \cdot \sum_{u \in U} \left[q_u^{(i)} q_u^{(j)} - \frac{1}{n}(\mathbf{Q}^T \mathbf{Q})_{i,j}\right]^2. \quad (20)$$

We can loosely bound:

$$\sum_{u \in U} \left[q_u^{(i)} q_u^{(j)} - \frac{1}{n}(\mathbf{Q}^T \mathbf{Q})_{i,j}\right]^2 \leq \sum_{u \in U} \left[q_u^{(i)} q_u^{(j)}\right]^2 = \|\mathbf{Q}\|_F^2 \leq n \|\mathbf{Q}\|_2^2.$$

The lemma follows by plugging into (20) and taking a square root of the error bound.  $\square$

**Remark C.3.** We note that, as shown in Lemma B.1, the accuracy of averaging to approximate  $\mathbf{Q}^T \mathbf{Q}$  only decreases with each round. Thus, *AsynchOrth* (Algorithm 4) does not need to terminate the averaging after  $T$  rounds, but can continuously maintain approximations to  $\mathbf{Q}^T \mathbf{Q}$  and  $[\hat{v}_u^{(1)}, \dots, \hat{v}_u^{(k)}]$ , which will only become more accurate over time.

We next show how the entrywise approximation bound of Lemma C.2 translates into error in computing  $\tilde{\mathbf{V}} = \mathbf{Q}(\mathbf{L}_u^T)^{-1}$ . We use a forward stability result on the Cholesky decomposition:

**Theorem C.4** (Theorem 10.8 of [18], from [35]). *Let  $\mathbf{R} \in \mathbb{R}^{k \times k}$  be positive definite with Cholesky decomposition  $\mathbf{R} = \mathbf{L}\mathbf{L}^T$ . Let  $\Delta_R$  be a symmetric matrix satisfying  $\|\mathbf{R}^{-1} \cdot \Delta_R\|_2 < 1$ . Then  $\mathbf{R} + \Delta_R$  has the Cholesky decomposition  $\mathbf{R} + \Delta_R = (\mathbf{L} + \Delta_L)(\mathbf{L} + \Delta_L)^T$  where:*

$$\frac{\|\Delta_L\|_F}{\|\mathbf{L}\|_2} \leq \frac{\|\mathbf{R}^{-1}\|_2 \cdot \|\Delta_R\|_F}{\sqrt{2}(1 - \|\mathbf{R}^{-1}\|_2 \cdot \|\Delta_R\|_F)}.$$

Using this result we can show:

**Lemma C.5.** *Consider  $\mathbf{Q} \in \mathbb{R}^{n \times k}$  with maximum and minimum singular values  $\sigma_{\max}(\mathbf{Q})$  and  $\sigma_{\min}(\mathbf{Q})$ . For all  $u \in [n]$ , let  $\mathbf{R}_u \in \mathbb{R}^{k \times k}$  be any symmetric matrix with  $|(\mathbf{R}_u)_{i,j} - (\mathbf{Q}^T \mathbf{Q})_{i,j}| \leq \epsilon$  for all  $i, j$  and some  $\epsilon \leq \frac{n \min(1, \sigma_{\min}(\mathbf{Q})^3)}{2\sqrt{2} \max(1, \sigma_{\max}(\mathbf{Q}))}$ . Let  $\mathbf{L} = \text{chol}(\mathbf{Q}^T \mathbf{Q})$ ,  $\mathbf{L}_u = \text{chol}(\mathbf{R}_u)$ ,  $\tilde{\mathbf{V}} = \mathbf{Q}(\mathbf{L}^T)^{-1}$ , and  $\hat{\mathbf{V}}$  have  $u^{\text{th}}$  row equal to the  $u^{\text{th}}$  row of  $\mathbf{Q}(\mathbf{L}_u^T)^{-1}$ . Then letting  $\tilde{\mathbf{v}}_i$  and  $\hat{\mathbf{v}}_i$  be the  $i^{\text{th}}$  columns of  $\tilde{\mathbf{V}}$  and  $\hat{\mathbf{V}}$  respectively, for all  $i \in [k]$ ,*

$$\|\hat{\mathbf{v}}_i - \tilde{\mathbf{v}}_i\|_2 \leq \frac{2\sqrt{2}\epsilon n^2 \sigma_{\max}(\mathbf{Q})}{\sigma_{\min}(\mathbf{Q})^3}.$$

*Proof.* Denote  $\mathbf{R} = \mathbf{Q}^T \mathbf{Q}$ . By the bound  $|(\mathbf{R}_u)_{i,j} - (\mathbf{Q}^T \mathbf{Q})_{i,j}| \leq \epsilon$  we can write  $\mathbf{R}_u = \mathbf{R} + \Delta_R$  where  $\Delta_R$  is symmetric with  $\|\Delta_R\|_F \leq n\epsilon$ . We have

$$\|\mathbf{R}^{-1} \Delta_R\|_2 \leq \|\mathbf{R}^{-1}\|_2 \cdot \|\Delta_R\|_F < \frac{n\epsilon}{\sigma_{\min}(\mathbf{Q})^2} < \frac{1}{2}$$

where the last step follows from our upper bound on  $\epsilon$ . Plugging into Theorem C.4 gives that  $\mathbf{R}_u$  has Cholesky decomposition:  $\mathbf{R}_u = \mathbf{L}_u \mathbf{L}_u^T$  where  $\mathbf{L}_u = \mathbf{L} + \Delta_L$  and

$$\|\Delta_L\|_2 \leq \|\Delta_L\|_F \leq \|\mathbf{L}\|_2 \cdot \sqrt{2} \|\mathbf{R}^{-1}\|_2 \|\Delta_R\|_F \leq \frac{\sqrt{2} \cdot \epsilon n \cdot \sigma_{\max}(\mathbf{Q})}{\sigma_{\min}(\mathbf{Q})^2}. \quad (21)$$

Let  $\tilde{\mathbf{V}}^{(u)}$  be the  $u^{th}$  row of  $\tilde{\mathbf{V}} = \mathbf{Q}(\mathbf{L}^T)^{-1}$  and  $\hat{\mathbf{V}}^{(u)}$  be the  $u^{th}$  row of  $\hat{\mathbf{V}}$  equal to the  $u^{th}$  row of  $\mathbf{Q}(\mathbf{L}_u^T)^{-1}$ . Using (21), we can bound  $\|\mathbf{L}^{-1}\|_2 \cdot \|\Delta_L\|_2 \leq \frac{\sqrt{2\epsilon n} \sigma_{\max}(\mathbf{Q})}{\sigma_{\min}(\mathbf{Q})^3} \leq \frac{1}{2}$  by our upper bound on  $\epsilon$ . By a standard linear system stability bound (e.g., Theorem 7.2 of [18]):

$$\|\tilde{\mathbf{V}}^{(u)} - \hat{\mathbf{V}}^{(u)}\|_2 \leq \|\tilde{\mathbf{V}}^{(u)}\|_2 \cdot \frac{\|\mathbf{L}^{-1}\|_2 \cdot \|\Delta_L\|_2}{1 - \|\mathbf{L}^{-1}\|_2 \cdot \|\Delta_L\|_2} \leq \frac{2\sqrt{2\epsilon n} \sigma_{\max}(\mathbf{Q})}{\sigma_{\min}(\mathbf{Q})^3}$$

where we bound  $\|\tilde{\mathbf{V}}^{(u)}\|_2 \leq 1$  since  $\tilde{\mathbf{V}}$  is orthonormal. Since every entry in  $\tilde{\mathbf{V}}^{(u)} - \hat{\mathbf{V}}^{(u)}$  can be loosely bounded in magnitude by the vector's norm, summing over all  $n$  rows gives  $\|\hat{\mathbf{v}}_i - \tilde{\mathbf{v}}_i\|_2 \leq \frac{2\sqrt{2\epsilon n^2} \sigma_{\max}(\mathbf{Q})}{\sigma_{\min}(\mathbf{Q})^3}$ , completing the lemma.  $\square$

To apply Lemma C.5 we must bound  $\sigma_{\max}(\mathbf{Q})$  and  $\sigma_{\min}(\mathbf{Q})$ . We have the following:

**Lemma C.6** (Conditioning of  $\mathbf{Q}$ ). *Consider a set of nodes executing `AsynchOja`( $T, T', \eta$ ) (Algorithm 2) in the asynchronous communication model with weight matrix  $\mathbf{W}$  and with step size  $\eta$  and stopping time  $T$  as specified in Corollary C.1. Let  $\mathbf{Q} \in \mathbb{R}^{n \times k}$  be given by  $(\mathbf{Q})_{u,j} = q_u^{(j)}$  after round  $T$ . For some large enough constant  $c$ , with probability  $1 - \delta - e^{-\Theta(n)}$ :*

$$\sigma_{\max}(\mathbf{Q}) \leq e^{c(\log \xi + 1/\epsilon) \cdot \left[ \frac{\lambda_1(\mathbf{D} + \mathbf{W})}{\text{gap}} + 1 \right]} \quad \text{and} \quad \sigma_{\min}(\mathbf{Q}) \geq \delta/c.$$

*Proof.* We can write  $\mathbf{Q} = \left( \prod_{j=1}^T (\mathbf{I} + \eta \mathbf{A}_j) \right) \mathbf{Q}_0$  for  $T = \frac{c_2(\log \xi + 1/\epsilon)}{\text{gap} \cdot \eta}$  where  $\mathbf{A}_j = \mathbf{e}_{u,v} \mathbf{e}_{u,v}^T$  for some pair  $u, v$ . Recall that  $\mathbf{Q}_0$  has all entries independently selected from  $\mathcal{N}(0, 1)$ . We can loosely bound that, with probability  $\geq 1 - \delta/3$ ,  $\|\mathbf{Q}_0\|_2 \leq \|\mathbf{Q}_0\|_F \leq cnk \sqrt{\log(1/\delta)}$  for some sufficiently large  $c$ . Additionally, by Theorem 1.1 of [32] we have  $\sigma_{\min}(\mathbf{Q}_0) \geq \delta/c$  with probability  $\geq 1 - \delta/3 - e^{-\Theta(n)}$  for sufficiently large  $c$ . Starting with the lower bound, since  $\mathbf{A}_j$  is always PSD and so all eigenvalues of  $\mathbf{I} + \eta \mathbf{A}_j$  are  $\geq 1$ ,

$$\begin{aligned} \sigma_{\min}(\mathbf{Q}) &\geq \prod_{j=1}^T \sigma_{\min}(\mathbf{I} + \eta \mathbf{A}_j) \cdot \sigma_{\min}(\|\mathbf{Q}_0\|_2) \\ &\geq \sigma_{\min}(\|\mathbf{Q}_0\|_2) \geq \delta/c. \end{aligned}$$

For the upper bound, for any  $0 \leq t \leq T$  let  $\mathbf{B}_t = \prod_{j=1}^t (\mathbf{I} + \eta \mathbf{A}_j)$ . Let  $\mathbf{B}_0 = \mathbf{I}$ . Define  $\alpha_t \stackrel{\text{def}}{=} \mathbb{E}[\text{tr}(\mathbf{B}_t \mathbf{B}_t^T)] = \text{tr}(\mathbb{E}[\mathbf{B}_t \mathbf{B}_t^T])$ . We have  $\alpha_t \geq \mathbb{E}[\|\mathbf{B}_t\|_2^2]$ . Further, we can bound  $\alpha_t$  by:

$$\begin{aligned} \alpha_t &= \mathbb{E}[\text{tr}((\mathbf{I} + \eta \mathbf{A}_t) \mathbf{B}_{t-1} \mathbf{B}_{t-1}^T (\mathbf{I} + \eta \mathbf{A}_t)^T)] \\ &= \mathbb{E}[\text{tr}(\mathbf{B}_{t-1} \mathbf{B}_{t-1}^T (\mathbf{I} + \eta \mathbf{A}_t)^T (\mathbf{I} + \eta \mathbf{A}_t))] && \text{(By cyclic property of the trace)} \\ &= \text{tr}(\mathbb{E}[\mathbf{B}_{t-1} \mathbf{B}_{t-1}^T] \mathbb{E}[(\mathbf{I} + \eta \mathbf{A}_t)^T (\mathbf{I} + \eta \mathbf{A}_t)]) && \text{(By independence of samples } \mathbf{A}_t) \\ &= \text{tr}(\mathbb{E}[\mathbf{B}_{t-1} \mathbf{B}_{t-1}^T] (\mathbf{I} + 2(\eta + \eta^2)(\mathbf{D} + \mathbf{W}))) && \text{(Since } \mathbb{E}[\mathbf{A}_t] = \mathbf{D} + \mathbf{W} \text{ and } \mathbf{A}_t^2 = 2\mathbf{A}_t) \\ &\leq \text{tr}(\mathbb{E}[\mathbf{B}_{t-1} \mathbf{B}_{t-1}^T] \cdot \|\mathbf{I} + 2(\eta + \eta^2)(\mathbf{D} + \mathbf{W})\|_2) \\ &\leq \alpha_{t-1} \cdot (1 + 4\eta\lambda_1(\mathbf{D} + \mathbf{W})) \end{aligned}$$

where the second to last bound follows since  $\mathbf{D} + \mathbf{W}$  and  $\mathbf{B}_t \mathbf{B}_t^T$  are PSD. The last bound follows since  $\eta < 1$ . We thus have, applying this argument inductively,  $\alpha_T \leq (1 + 4\eta\lambda_1(\mathbf{D} + \mathbf{W}))^T$ , and so,

$\|\mathbf{B}_t\|_2 \leq \|\mathbf{B}_t\|_2^2 \leq \text{tr}(\mathbf{B}_t \mathbf{B}_t^T) \leq \frac{(1+4\eta\lambda_1(\mathbf{D}+\mathbf{W}))^T}{\delta/3}$  with probability  $\geq 1 - \delta/3$  by Markov's inequality. Combined with our bound on  $\|\mathbf{Q}_0\|_2$  this gives, with probability  $\geq 1 - \delta/3 - \delta/3$ :

$$\begin{aligned} \sigma_{\max}(\mathbf{Q}) = \|\mathbf{Q}\|_2 &\leq \prod_{j=1}^T \|\mathbf{I} + \eta \mathbf{A}_j\|_2 \cdot \|\mathbf{Q}_0\|_2 \\ &\leq \frac{(1 + 4\eta\lambda_1(\mathbf{D} + \mathbf{W}))^{\frac{c_2(\log \xi + 1/\epsilon)}{\text{gap} \cdot \eta}}}{\delta} \cdot \|\mathbf{Q}_0\|_2 \\ &\leq e^{\frac{c(\log \xi + 1/\epsilon) \cdot \lambda_1(\mathbf{D} + \mathbf{W})}{\text{gap}}} \cdot \frac{\|\mathbf{Q}_0\|_2}{\delta/3} \\ &\leq e^{\frac{c(\log \xi + 1/\epsilon) \cdot \lambda_1(\mathbf{D} + \mathbf{W})}{\text{gap}} + c \log \xi} \\ &\leq e^{c(\log \xi + 1/\epsilon) \cdot \left[ \frac{\lambda_1(\mathbf{D} + \mathbf{W})}{\text{gap}} + 1 \right]} \end{aligned}$$

for some large enough  $c$ . In the second to last step we bound  $\frac{\|\mathbf{Q}_0\|_2}{\delta/3} \leq \frac{cnk\sqrt{\log(1/\delta)}}{\delta/3} \leq c \log \xi$  for large enough  $c$ . The theorem follows after union bounding, which gives that both our upper and lower bounds hold with probability  $\geq 1 - \delta$ .  $\square$

We can finally complete our analysis, proving our main asynchronous eigenvector approximation theorem, restated below:

**Theorem 3.4.** *Let  $\mathbf{v}_1, \dots, \mathbf{v}_k$  be the top  $k$  eigenvectors of the communication matrix  $\mathbf{D} + \mathbf{W}$  in an asynchronous communication model, and let  $\Lambda, \text{gap}, \gamma_{\text{mix}}$  be bounds satisfying:  $\Lambda \geq \sum_{j=1}^k \lambda_j(\mathbf{D} + \mathbf{W})$ ,  $\text{gap} \leq \min_{j \in [k]} [\lambda_j(\mathbf{D} + \mathbf{W}) - \lambda_{j+1}(\mathbf{D} + \mathbf{W})]$ , and  $\gamma_{\text{mix}} \leq \min \left[ \frac{1}{n}, \log \left( \lambda_2^{-1}(\mathbf{I} - \frac{1}{2}\mathbf{D} + \frac{1}{2}\mathbf{W}) \right) \right]$ .*

*For any  $\epsilon, \delta \in (0, 1)$ , let  $\xi = \frac{n}{\delta \epsilon \cdot \text{gap}}$ . Let  $\eta = \frac{c_1 \epsilon^2 \cdot \text{gap} \cdot \delta^2}{\Lambda k^3 \log^3 \xi}$  for sufficiently small  $c_1$ , and  $T = \frac{c_2 \cdot (\log \xi + 1/\epsilon)}{\text{gap} \cdot \eta}$ ,  $T' = \frac{c_3 (\log \xi + 1/\epsilon) \cdot \lambda_1(\mathbf{D} + \mathbf{W})}{\text{gap} \cdot \gamma_{\text{mix}}}$  for sufficiently large  $c_2, c_3$ . For all  $u \in [n], i \in [k]$ , let  $\hat{v}_u^{(j)}$  be the local state computed by Algorithm 2. If  $\hat{\mathbf{V}} \in \mathbb{R}^{n \times k}$  is given by  $(\hat{\mathbf{V}})_{u,j} = \hat{v}_u^{(j)}$  and  $\hat{\mathbf{v}}_i$  is the  $i^{\text{th}}$  column of  $\hat{\mathbf{V}}$ , then with probability  $\geq 1 - \delta - e^{-\Theta(n)}$ , for all  $i \in [k]$ :*

$$|\hat{\mathbf{v}}_i^T \mathbf{v}_i| \geq 1 - \epsilon \quad \text{and} \quad \|\hat{\mathbf{v}}_i\|_2 \leq 1 + \epsilon.$$

*Proof.* Let  $\mathbf{Q} \in \mathbb{R}^{n \times k}$  be given by  $(\mathbf{Q})_{u,j} = q_u^{(j)}$ , after round  $T$  (i.e.  $\mathbf{Q}$  is the input passed to `AsynchOrth` by `AsynchOja`). Let  $\tilde{\mathbf{V}}$  be the output given by running exact Cholesky orthogonalization (Algorithm 3) on  $\mathbf{Q}$ . Let  $\tilde{\mathbf{v}}_i$  be its  $i^{\text{th}}$  column. Let  $\epsilon' = \frac{\epsilon \delta^3}{n^2 \cdot \exp\left(\frac{c(\log \xi + 1/\epsilon) \cdot \lambda_1(\mathbf{D} + \mathbf{W})}{\text{gap}}\right)}$ . By our bound on

$$\sigma_{\max}(\mathbf{Q}) = \|\mathbf{Q}\|_2 \text{ in Lemma C.6, with probability } 1 - \delta - \epsilon^{-\Theta(n)}, T' \text{ satisfies } T' = \Omega\left(\frac{\log\left(\frac{n\|\mathbf{Q}\|_2}{\epsilon'\delta}\right)}{\gamma_{\text{mix}}}\right).$$

So by Lemma C.2, conditioned on the previous bound holding, with probability  $\geq 1 - \delta$ , for all  $u$  and  $i, j \in [k]$ ,  $\mathbf{R}_u$  computed in Algorithm 4 satisfies:

$$|(\mathbf{R}_u)_{i,j} - (\mathbf{Q}^T \mathbf{Q})_{i,j}| \leq \epsilon'.$$

We can then apply Lemma C.5 and our bounds on the maximum and minimum singular values of  $\mathbf{Q}$  in Lemma C.6. With probability  $1 - 2\delta - e^{-\Theta(n)}$ , for all  $i \in [k]$ :

$$\|\hat{\mathbf{v}}_i - \tilde{\mathbf{v}}_i\|_2 \leq \frac{2\sqrt{2}\epsilon' n^2 e^{\frac{c(\log \xi + 1/\epsilon) \cdot \lambda_1(\mathbf{D} + \mathbf{W})}{\text{gap}}}}{(\delta/c)^3} = O(\epsilon). \quad (22)$$

Now, by [Corollary C.1](#) and our setting of  $\eta$  and  $T$ , with probability  $1 - \delta$ , for all  $i \in [k]$ ,  $|\tilde{\mathbf{v}}_i^T \mathbf{v}_i| \geq 1 - \epsilon$ . Since  $\|\mathbf{v}_i\|_2 = 1$ , (22) gives, with overall probability  $\geq 1 - 3\delta - e^{-\Theta(n)}$ :

$$\begin{aligned} |\hat{\mathbf{v}}_i^T \mathbf{v}_i| &= |\tilde{\mathbf{v}}_i^T \mathbf{v}_i - (\tilde{\mathbf{v}}_i^T - \hat{\mathbf{v}}_i^T) \mathbf{v}_i| \geq |\tilde{\mathbf{v}}_i^T \mathbf{v}_i| - |(\tilde{\mathbf{v}}_i^T - \hat{\mathbf{v}}_i^T) \mathbf{v}_i| \\ &\geq 1 - \epsilon - \|\hat{\mathbf{v}}_i - \tilde{\mathbf{v}}_i\|_2 \\ &\geq 1 - O(\epsilon). \end{aligned}$$

Additionally, using (22) and  $\|\tilde{\mathbf{v}}_i\|_2 = 1$  (shown in [Corollary C.1](#)), we can apply triangle inequality to show  $\|\hat{\mathbf{v}}_i\|_2 \leq 1 + O(\epsilon)$ . The theorem follows by adjusting constants on  $\epsilon, \delta$ .  $\square$

## D Community Detection in the $(n, p, q)$ -weighted communication model

To convert our matrix concentration bound to a bound on the closeness of the two eigenvectors, we apply the Davis-Kahan theorem:

**Theorem D.1** (Davis-Kahan Theorem – [13] Theorem 5.4). *Let  $\mathbf{M}$  and  $\hat{\mathbf{M}}$  be symmetric matrices with eigenvectors  $\mathbf{v}_1, \dots, \mathbf{v}_n$  and  $\hat{\mathbf{v}}_1, \dots, \hat{\mathbf{v}}_n$  respectively. We have:*

$$1 - |\hat{\mathbf{v}}_i^T \mathbf{v}_i| \leq \frac{2\|\mathbf{M} - \hat{\mathbf{M}}\|_2}{\min[\lambda_{i-1}(\mathbf{M}) - \lambda_i(\mathbf{M}), \lambda_i(\mathbf{M}) - \lambda_{i+1}(\mathbf{M})]}.$$

*Proof.* Theorem 5.4 of [13] states the above bound with the lefthand side equal to  $\sin 2\theta_i$ , where  $\theta_i$  is the angle between  $\mathbf{v}_i$  and  $\hat{\mathbf{v}}_i$ . It is noted in the proof that the bound can also be proven on  $\sin \theta_i = \sqrt{1 - \cos^2 \theta_i} = \sqrt{1 - (\hat{\mathbf{v}}_i^T \mathbf{v}_i)^2} \geq 1 - |\hat{\mathbf{v}}_i^T \mathbf{v}_i|$ , giving our statement of the bound.  $\square$

Using this theorem we show:

**Lemma D.2** (Concentration of  $G(n, p, q)$  communication matrix second eigenvector). *Let  $\mathbf{v}_2$  be the second eigenvector of  $\mathbf{C}^{(p, q)}$  for any  $p > q > 0$ . Let  $\tilde{\mathbf{v}}_2$  be the second eigenvector of  $(\mathbf{D} + \mathbf{W})$ , where  $\mathbf{W}$  is the communication weight matrix and  $\mathbf{D}$  is the degree matrix in the  $G(n, p, q)$ -communication model. Then if  $\frac{\min[q, p-q]}{\sqrt{p+q}} \geq \frac{9\sqrt{\log(n/\delta)}}{\epsilon\sqrt{n}}$ , with probability  $\geq 1 - \delta$ :*

$$|\tilde{\mathbf{v}}_2^T \mathbf{v}_2| \geq 1 - \epsilon.$$

*Proof.* Consider the  $n \times \binom{n}{2}$  matrix  $\mathbf{B}$  with columns indexed by unordered pairs of vertices,  $(u, v)$  with  $u \neq v$ . If  $u$  and  $v$  are in the same cluster, the  $(u, v)$  column is  $\sqrt{p} \cdot \mathbf{e}_{u,v}$ . If they are in different clusters it is  $\sqrt{q} \cdot \mathbf{e}_{u,v}$ . We can see that  $\mathbf{B}\mathbf{B}^T = \mathbf{C}^{(p, q)} + \frac{n(p+q)-2p}{2} \cdot \mathbf{I}_{n \times n}$ . Thus, the second eigenvector of  $\mathbf{B}\mathbf{B}^T$  is identical to that of  $\mathbf{C}^{(p, q)}$ ,  $\mathbf{v}_2$ . Further we can compute the eigenvalues:

$$\begin{aligned} \lambda_1(\mathbf{B}\mathbf{B}^T) &= \lambda_1(\mathbf{C}^{(p, q)}) + \frac{n(p+q)-2p}{2} = n(p+q) - p \\ \lambda_2(\mathbf{B}\mathbf{B}^T) &= \lambda_2(\mathbf{C}^{(p, q)}) + \frac{n(p+q)-2p}{2} = np - p \\ \lambda_i(\mathbf{B}\mathbf{B}^T) &= \frac{n(p+q)}{2} - p \text{ for all } i \in \{3, \dots, n\}. \end{aligned}$$

Let  $\tilde{\mathbf{B}}\tilde{\mathbf{B}}^T$  be obtained from  $\mathbf{B}$  by independently setting each column to  $\mathbf{b}_i/\sqrt{p_i}$  with probability  $p_i$  equal to its squared norm ( $p$  for intracluster columns and  $q$  for intercluster columns) and to  $\mathbf{0}$  otherwise. In this way  $\mathbb{E}[\tilde{\mathbf{B}}\tilde{\mathbf{B}}^T] = \mathbf{B}\mathbf{B}^T$ . Further, we can see that sampling  $\tilde{\mathbf{B}}\tilde{\mathbf{B}}^T$  is identical to sampling  $\mathbf{D} + \mathbf{W}$  in the  $G(n, p, q)$ -communication model, up to a scaling. Thus, to prove the lemma, it suffices to show that  $\tilde{\mathbf{v}}_2$ , the second eigenvector of  $\tilde{\mathbf{B}}\tilde{\mathbf{B}}^T$  is close to  $\mathbf{v}_2$  with good probability. We do this via a matrix Bernstein bound, which shows that  $\tilde{\mathbf{B}}\tilde{\mathbf{B}}^T$  is close to  $\mathbf{B}\mathbf{B}^T$  with good probability.

Specifically let  $\mathbf{b}_i$  and  $\tilde{\mathbf{b}}_i$  be the  $i^{\text{th}}$  columns of  $\mathbf{B}$  and  $\tilde{\mathbf{B}}$  respectively. We can apply Theorem 1.4 of [36], where in the notation of the theorem we set  $\mathbf{X}_k = \tilde{\mathbf{b}}_k\tilde{\mathbf{b}}_k^T - \mathbf{b}_k\mathbf{b}_k^T$ . We have  $\mathbb{E}[\mathbf{X}_k] = \mathbf{0}$  and  $\|\mathbf{X}_k\| \leq 2$  always, and

$$\begin{aligned} \sigma^2 &= \left\| \sum_k \mathbb{E}(\mathbf{X}_k^2) \right\|_2 = \left\| \sum_k \left( \mathbb{E}[\|\tilde{\mathbf{b}}_k\|_2^2 \cdot \tilde{\mathbf{b}}_k\tilde{\mathbf{b}}_k^T] - 2\mathbb{E}[\tilde{\mathbf{b}}_k\tilde{\mathbf{b}}_k^T \mathbf{b}_k\mathbf{b}_k^T] + \|\mathbf{b}_k\|_2^2 \mathbf{b}_k\mathbf{b}_k^T \right) \right\|_2 \\ &= \left\| \sum_k (2\mathbf{b}_k\mathbf{b}_k^T - \|\mathbf{b}_k\|_2^2 \cdot \mathbf{b}_k\mathbf{b}_k^T) \right\|_2 \\ &\quad \text{(Since } \|\tilde{\mathbf{b}}_k\|_2^2 = 2 \text{ with probability } \|\mathbf{b}_k\|_2^2, 0 \text{ otherwise.)} \\ &\leq \|\mathbf{B}\mathbf{B}^T\|_2 \leq 2n(p+q). \end{aligned}$$

where the final bound follows by our computation of  $\lambda_1(\mathbf{B}\mathbf{B}^T) = \|\mathbf{B}\mathbf{B}^T\|_2$  above. The inequality follows from the fact that  $2 - \|\mathbf{b}_k\|_2^2 \in [2 - p^2, 2 - q^2]$  and so  $\sum_k (2\mathbf{b}_k\mathbf{b}_k^T - \|\mathbf{b}_k\|_2^2 \cdot \mathbf{b}_k\mathbf{b}_k^T)$  is PSD and  $\preceq (2 - q^2)\mathbf{B}\mathbf{B}^T \preceq 2\mathbf{B}\mathbf{B}^T$ .

Plugging the above bounds into Theorem 1.4 of [36] gives:

$$\mathbb{P} \left[ \|\tilde{\mathbf{B}}\tilde{\mathbf{B}}^T - \mathbf{B}\mathbf{B}^T\|_2 > \Delta \right] \leq n \cdot \exp \left( \frac{-\Delta^2/2}{2n(p+q) + 2\Delta/3} \right). \quad (23)$$

Applying the Davis-Kahan theorem (Theorem D.1) to  $\mathbf{B}\mathbf{B}^T$  and  $\tilde{\mathbf{B}}\tilde{\mathbf{B}}^T$ , we have using our eigenvalue calculations for  $\mathbf{B}\mathbf{B}^T$ :

$$\min \left[ \lambda_1(\mathbf{B}\mathbf{B}^T) - \lambda_2(\mathbf{B}\mathbf{B}^T), \lambda_2(\mathbf{B}\mathbf{B}^T) - \lambda_3(\mathbf{B}\mathbf{B}^T) \right] = \min \left( q \cdot n, \frac{n(p-q)}{2} \right). \quad (24)$$

and so  $1 - |\tilde{\mathbf{v}}_2^T \mathbf{v}_2| \leq \epsilon$  as long as  $\Delta \leq \epsilon \cdot \min \left[ \frac{q \cdot n}{2}, \frac{n(p-q)}{4} \right]$ . Plugging this in to (23) and simplifying:

$$\begin{aligned} \mathbb{P} \left[ \|\tilde{\mathbf{B}}\tilde{\mathbf{B}}^T - \mathbf{B}\mathbf{B}^T\|_2 > \epsilon \cdot \min \left[ \frac{q \cdot n}{2}, \frac{n(p-q)}{4} \right] \right] &\leq n \cdot \exp \left( \frac{-\epsilon^2 \cdot \min \left[ \frac{q^2 \cdot n^2}{4}, \frac{n^2(p-q)^2}{16} \right] / 2}{2.166 \cdot n(p+q)} \right) \\ &\leq n \cdot \exp \left( \frac{-\epsilon^2 n \cdot \min [q^2, (p-q)^2]}{70(p+q)} \right) \end{aligned} \quad (25)$$

This probability is bounded by  $\delta$  if  $\frac{\min[q, p-q]}{\sqrt{p+q}} \geq \frac{\sqrt{70 \log(n/\delta)}}{\epsilon \sqrt{n}}$ , giving the lemma.  $\square$

With Lemma D.2 ensuring that the second eigenvector of  $\mathbf{D} + \mathbf{W}$  in fact approximates the cluster indicator vector  $\chi$ , we can now show that approximately computing this eigenvector using the **Asynch0ja** algorithm and thresholding its entries by their signs gives an approximately correctly distributed community detection protocol.

**Theorem 4.3.** Consider a set of nodes executing *AsynchCD*( $T, T', \eta$ ) (*Algorithm 5*) in the  $G(n, p, q)$ -communication model. Let  $\rho = \min\left(\frac{q}{p+q}, \frac{p-q}{p+q}\right)$ . If for sufficiently small constant  $c_1$  and sufficiently large constants  $c_2, c_3$ ,

$$\eta = \frac{c_1 \epsilon^2 \delta^2 \cdot \rho}{\log^3\left(\frac{n}{\epsilon \delta \rho}\right)}, \quad T = \frac{c_2 \cdot n \cdot \left(\log^3\left(\frac{n}{\epsilon \delta \rho}\right) + \frac{\log\left(\frac{n}{\epsilon \delta \rho}\right)}{\epsilon}\right)}{\epsilon^2 \delta^2 \rho^2}, \quad \text{and} \quad T' = \frac{c_3 \cdot n \cdot \left(\log\left(\frac{n}{\epsilon \delta \rho}\right) + \frac{1}{\epsilon}\right)}{\rho^2},$$

and if  $\frac{\min[q, p-q]}{\sqrt{p+q}} \geq \frac{c_4 \sqrt{\log(n/\delta)}}{\epsilon \sqrt{n}}$  for large enough constant  $c_4$ , then, with probability  $1 - \delta$ , after ignoring  $\epsilon \cdot n$  nodes, all remaining nodes in  $V_1$  terminate in some state  $s_1$ , and all remaining nodes in  $V_2$  terminate in some different state  $s_2$ . Suppressing polylogarithmic factors in the parameters, the total number of global rounds and local rounds required are:  $T + T' = \tilde{O}\left(\frac{n}{\epsilon^3 \delta^2 \rho^2}\right)$  and  $L = \tilde{O}\left(\frac{1}{\epsilon^3 \delta^3 \rho^2}\right)$ .

*Proof.* Let  $\mathbf{v}_2 = \chi$  be the second eigenvector of  $\mathbf{C}^{(p,q)}$  and  $\tilde{\mathbf{v}}_2$  be the second eigenvector of  $\mathbf{D} + \mathbf{W}$ . By [Lemma D.2](#), with probability  $1 - \delta$ ,  $|\tilde{\mathbf{v}}_2^T \mathbf{v}_2| \geq 1 - \epsilon$  which gives if  $\tilde{\mathbf{v}}_2^T \mathbf{v}_2 \geq 0$ ,

$$\epsilon \geq 1 - \tilde{\mathbf{v}}_2^T \mathbf{v}_2 = \tilde{\mathbf{v}}_2^T (\tilde{\mathbf{v}}_2 - \mathbf{v}_2)$$

and so  $\|\tilde{\mathbf{v}}_2 - \mathbf{v}_2\|_2 \leq \epsilon$ . Similarly, if  $\tilde{\mathbf{v}}_2^T \mathbf{v}_2 \leq 0$  we have  $\|\tilde{\mathbf{v}}_2 + \mathbf{v}_2\|_2 \leq \epsilon$ . If  $\hat{\mathbf{v}}_2$  satisfies  $|\hat{\mathbf{v}}_2^T \tilde{\mathbf{v}}_2| \geq 1 - \epsilon$  and  $\|\hat{\mathbf{v}}_2\|_2 \leq 1 + \epsilon$  then this gives:

$$|\hat{\mathbf{v}}_2^T \mathbf{v}_2| \geq |\hat{\mathbf{v}}_2^T \tilde{\mathbf{v}}_2| - |\hat{\mathbf{v}}_2^T (\mathbf{v}_2 - \tilde{\mathbf{v}}_2)| \geq 1 - (2 + \epsilon)\epsilon$$

in the case where  $\tilde{\mathbf{v}}_2^T \mathbf{v}_2 \geq 0$  and

$$|\hat{\mathbf{v}}_2^T \mathbf{v}_2| \geq |-\hat{\mathbf{v}}_2^T \tilde{\mathbf{v}}_2| - |\hat{\mathbf{v}}_2^T (\mathbf{v}_2 + \tilde{\mathbf{v}}_2)| \geq 1 - (2 + \epsilon)\epsilon$$

in the case where  $\tilde{\mathbf{v}}_2^T \mathbf{v}_2 \leq 0$ . Either way, we have:

$$|\hat{\mathbf{v}}_2^T \mathbf{v}_2| = 1 - O(\epsilon). \quad (26)$$

Now, in the  $G(n, p, q)$ -communication model  $\mathbf{D} + \mathbf{W}$  is a scaling of  $\tilde{\mathbf{B}}\tilde{\mathbf{B}}^T$ , as defined in the proof of [Lemma D.2](#). The scale factor is the inverse of the number of edges in the sampled graph which is  $\frac{n^2(p+q)-2np}{4}$  in expectation. As long as  $p + q \geq \frac{c \log(n/\delta)}{n}$  for large enough  $c$  (which is implied by our assumption  $\frac{\min[q, p-q]}{\sqrt{p+q}} \geq \frac{c_4 \sqrt{\log(n/\delta)}}{\epsilon \sqrt{n}}$ ), then by a simple Chernoff bound, the number of sampled edges will be within a 2 factor of this expected value with probability  $1 - \delta$ . Thus:

$$\mathbf{D} + \mathbf{W} = s \cdot \tilde{\mathbf{B}}\tilde{\mathbf{B}}^T \quad \text{for some} \quad s \geq \frac{2}{n^2(p+q)}. \quad (27)$$

Again applying the matrix Bernstein bound in [Lemma D.2](#), equation (25), with probability  $\geq 1 - \delta$ :

$$\|\tilde{\mathbf{B}}\tilde{\mathbf{B}}^T - \mathbf{B}\mathbf{B}^T\|_2 \leq \epsilon \min\left[\frac{q \cdot n}{2}, \frac{n(p-q)}{4}\right]. \quad (28)$$

Combined with the eigenvalue gap calculations for  $\mathbf{B}\mathbf{B}^T$  shown in (24), we have:

$$\min \left[ \lambda_1(\tilde{\mathbf{B}}\tilde{\mathbf{B}}^T) - \lambda_2(\tilde{\mathbf{B}}\tilde{\mathbf{B}}^T), \lambda_2(\tilde{\mathbf{B}}\tilde{\mathbf{B}}^T) - \lambda_3(\tilde{\mathbf{B}}\tilde{\mathbf{B}}^T) \right] \geq (1 - \epsilon) \cdot \min \left( q \cdot n, \frac{n(p - q)}{2} \right). \quad (29)$$

After scaling, by (27):

$$\begin{aligned} \min [\lambda_1(\mathbf{D} + \mathbf{W}) - \lambda_2(\mathbf{D} + \mathbf{W}), \lambda_2(\mathbf{D} + \mathbf{W}) - \lambda_3(\mathbf{D} + \mathbf{W})] &\geq \frac{2(1 - \epsilon)}{n^2(p + q)} \cdot \min \left( qn, \frac{n(p - q)}{2} \right) \\ &\geq \frac{\rho}{2n} \end{aligned}$$

where in the final bound we assume  $1 - \epsilon \geq 1/2$  which is without loss of generality, since we can always scale  $\epsilon$  down by a  $1/2$  factor.

We similarly use the perturbation bound of (28), the scale bound of (27) and our eigenvalue calculations for  $\mathbf{B}\mathbf{B}^T$  to argue that that  $\lambda_1(\mathbf{D} + \mathbf{W}) \leq \frac{16}{n}$  and that  $\lambda_2(\mathbf{I} - 1/2\mathbf{D} + 1/2\mathbf{W}) \leq 1 - \frac{q}{2n(p + q)}$  and so  $\log(\lambda_2^{-1}(\mathbf{I} - 1/2\mathbf{D} + 1/2\mathbf{W})) \geq \frac{q}{2n(p + q)} \geq \frac{\rho}{2n}$ .

We can thus apply Theorem 3.4 with  $k = 2$ ,  $\Lambda = \frac{32}{n}$ ,  $\overline{\text{gap}} = \frac{\rho}{2n}$ , and  $\gamma_{\text{mix}} = \frac{\rho}{2n}$ . With these parameters we can set, for sufficiently small  $c_1$  and large  $c_2, c_3$ ,

$$\eta = \frac{c_1 \epsilon^2 \delta^2 \cdot \rho}{\log^3 \left( \frac{n}{\epsilon \delta \rho} \right)}, \quad T = \frac{c_2 \cdot n \cdot \left( \log^3 \left( \frac{n}{\epsilon \delta \rho} \right) + \frac{\log \left( \frac{n}{\epsilon \delta \rho} \right)}{\epsilon} \right)}{\epsilon^2 \delta^2 \rho^2}, \quad \text{and} \quad T' = \frac{c_3 \cdot n \cdot \left( \log \left( \frac{n}{\epsilon \delta \rho} \right) + \frac{1}{\epsilon} \right)}{\rho^2}.$$

where to bound  $T'$  we use that  $\frac{\lambda_1(\mathbf{D} + \mathbf{W})}{\overline{\text{gap}}} \leq \frac{32}{\rho}$ . Letting  $\hat{\mathbf{V}} \in \mathbb{R}^{n \times k}$  be given by  $(\hat{\mathbf{V}})_{u,j} = \hat{v}_u^{(j)}$  where  $\hat{v}_u^{(j)}$  are the outputs of  $\text{AsynchOja}(T, T', \eta)$  and letting  $\hat{\mathbf{v}}_2$  be the second column of  $\hat{\mathbf{V}}$ , with these parameters, Theorem 3.4 gives that with probability  $1 - \delta$ ,

$$|\hat{\mathbf{v}}_2^T \tilde{\mathbf{v}}_2| \geq 1 - \epsilon \quad \text{and} \quad \|\hat{\mathbf{v}}_2\|_2 \leq 1 + \epsilon$$

where  $\tilde{\mathbf{v}}_2$  is the second eigenvector of  $\mathbf{D} + \mathbf{W}$ . By (26) we thus have  $|\hat{\mathbf{v}}_2^T \mathbf{v}_2| \geq 1 - O(\epsilon)$ .

Applying Lemma 4.1 after adjusting  $\epsilon$  by a constant factor then gives the theorem. Additionally, in expectation, each node is involved in  $L = \Theta \left( \frac{T + T'}{n} \right)$  interactions. This bound also holds for all nodes with probability  $1 - \delta$  by a Chernoff bound, since  $L = \Omega(\log(n/\delta))$ . We can union bound over the various events required for the theorem to hold, all occurring with probability  $\geq 1 - \delta$ , which gives the final theorem after adjusting  $\delta$  by a constant factor.  $\square$

## E Analysis of CLEANUP PHASE in $(n, p, q)$ -weighted communication model

As we will argue in Appendix F, if  $q \leq p/2$ , then the global number of rounds simply becomes  $O(n \log^2 n)$ .

**Lemma E.1.** *Consider  $(n, p, q)$ -weighted communication model. After  $O \left( \frac{72n \ln n}{(\sqrt{p'} - \sqrt{q'})^2} \right)$  global rounds Algorithm CLEANUP PHASE with parameters  $k = 1$  and  $r = \frac{72n \ln n}{(\sqrt{p'} - \sqrt{q'})^2}$ , where  $p' = \frac{(1 - \epsilon)p}{p + q} > 0$  and  $q' = \frac{q + \epsilon p}{p + q}$ ,  $p' > q'$  we have that all nodes are correctly labeled.*

*Proof.* Fix an arbitrary node  $u$ . Let  $X_u$  ( $X'_u$ ), respectively) denote the number of times an edge  $(u, v)$  was chosen where  $v$ 's current label is the same as  $u$ 's ground-truth label meaning that both are in  $V_1$  or  $V_2$  (opposite, respectively). Such an edge is scheduled at a given time step with probability at least  $p_1 \stackrel{\text{def}}{=} \frac{(1-\epsilon)n/2p}{p(\frac{n}{2})^2 + q(\frac{n}{2})^2} = \frac{2(1-\epsilon)p}{(p+q)n}$  (at most  $p_2 \stackrel{\text{def}}{=} \frac{2(q+\epsilon p)}{(p+q)n}$ , respectively).

We show that w.h.p.  $X_u > X'_u$ ; taking Union bound over the complementary events for all nodes yields the claim.

We distinguish between two cases. First assume  $q' \geq p'/12$ . Let  $\delta = \sqrt{6 \ln n / \mathbb{E}[X_u]}$ ,  $\delta' = \sqrt{6 \ln n / \mathbb{E}[X'_u]}$  and observe that  $\delta, \delta' \leq 1$ . By Chernoff bounds with, we get,

$$\mathbb{P}[X_u \geq (1-\delta)\mathbb{E}[X_u]] + \mathbb{P}[X'_u \leq (1+\delta')\mathbb{E}[X'_u]] \leq 2e^{-2 \ln n}. \quad (30)$$

Conditioning on this, we have

$$X_u - X'_u \geq (1-\delta)\mathbb{E}[X_u] - (1+\delta')\mathbb{E}[X'_u] = rp_1 - \sqrt{6 \ln n rp_1} - \left(rp_2 + \sqrt{6 \ln n rp_2}\right) > 0, \quad (31)$$

for  $r = \frac{72n \ln n}{(\sqrt{p'} - \sqrt{q'})^2} \geq \frac{6 \ln n}{(\sqrt{p_1} - \sqrt{p_2})^2}$ . Similarly, if  $q' \leq p'/12$ , then, by [Theorem F.5](#),

$$\mathbb{P}[X_u < X'_u] \leq \mathbb{P}[X_u < \mathbb{E}[X_u]/2] + \mathbb{P}[X'_u \geq 6 \max\{\ln n, \mathbb{E}[X'_u]\}] \leq 2e^{-2 \ln n}. \quad (32)$$

Thus, in both cases we get

$$\mathbb{P}[X_u \leq X'_u] \leq 2e^{-2 \ln n}. \quad (33)$$

This finishes the proof.  $\square$

*Proof of [Theorem 5.1](#).* The lemma is a direct consequence of [Lemma E.1](#).  $\square$

## F Analysis of CLEANUP PHASE in $G(n, p, q)$ -communication model

Let  $E(u, S)$  denote the number of edges between  $u \in V$  and  $S \subseteq V$ . Recall that

$$\Delta = \frac{p}{2} - \frac{q}{2} - \sqrt{12p \ln n / n} - \sqrt{12q \ln n / n}.$$

Throughout this section we will assume that

$$\Delta \geq 2^{15} \ln n / n. \quad (34)$$

Let  $n^* = \frac{n\Delta}{240p}$ . Fix an arbitrary set  $S \subseteq V$  with  $|S| \leq n^*$ . For  $u \in V$  define

$$Y_u^S = \begin{cases} 1 & E(u, S) \geq \Delta n / 12 \\ 0 & \text{otherwise} \end{cases}.$$

Let

$$Y^S = \begin{cases} 1 & \sum_{u \in V} Y_u^S \geq |S|/3 \\ 0 & \text{otherwise} \end{cases}.$$

**Lemma F.1.** Assume that  $|S| \leq n^*$ . We have that

$$\mathbb{P}[Y_S = 1] \leq e^{-\Delta|S|n/2^{11}}.$$

*Proof.* Observe that the  $\{Y_u^S : u \notin S\}$  are independent 0/1-variables. We remark that  $\{Y_u^S : u \in S\}$ , are correlated and we will need a different approach to bound  $\sum_{u \in S} Y_u^S$  which we save for later. Applying Chernoff with  $\delta = \frac{n\Delta}{p|S|12} - 1$ ,  $\mathbb{E}[E(u, S)] = p|S|$ . Note that  $\frac{n\Delta}{p|S|12} - 1 \geq \frac{n\Delta}{p|S|24} \geq 1$ , due to the bound on  $|S|$ . We get

$$p' \stackrel{\text{def}}{=} \mathbb{P}[Y_u^S = 1] = \mathbb{P}[E(u, S) \geq n\Delta/12] = \mathbb{P}[E(u, S) \geq (1 + \delta)\mathbb{E}[E(u, S)]] \quad (35)$$

$$\leq \exp\left(-\frac{p|S|}{3} \left(\frac{n\Delta}{p|S|12} - 1\right)\right) \leq e^{-\Delta n/72}. \quad (36)$$

where the penultimate inequality stems from Chernoff bounds. By [Theorem F.4](#), with parameters  $\alpha = \frac{|S|}{6m}$ ,  $m = n - |S|$ ,  $p = p'$ , we get that

$$\begin{aligned} \mathbb{P}\left[\sum_{u \in V \setminus S} Y_u^S \geq |S|/6\right] &\leq \left(\left(\frac{p'}{\frac{|S|}{6m}}\right)^{\frac{|S|}{6m}} \left(\frac{1-p'}{1-\frac{|S|}{6m}}\right)^{1-\frac{|S|}{6m}}\right)^m \\ &\leq \left(\left(\frac{p'}{\frac{|S|}{6m}}\right)^{\frac{|S|}{6m}} \left(\frac{1}{1-\frac{|S|}{6m}}\right)\right)^m \leq \left((6mp')^{\frac{|S|}{6m}}\right)^m \left(\frac{1}{1-\frac{|S|}{6m}}\right)^m \\ &\leq (6m)^{\frac{|S|}{6}} (p')^{\frac{|S|}{6}} \left(\frac{1}{1-\frac{|S|}{6n}}\right)^m \\ &\leq \exp\left(\frac{|S|}{6} \ln(6m) - \frac{|S|}{6} \Delta n/72 + m \ln\left(\frac{1}{1-\frac{|S|}{6m}}\right)\right) \\ &\leq \exp\left(\frac{|S|}{6} \ln(6m) - \frac{|S|}{6} \Delta n/72 + m \frac{e|S|}{6m}\right) \\ &\leq \exp\left(-\frac{|S|\Delta n}{2^{10}}\right), \end{aligned}$$

where the penultimate inequality comes from  $\ln(1/(1-x)) \leq ex$  for  $x \in (0, 0.8]$  and the last inequality comes from [\(34\)](#).

We now turn to bounding  $\{Y_u^S : u \in S\}$ . In order for  $\sum_{u \in S} Y_u^S \geq |S|/6$  a counting argument shows that the number of required edges with both endpoints in  $S$  needs to be at least  $\frac{\Delta n}{12} \frac{|S|}{6} \frac{1}{2}$ . we have  $\mathbb{E}[E(S, S)] = p|S|^2/2$ . In Applying Chernoff bounds yields  $\delta = \frac{n\Delta}{72p|S|} - 1 \geq \frac{n\Delta}{144p|S|} \geq 1$

$$\begin{aligned} \mathbb{P}\left[\sum_{u \in S} Y_u^S \geq |S|/6\right] &\leq \mathbb{P}\left[E(S, S) \geq \frac{\Delta n}{12} \frac{|S|}{6}\right] = \mathbb{P}[E(S, S) \geq (1 + \delta)\mathbb{E}[E(S, S)]] \\ &\leq \exp\left(-\frac{|S|\Delta n}{2^{10}}\right). \end{aligned}$$

We have

$$\mathbb{P}[Y^S = 1] \leq \mathbb{P}\left[\sum_{u \in V \setminus S} Y_u^S \geq |S|/6\right] + \mathbb{P}\left[\sum_{u \in S} Y_u^S \geq |S|/6\right] \leq 2 \exp\left(-\frac{|S|\Delta n}{2^{10}}\right)$$

□

We say a graph is *smooth* if for all subsets  $S$  of size  $|S| \in [1, n^*]$  we have  $Y^S = 0$ .

**Lemma F.2.** *Let  $G_{n,p}$  be an Erdős-Rényi graph with parameters  $n, p$  satisfying (34). Then,  $G_{n,p}$  is smooth w.h.p..*

*Proof.* Applying Union bound,

$$\mathbb{P}\left[\sum_{S, |S| \in [1, n^*]} Y_S = 0\right] \leq \sum_{S, |S| \in [1, n^*]} \mathbb{P}[Y_S = 0] \leq \sum_{i=1}^{n^*} \sum_{S, |S|=i} \mathbb{P}[Y_S = 0] \quad (37)$$

$$\leq \sum_{i=1}^{n^*} \binom{n}{i} \mathbb{P}[Y_S = 0 \mid |S| = i] \leq \sum_{i=1}^{n^*} \left(\frac{en}{i}\right)^i e^{-\Delta ni/2^{11}} \quad (38)$$

$$\leq \sum_{i=1}^{n^*} \exp(i \ln n - \Delta ni/2^{11}) \leq \sum_{i=1}^{n^*} \frac{1}{n^3} \leq \frac{1}{n^2}. \quad (39)$$

□

**Lemma F.3.** *Consider  $G(n, p, q)$ -communication model. Assume the graph is smooth. Let  $S_t$  be the set of nodes that are incorrectly labeled after phase  $t$ . If  $|S_t| \in [1, n^*]$ , then after one additional phase of  $O\left(\frac{np \log n}{(\sqrt{p''} - \sqrt{q''})^2}\right)$  global rounds Algorithm CLEANUP PHASE with parameter  $r = \frac{72pn \log n}{(\sqrt{p''} - \sqrt{q''})^2}$ , where*

1.  $p'' = \frac{p}{2} - \sqrt{\frac{6p \log n}{n}} - \frac{\Delta}{12}$  and
2.  $q'' = \frac{q}{2} + \sqrt{\frac{6q \log n}{n}} + \frac{\Delta}{12}$  and
3.  $\Delta = \Omega(\log n/n)$ ,

conditioning on  $\mathcal{F}_t$  we get that w.h.p.  $|S_{t+1}| \leq (2/3)|S_t|$ , where  $\mathcal{F}_t$  denotes the filtration up to time  $t$ .

*Proof.* The proof idea is as follows. Since the graph is smooth, there are at most  $|S_t|/3$  nodes with a large number of edges to  $S_t$ . In the remainder of the proof we will show that w.h.p. all other nodes will set their label correctly (including most of the nodes of  $S_t$  itself). This implies, by taking Union bound over all nodes, that w.h.p.  $|S_{t+1}| \leq |S_t|/3$ .

Fix an arbitrary node  $u$  with  $Y_u^S = 0$ . The number of edges  $u$  has within its own cluster is at least  $np/2 - \sqrt{12np \log n}$  w.p. at least  $1 - n^2$ , by Chernoff bounds with parameter  $\delta = \sqrt{\frac{12 \log n}{np}}$ .

Again, let  $X_u$  ( $X'_u$ ), respectively) denote the number of times an edge  $(u, v)$  was chosen where  $v$ 's current label is the same as  $u$ 's ground-truth label (opposite, respectively). We get, by Lemma F.2 that

$$\mathbb{P}[X_u] = \frac{np/2 - \sqrt{12np \log n} - \Delta n/12}{|E|}.$$

We assume w.l.o.g. that  $q = \Omega(\log n/n)$  otherwise, the proof follows trivially since (34) implies that  $p = \Omega(\log n/n)$ .

$$\mathbb{P}[X'_u] = \frac{nq/2 + \sqrt{12nq \log n} + \Delta n/12}{|E|}$$

Note that  $n^2p/8 \leq |E| \leq 2n^2p$  w.h.p. Applying the same argument as in Lemma E.1 with  $p_1 = \mathbb{P}[X_u]$  and  $p_2 = \mathbb{P}[X'_u]$ , we get that  $r \geq \frac{72pn \log n}{(\sqrt{p''} - \sqrt{q''})^2}$  rounds are sufficient w.h.p.

For completeness we give the proof. W.l.o.g.  $q = \Omega(\log n/n)$ . Let  $\delta = \sqrt{6 \log n / \mathbb{E}[X_u]}$ ,  $\delta' = \sqrt{6 \log n / \mathbb{E}[X'_u]}$  and observe that  $\delta, \delta' \leq 1$ . By Chernoff bounds, we get,

$$\mathbb{P}[X_u \geq (1 - \delta)\mathbb{E}[X_u]] + \mathbb{P}[X'_u \leq (1 + \delta')\mathbb{E}[X'_u]] \leq 2e^{-2 \log n}. \quad (40)$$

Conditioning on this, we have

$$X_u - X'_u \geq (1 - \delta)\mathbb{E}[X_u] - (1 + \delta')\mathbb{E}[X'_u] = rp_1 - \sqrt{6 \log n rp_1} - \left(rp_2 + \sqrt{6 \log n rp_2}\right) > 0, \quad (41)$$

where the last inequality holds as long as  $r \geq \frac{6 \log n}{(\sqrt{p_1} - \sqrt{p_2})^2}$ .

We have

$$r = \frac{72pn \log n}{(\sqrt{p''} - \sqrt{q''})^2} > \frac{6 \log n}{\frac{n}{|E|}(\sqrt{p''} - \sqrt{q''})^2} \geq \frac{6 \log n}{(\sqrt{p_1} - \sqrt{p_2})^2}.$$

We remark that if  $q = p/2$ , then number of rounds becomes  $r = O(n \log n)$ . First, observe that (34) implies that  $p \geq 2^{12} \log n/n$  and  $\Delta \leq p/2$ . Hence,  $\sqrt{p''} - \sqrt{q''} \geq \sqrt{\frac{p}{2} - p\sqrt{12/2^{12}} - p/12} - \sqrt{\frac{p}{4} + \frac{p}{2}\sqrt{12/2^{12}} + p/24} \geq \frac{\sqrt{p}}{2^{11}}$  and hence  $r = O(n \log n)$ .

This completes the proof.  $\square$

*Proof of Theorem 5.2.* The proof from Lemma F.2 together Lemma F.3 and applying Union bound over at most  $6 \log n$  rounds.  $\square$

**Theorem F.4** ([17, Equation 10]). *Let  $Y = \sum_{i=1}^m Y_i$  be the sum of  $m$  i.i.d. random variables with  $\mathbb{P}[Y_i = 1] = p$  and  $\mathbb{P}[Y_i = 0] = 1 - p$ . We have for any  $\alpha \in (0, 1)$  that*

$$\mathbb{P}[Y \geq \alpha \cdot m] \leq \left( \left( \frac{p}{\alpha} \right)^\alpha \left( \frac{1-p}{1-\alpha} \right)^{1-\alpha} \right)^m.$$

**Theorem F.5** (Chernoff bound [26, Theorem 4.4 and 4.5]). *Let  $X = \sum_i X_i$  be the sum of 0/1 independent random variables. Then,*

1. for any  $\delta > 0$ ,

$$\mathbb{P}[X \geq (1 + \delta)\mathbb{E}[X]] < \left( \frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^{\mathbb{E}[X]}.$$

2. for any  $0 < \delta \leq 1$ ,

$$\mathbb{P}[X \geq (1 + \delta)\mathbb{E}[X]] \leq e^{-\mathbb{E}[X]\delta^2/3}.$$

3. for  $R \geq 6\mathbb{E}[X]$ ,

$$\mathbb{P}[X \geq R] \leq 2^{-R}.$$

4. for  $0 < \delta < 1$ ,

$$\mathbb{P}[X \leq (1 - \delta)\mathbb{E}[X]] \leq \left( \frac{e^{-\delta}}{(1 - \delta)^{1-\delta}} \right)^{\mathbb{E}[X]}.$$

5. for  $0 < \delta < 1$ ,

$$\mathbb{P}[X \leq (1 - \delta)\mathbb{E}[X]] \leq e^{-\mathbb{E}[X]\delta^2/2}.$$