Community Recovery in Graphs with Locality
— Supplemental Materials —

Yuxin Chen∗ Govinda Kamath† Changho Suh‡ David Tse§

Abstract

This supplemental document presents details concerning (1) analytical derivations that support the
theorems made in the main text “Community Recovery in Graphs with Locality”. One can find here a
discussion of Algorithm 1 for multi-linked samples, as well as detailed proofs of all main theorems.

1 Main Theorems

We repeat the algorithm and the main theorems as follows for convenience of presentation.

**Theorem 1.** Fix $\theta > 0$ and any small $\epsilon > 0$. Let $G$ be a ring $R_r$ with connectivity radius $r \gtrsim \log^3 n$. Then with probability approaching one, Algorithm 1 converges to the exact solution within $T = O(\log n)$ iterations, provided that the sample size
\[ m \geq (1 + \epsilon) m^*, \]
where
\[ m^* = \frac{n \log n}{2(1 - e^{-\text{KL}(0.5\|\theta)})}. \]

Conversely, if $m < (1 - \epsilon) m^*$, then the probability of error $P_e(\psi)$ is approaching one for any algorithm $\psi$.

**Theorem 2.** Theorem 1 continues to hold for the following families of measurement graphs:
(1) Lines with $r = n^\beta$ for some constant $0 < \beta < 1$, where
\[ m^* = \frac{(1 + o(1)) \max \{\frac{1}{2}, \beta\} n \log n}{1 - e^{-\text{KL}(0.5\|\theta)}}; \]
(2) Grids with $r = n^\beta$ for some constant $0 < \beta < 0.5$, where
\[ m^* = \frac{(1 + o(1)) \max \{\frac{1}{2}, 2\beta\} n \log n}{1 - e^{-\text{KL}(0.5\|\theta)}}. \]
(3) Lines with $r = \gamma n$ for some constant $0 < \gamma \leq 1$, where
\[ m^* = \frac{(1 + o(1)) \max \{1 - \frac{1}{\gamma}, \gamma\} n \log n}{1 - e^{-\text{KL}(0.5\|\theta)}}. \]

**Theorem 3.** Theorem 1 and Theorem 2(1)(2) continue to hold under the nonuniform sampling model, provided that $\max_{(i,j) \in E} w_{i,j}$ is bounded by some large constant.

∗Y. Chen is with the Department of Statistics, Stanford University, Stanford, CA 94305, USA (email: yxchen@stanford.edu).
†G. Kamath is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA (email: gkamath@stanford.edu).
‡C. Suh is with the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea (e-mail: chsuh@kaist.ac.kr).
§D. Tse is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA (email: dntse@stanford.edu).
Algorithm 1

1. Run spectral method (Algorithm 3 of [1]) on a core subgraph induced by \( V_c \), which yields estimates \( X_j^{(0)}, 1 \leq j \leq |V_c| \).

2. Progressive estimation: for \( i = |V_c| + 1, \ldots, n \),

\[
X_i^{(0)} \leftarrow \text{majority} \left\{ Y_{i,j}^{(l)} \oplus X_j^{(0)} \mid j < i, (i,j) \in E, 1 \leq l \leq N_{i,j} \right\}.
\]

3. Successive local refinement: for \( t = 0, \ldots, T-1 \),

\[
X_i^{(t+1)} \leftarrow \text{majority} \left\{ Y_{i,j}^{(l)} \oplus X_j^{(t)} \mid j \neq i, (i,j) \in E, 1 \leq l \leq N_{i,j} \right\}, \quad 1 \leq i \leq n.
\]

4. Output \( X_i^{(T)}, 1 \leq i \leq n \).

Here, \( \text{majority} \{ \cdot \} \) represents the majority voting rule: for any sequence \( s_1, \ldots, s_k \in \{0,1\} \), \( \text{majority} \{s_1, \ldots, s_k\} \) is equal to 1 if \( \sum_{i=1}^k s_i > k/2 \); and 0 otherwise.

\*

Theorem 4. Theorem 1 continues to hold under the small-world graph model stated in the main text.

Theorem 5. Fix \( L \geq 2 \). Theorem 1 continues to hold under the \( L \)-wise sampling model introduced in the main text, with \( m^* \) replaced by

\[
m^* := \frac{n \log n}{L (1 - e^{-D(P_0,P_1)})},
\]

where

\[
P_0 := (1-p) \text{Binomial} (L - 1, 1-p) + p \text{Binomial} (L - 1, p);
\]

\[
P_1 := (1-p) \text{Binomial} (L - 1, p) + p \text{Binomial} (L - 1, 1-p).
\]

Remark 1. \( D(P_0,P_1) \) can be expressed in closed form as

\[
D(P_0,P_1) = -\log \left\{ \sum_{i=0}^{L-1} \binom{L-1}{i} \left[ p^i (1-p)^{L-i} + (1-p)^i p^{L-1} \right] \left[ p^{i+1} (1-p)^{L-i-1} + (1-p)^{i+1} p^{L-i-1} \right] \right\}.
\]

When \( L = 2 \), this reduces to

\[
D(P_0,P_1) = -\log \left\{ 2 \sqrt{(1-p)^2 + p^2} (2p (1-p)) \right\} = -\log \left\{ 2 \sqrt{1 - \theta} \right\} = \text{KL}(0.5 \| \theta)
\]

for \( \theta := 2p(1-p) \).

The analyses for all cases follow almost the same arguments. As a result, we separate the proofs into two parts: (1) the minimax lower bound, and (2) the achievability of Algorithm 1. Each part accommodates all cases considered in the paper.

2 Preliminaries

Before continuing, we gather a few facts that will be useful throughout. First of all, recall that the maximum likelihood (ML) decision rule achieves the lowest Bayesian probability of error, assuming uniform prior over two hypotheses of interest. The resulting error exponent is determined by the Chernoff information, as given in the following lemma.

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*
Lemma 1. Fix any small $\epsilon > 0$. Suppose we observe a collection of $N_z$ random variables $Z = \{Z_1, \cdots, Z_{N_z}\}$ that are i.i.d. given $N_z$. Consider two hypotheses $H_0$: $Z_i \sim P_0$ and $H_1$: $Z_i \sim P_1$ for two given probability measures $P_0$ and $P_1$. Assume that the Chernoff information $D^* = D(P_0, P_1) > 0$ and the alphabet of $Z_i$ are both finite and fixed, and that $\max_z \frac{P_1(z)}{P_0(z)} < \infty$.

(a) Suppose $N_z$ is sufficiently large. Then one has

$$\exp \left\{- (1 + \epsilon) N_z D^* \right\} \leq P_0 \left( \frac{P_1(Z)}{P_0(Z)} \geq 1 \bigg| N_z \right) \leq \exp \left\{- N_z D^* \right\}. \quad (7)$$

(b) If $N_z \sim \text{Poisson}(N)$ for some sufficiently large quantity $N$, then

$$\exp \left\{- (1 + \epsilon) N \left(1 - e^{-D^*} \right) \right\} \leq P_0 \left( \frac{P_1(Z)}{P_0(Z)} \geq 1 \bigg| N \right) \leq \exp \left\{- N \left(1 - e^{-D^*} \right) \right\}. \quad (8)$$

Proof. See Appendix A

We emphasize that the best achievable error exponent coincides with the Chernoff information $D^*$ when the sample size is fixed, while it becomes $1 - e^{-D^*}$ when the sample size is Poisson distributed.

Next, we consider the robustness of the ML test. In particular, we control the probability of error when the ML decision boundary is slightly shifted, as stated below.

Lemma 2. Consider any $\epsilon > 0$, and let $N \sim \text{Poisson}(\lambda)$.

(a) Fix any $0 < \theta < 0.5$. Conditional on $N$, draw $N$ independent samples $Z_1, \cdots, Z_N$, where $Z_i \sim \text{Bernoulli}(\theta)$, $1 \leq i \leq N$. Then one has

$$\Pr \left\{ \sum_{i=1}^{N} Z_i \geq \frac{1}{2} N - \epsilon \theta \right\} \leq \exp \left( \epsilon \cdot 2 \log \frac{1 - \theta}{\theta} \right) \exp \left\{- \lambda \left(1 - e^{-\text{KL}(0.5|\theta)} \right) \right\}. \quad (9)$$

(b) Let $P_0$ and $P_1$ be two distributions obeying $\max_z \frac{P_1(z)}{P_0(z)} < \infty$. Conditional on $N$, draw $N$ independent samples $Z_i \sim P_0$ ($1 \leq i \leq N$). Then one has

$$P_0 \left( \sum_{j=1}^{N} \log \frac{P_j(Z_i)}{P_j(Z_i)} \geq -\epsilon \theta \right) \leq \exp (\epsilon \lambda) \exp \left\{- \lambda \left(1 - e^{-D^*} \right) \right\}, \quad (10)$$

where $D^* = D(P_0, P_1)$.

Proof. See Appendix B

In addition, the following lemma develops an upper bound on the tail of Poisson random variables.

Lemma 3. Suppose that $N \sim \text{Poisson}(\epsilon \lambda)$ for some $0 < \epsilon < 1$. Then for any $c_1 > 2\epsilon$, one has

$$\Pr \left\{ N \geq \frac{c_1 \lambda}{\log \frac{1}{\epsilon}} \right\} \leq 2 \exp \left\{ - \frac{c_1 \lambda}{2} \right\}. \quad (11)$$

Proof. See Appendix C

Our analysis also relies on the well-known Chernoff-Hoeffding inequality [6, Theorem 1, Eqn (2.1)].

Lemma 4 (Chernoff-Hoeffding Inequality). Suppose $Z_1, \cdots, Z_n$ are independent Bernoulli random variables with mean $E[Z_i] \leq p$. Then for any $q \geq p$, one has

$$\Pr \left\{ \frac{1}{n} \sum_{j=1}^{n} Z_j \geq q \right\} \leq \exp \left\{ - n \text{KL}(q \| p) \right\},$$

where $\text{KL}(q \| p) := q \log \frac{q}{p} + (1 - q) \log \frac{1 - q}{1 - p}$.
We end this subsection with a lower bound on the KL divergence between two Bernoulli distributions.

**Fact 1.** For any $0 \leq q \leq \tau \leq 1$,

$$\text{KL}(\tau \| q) := \tau \log \frac{\tau}{q} + (1 - \tau) \log \frac{1 - \tau}{1 - q} \geq \tau \log (\tau/q) - \tau.$$  

**Proof.** By definition,

$$\text{KL}(\tau \| q) \overset{(i)}{=} \tau \log \frac{\tau}{q} + (1 - \tau) \log (1 - \tau) \overset{(ii)}{=} \tau \log (\tau/q) - \tau,$$

where (i) follows since $\log \frac{1}{1 - q} \geq 0$, and (ii) arises since $(1 - \tau) \log (1 - \tau) \geq - (1 - \tau) \tau \geq -\tau$. \qed

Here and below, we will denote

$$D^* = D(P_0, P_1),$$

where $P_0$ (resp. $P_1$) denotes the distribution of $Y$ given $X_1 = 0$ (resp. $X_1 = 1$), with $Y$ indicating any (multi-linked) sample that involves $X_1$. In particular, for the pairwise case one has

$$D^* = D(\text{Bernoulli}(\theta), \text{Bernoulli}(1 - \theta)) = \text{KL}(0.5 \| \theta).$$

The characterization of $D^*$ beyond the pairwise case is deferred to Section 5.2.

### 3 Minimax Lower Bound

The key ingredient of the proof is to establish the following lemma.

**Lemma 5.** Fix any constant $\epsilon > 0$, and consider any vertex subset $V_1 \subseteq V$ with $|V_1| \geq n^\epsilon$. Let $\tilde{d}$ denote the maximum vertex degree within $V_1$. If

$$\lambda \tilde{d} \leq (1 - \epsilon) \frac{\log |V_1|}{1 - e^{-D^*}},$$

then the probability of error $\inf_\psi P_e(\psi) \to 1$ as $n \to \infty$.

**Proof.** See Appendix D \qed

We are now in position to demonstrate how Lemma 5 leads to tight lower bounds.

- **Rings.** For rings, take $V_1 = \{1, \cdots, n\}$ for some small constant $\epsilon > 0$, which obeys $\tilde{d} = d_{\text{avg}}$. Applying Lemma 5 leads to a necessary recovery condition

$$\lambda d_{\text{avg}} \geq (1 - \epsilon) \frac{\log n}{1 - e^{-D^*}}.$$  

Since $m = \frac{1}{2} \lambda nd_{\text{avg}}$, this is equivalent to

$$m \geq (1 - \epsilon) \cdot \frac{n \log n}{2(1 - e^{-D^*})}.$$  

- **Lines with $r = n^\beta$ ($0 < \beta < 1$).** Take $V_1 = \{1, \cdots, cn^\beta\}$ for some small constant $\epsilon > 0$, which obeys $\tilde{d} = (1 + o(1)) (1 + \epsilon) \frac{d_{\text{avg}}}{2^\beta}$. According to Lemma 5, a necessary recovery condition is

$$\lambda \tilde{d} \geq (1 - \epsilon) \frac{\log |V_1|}{1 - e^{-D^*}}.$$  

$$\iff \lambda d_{\text{avg}} \geq (1 + o(1)) \frac{1 - \epsilon}{1 + \epsilon} \frac{2(\beta \log n + \log \epsilon)}{1 - e^{-D^*}}.$$
In addition, if we take \( V_1 = V \), then \( \tilde{d} = (1 + o(1)) \) \( d_{\text{avg}} \). Lemma 5 leads to another necessary condition:

\[
\lambda d_{\text{avg}} \geq (1 + o(1)) (1 - \epsilon) \cdot \frac{\log n}{1 - e^{-D^*}}.
\]

Combining these conditions and recognizing that \( \epsilon \) can be arbitrary, we arrive at the following necessary condition

\[
\lambda d_{\text{avg}} \geq (1 - \epsilon) \max \left\{ 2\beta, 1 \right\} \frac{n \log n}{1 - e^{-D^*}},
\]

which is equivalent to

\[
m = \frac{1}{2} n \lambda d_{\text{avg}} \geq (1 - \epsilon) \max \left\{ \beta, \frac{1}{2} \right\} \frac{n \log n}{1 - e^{-D^*}}.
\]

**Lines with** \( r = \gamma n \) \((0 < \gamma \leq 1)\). Take \( V_1 = \{1, \ldots, \epsilon \gamma n\} \) for some small constant \( \epsilon > 0 \), which obeys \( \tilde{d} = \epsilon \gamma n + r = (1 + O(\epsilon)) r \). According to Lemma 5, a necessary recovery condition is

\[
\lambda \tilde{d} \geq (1 - \epsilon) \cdot \frac{\log |V_1|}{1 - e^{-D^*}} \iff \lambda r \geq \frac{1 - \epsilon}{1 + O(\epsilon)} \cdot \frac{2 (\log n + \log (\epsilon \gamma))}{1 - e^{-D^*}}.
\]

(13)

On the other hand, the total number of edges in \( G \) is given by

\[
\frac{1 + o(1)}{2} \left( n^2 - (n - r)^2 \right) = (1 + o(1)) nr \left( 1 - \frac{1}{2} \frac{r}{n} \right),
\]

and hence

\[
d_{\text{avg}} = (1 + o(1)) \left( 1 - \frac{1}{2} \gamma \right) r.
\]

(14)

This taken collectively with (13) and \( m = \frac{1}{2} nd_{\text{avg}} \lambda \) establishes the necessary recovery condition

\[
m \geq (1 - O(\epsilon)) \left( 1 - \frac{1}{2} \gamma \right) \frac{n \log n}{1 - e^{-D^*}}.
\]

This completes the proof for this case since \( \epsilon \) can be arbitrary.

**Grids with** \( r = n^\beta \) \((0 < \beta < 1)\). Pick \( V_1 = \{1, \ldots, \epsilon n^\beta\} \) for some small constant \( \epsilon > 0 \), which obeys \( \tilde{d} = (1 + \epsilon) \frac{d_{\text{avg}}}{4} \). According to Lemma 5, a necessary recovery condition is

\[
\lambda \tilde{d} \geq (1 - \epsilon) \cdot \frac{\log |V_1|}{1 - e^{-D^*}} \iff \lambda d_{\text{avg}} \geq \frac{1 - \epsilon}{1 + \epsilon} \cdot \frac{8 (\beta \log n + \log \epsilon)}{1 - e^{-D^*}}.
\]

In addition, by taking \( V_1 = V \) one has \( \tilde{d} = d_{\text{avg}} \). Lemma 5 then leads to another necessary condition:

\[
\lambda d_{\text{avg}} \geq (1 - \epsilon) \cdot \frac{\log n}{1 - e^{-D^*}}.
\]

Putting these conditions together we derive

\[
\lambda d_{\text{avg}} \geq (1 - O(\epsilon)) \max \left\{ 8\beta, 1 \right\} \frac{n \log n}{1 - e^{-D^*}},
\]

which in turn gives

\[
m = \frac{1}{2} n d_{\text{avg}} \lambda \geq (1 - O(\epsilon)) \max \left\{ 4\beta, \frac{1}{2} \right\} \frac{n \log n}{1 - e^{-D^*}}.
\]
Furthermore, Lemma 5 immediately extends to the nonuniform sampling case with slight modification. To be precise, define the weighted degree of a node $v$ as

$$d_v^w := \sum_{i: (i,v) \in E} w_{i,v},$$

then the necessary recovery condition is given as follows.

**Lemma 6.** Consider the nonuniform sampling model introduced in the main text, and suppose that $\frac{\max_{(i,j) \in E} w_{i,j}}{\min_{(i,j) \in E} w_{i,j}}$ is bounded. Lemma 5 continues to hold, provided that $\tilde{d}$ is defined as the maximum weighted degree within $\mathcal{V}_1$.

**Proof.** See Appendix D.

This lemma allows us to accommodate the following graphs under nonuniform sampling.

- **Lines / rings / grids under nonuniform sampling.** The preceding proof continues to hold under the nonuniform sampling model with the assistance of Lemma 6, provided that $d_{avg}$ is replaced with the average weighted degree.

- **Small-world graphs.** The above argument for rings is applicable for small-world graphs as well, as long as $d_{avg}$ is replaced by the average weighted degree.

Finally, the above results immediately extend to the case with multi-linked samples.

**Lemma 7.** Lemma 5 continues to the hold for multi-linked samples, provided that the degree of each node $v$ is defined to be the number of hyper-edges incident to $v$.

**Proof.** See Appendix D.

When specialized to rings, setting $\mathcal{V}_1 = \{1, \ldots, n\}$ with $\tilde{d} = d_{avg}$ leads to the necessary condition

$$\lambda d_{avg} \geq (1 - \epsilon) \frac{\log n}{1 - e^{-D^*}},$$

where $d_{avg}$ represents the average degree defined with respect to the hyper-edges. Since each hyper-edge covers $L$ vertices, one sees that

$$m = \frac{1}{L} n \lambda d_{avg} \geq (1 - \epsilon) \frac{n \log n}{L (1 - e^{-D^*})},$$

where we have accounted for the over-count factor $L$. This establishes the lower bound for the case with multi-linked samples.

### 4 Optimal Performance Guarantees of Algorithm 1

Without loss of generality, we will assume $X_1 = \cdots = X_n = 0$ throughout this section. For simplicity of presentation, we will focus on the boundary case where $m \approx n \log n$, but all arguments easily extend to the regime where $m \gg n \log n$.

#### 4.1 Stage 1 gives approximate recovery for $\mathcal{G}_c$

According to [1, Theorem 1.6], the spectral method is successful in recovering a portion $1 - \frac{1}{2} \epsilon$ of variables in $\mathcal{V}_c$ with high probability, as long as $\lambda d_{avg}$ exceeds some sufficiently large constant. In our case, this condition is easily satisfied since

$$\lambda \gtrsim \frac{\log n}{d_{avg}} \implies \lambda d_{avg} \gtrsim \log n.$$
4.2 Stage 2 yields approximate recovery for $\mathcal{V}_c$\(\setminus\mathcal{V}_c\)

For concreteness, we start by establishing the achievability for lines and rings, which already contain all important ingredients for proving more general cases.

4.2.1 Lines / rings

We divide all vertices in $\mathcal{V}_c\setminus\mathcal{V}_c$ into small groups $\{\mathcal{V}_i\}$, each consisting of $\epsilon\log^3 n$ adjacent vertices:\footnote{Note that the errors occurring to distinct vertices are statistically dependent in the progressive estimation stage. The approach we propose is to look at a group of vertices simultaneously, and to bound the fraction of errors happening within this group. In order to exhibit sufficiently sharp concentration, we pick the group size to be at least $\epsilon\log^3 n$. A smaller group is possible via more refined arguments.}

$$\mathcal{V}_i := \{|\mathcal{V}_c| + (i-1)\epsilon\log^3 n + 1, \ldots, |\mathcal{V}_c| + i\epsilon\log^3 n\}.$$ 

In what follows, we will control the estimation errors happening within each group. For notational simplicity, we let $\mathcal{V}_0 := \mathcal{V}_c$. An important vertex set for the progressive step, denoted by $\mathcal{V}_0$, is the one encompassing all vertices preceding and connected to $\mathcal{V}_i$.

The proof is recursive, which mainly consists in establishing the claim below. To state the claim, we need to introduce a collection of events as follows

$$A_0 := \{\text{at most a fraction } \frac{\epsilon}{2} \text{ of progressive estimates } \left\{X_j^{(0)} : j \in \mathcal{V}_c\right\} \text{ is incorrect}\};$$

$$A_i := \{\text{at most a fraction } \epsilon \text{ of progressive estimates } \left\{X_j^{(0)} : j \in \mathcal{V}_i\right\} \text{ is incorrect}\}, \quad i \geq 1.$$

**Lemma 8.** For any $i \geq 0$, conditional on $A_0 \cap \cdots \cap A_i$, one has

$$\mathbb{P}\{A_{i+1} | A_0 \cap \cdots \cap A_i\} \geq 1 - O\left(n^{-2}\right). \quad (16)$$

As a result, one has

$$\mathbb{P}\{\cap_{i \geq 0} A_i\} \geq 1 - o(1). \quad (17)$$

Apparently, $A_0$ holds with high probability; see the analysis for Stage 1. Thus, if $(16)$ holds, then $(17)$ follows immediately from the union bound. In fact, $(17)$ suggests that for any group $\mathcal{V}_i$, only a small fraction of estimates obtained in this stage are incorrect, thus justifying approximate recovery for this stage. Moreover, since the neighborhood $\mathcal{N}(v)$ of each node $v \in \mathcal{V}_i$ is covered by at most $O\left(\frac{d_{\text{avg}}}{|\mathcal{V}_i|}\right)$ groups, the event $\cap_{i \geq 0} A_i$ leads the following observation that will be useful for analyzing Stage 3:

- There are no more than $O\left(\epsilon \cdot |\mathcal{V}_i|\right) O\left(\frac{d_{\text{avg}}}{|\mathcal{V}_i|}\right) = O\left(\epsilon d_{\text{avg}}\right)$ errors occurring to either the neighborhood $\mathcal{N}(v)$ or the backward neighborhood $\mathcal{N}(v) \cap \mathcal{V}_{i-1}$.

The rest of the section is devoted to establish the claim $(16)$.

**Proof of Lemma 8** As discussed above, it suffices to prove $(16)$ (which in turn justifies $(17)$). The following argument is conditional on $A_0 \cap \cdots \cap A_i$ and all estimates for $\mathcal{V}_0 \cup \cdots \cup \mathcal{V}_i$; we suppress the notation by dropping this conditional dependence whenever clear from the context.

Consider any vertex $u \in \mathcal{V}_{i+1}$. In the progressive estimation step, each $X_u^{(0)}$ relies on the preceding estimates $\left\{X_j^{(0)} : j < u, (j,u) \in \mathcal{E}\right\}$, as well as the set $\mathcal{B}_u$ of backward samples $\left\{Y_{u,j}^{(l)} : j < u, (j,u) \in \mathcal{E}, 1 \leq l \leq \mathcal{N}_{u,j}\right\}$ incident to $u$. We divide $\mathcal{B}_u$ into two parts

- $\mathcal{B}_{u,\text{good}}^u$: the set of samples $Y_{u,j}^{(l)}$ in $\mathcal{B}_u$ such that (i) $X_j^{(0)} = X_j$, and (ii) $j \in \mathcal{V}_{i+1}$;

- $\mathcal{B}_{u,\text{bad}}^u$: the remaining samples $\mathcal{B}_u \setminus \mathcal{B}_{u,\text{good}}^u$.

4.3.3 Additivity

The following additive property is the key to proving $(16)$. Given an event $A \subseteq \{1,2,\ldots,N\}$, define

$$\mathcal{P}(A) := \frac{|A|}{N}.$$
In words, favoring the truth exceeds wrong labels in for some universal constant \( D \) and we define the complement event as

\[
\text{It is self-evident to check that the above success condition would hold if}
\]

\[
\text{and set}
\]

\[
N_u^{\text{good}} := |B_u^{\text{good}}| \quad \text{and} \quad N_u^{\text{bad}} := |B_u^{\text{bad}}|.
\]

In words, \( B_u^{\text{good}} \) is associated with those preceding estimates in \( V_{i \rightarrow (i+1)} \) that are consistent with the truth, while \( B_u^{\text{bad}} \) entails the rest of the samples that might be unreliable. The purpose of this partition is to separate \( B_u^{\text{bad}} \), which only accounts for a small fraction of all samples.

We now proceed to analyze the majority voting procedure; by definition, it succeeds if the total votes favoring the truth exceeds \( \frac{1}{2} \left( N_u^{\text{good}} + N_u^{\text{bad}} \right) \). To preclude the effect of \( B_u^{\text{bad}} \), we pay particular attention to the part of votes obtained over \( B_u^{\text{good}} \), that is, the partial score

\[
\text{score}^\text{good}_u := \sum_{Y^{(l)}_{u,j} \in B_u^{\text{good}}} X^{(0)}_j \oplus Y^{(l)}_{u,j}.
\]

It is self-evident to check that the above success condition would hold if

\[
\text{score}^\text{good}_u < \frac{1}{2} \left( N_u^{\text{good}} + N_u^{\text{bad}} \right) - |B_u^{\text{bad}}| = \frac{1}{2} N_u^{\text{good}} - \frac{1}{2} N_u^{\text{bad}},
\]

and we define the complement event as

\[
D_u := \left\{ \text{score}^\text{good}_u \geq \frac{1}{2} N_u^{\text{good}} - \frac{1}{2} N_u^{\text{bad}} \right\}.
\]

The main point to work with \( D_u \) is that conditional on all prior estimates in \( V_{i \rightarrow (i+1)} \), the \( D_u \)'s are independent across all \( u \in V_{i+1} \). Furthermore, we decouple \( D_u \) into two events:

\[
\mathbb{P}\{D_u\} \leq \mathbb{P}\left\{ N_u^{\text{bad}} \geq \frac{c_0 \log n}{\log \frac{1}{\epsilon}} \right\} + \mathbb{P}\left\{ \text{score}^\text{good}_u \geq \frac{1}{2} N_u^{\text{true}} - \frac{1}{2} \frac{c_0 \log n}{\log \frac{1}{\epsilon}} \right\}
\]

for some universal constant \( c_0 > 0 \).

Recall that each edge is sampled at a Poisson rate \( \lambda \asymp \frac{m}{d_{\text{avg}}} \asymp \frac{\log n}{d_{\text{avg}}} \asymp \frac{\log n}{d_{\text{avg}}^{\text{true}}} \), and that the average number of samples connecting \( u \) and other nodes in \( V_{i+1} \) is \( \lambda \cdot O(\epsilon d_{\text{avg}}) \). On the event \( A_0 \cap \cdots \cap A_i \), the number of wrong labels in \( V_{i \rightarrow (i+1)} \) is \( O(\epsilon d_u) \), and hence

\[
\mathbb{E}\left[ N_u^{\text{bad}} \right] \leq O(\lambda \epsilon d_u) \leq \epsilon c_2 \log n
\]

\[\Box\]
for some constant $c_2 > 0$. This further gives
\[
\mathbb{E} \left[ N_u^{\text{good}} \right] \geq \lambda c_3 d_u - \mathbb{E} \left[ N_u^{\text{bad}} \right] \geq (1 - c_4 \epsilon) c_3 \lambda d_u
\]
for some constants $c_3, c_4 > 0$. Thus, Lemma 3 and the inequality (19) taken collectively yield
\[
\Pr \left\{ N_u^{\text{bad}} \geq \frac{c_1 c_2 \log n}{\log \frac{1}{\epsilon}} \right\} \leq 2 \exp \left\{ -\frac{c_1 c_2 \log n}{2} \right\}
\]
for any $c_1 > 2\epsilon$. In addition, in view of Lemma 2, there exists some function $\tilde{\xi}(\cdot)$ such that
\[
\Pr \left\{ \text{score}_u^{\text{good}} \geq \frac{1}{2} N_u^{\text{true}} - \frac{1}{2} c_0 \log n \right\} \leq \exp \left\{ - (1 - o_n(1)) \left( 1 - \tilde{\xi}(\epsilon) \right) c_3 \lambda d_u \left( 1 - e^{-D^*} \right) \right\}
\]
where $\tilde{\xi}(\epsilon)$ is independent of $n$ and vanishes as $\epsilon \to 0$. Putting these bounds together reveals that: when
\[
\lambda = \Theta \left( \frac{\log n}{d_{\text{avg}}} \right) \quad \text{and} \quad c_0 = c_1 c_2,
\]
there exists some function $\tilde{\xi}(\epsilon)$ independent of $n$ such that
\[
\Pr \left\{ \right\} \leq 2 \exp \left\{ -\frac{c_1 c_2 \log n}{2} \right\} + \exp \left\{ - (1 - o_n(1)) \left( 1 - \tilde{\xi}(\epsilon) \right) \frac{\lambda d_{\text{avg}}}{2} \left( 1 - e^{-D^*} \right) \right\}
\]
where $\tilde{\xi}(\epsilon)$ vanishes as $\epsilon \to 0$.

This in turn allows us to control the number of incorrect estimates within the group $\mathcal{V}_{i+1}$. Specifically,
\[
\Pr \left\{ \right\} \leq \exp \left\{ -\frac{c_1 c_2 \log n}{2} \right\} + \exp \left\{ - (1 - o_n(1)) \left( 1 - \tilde{\xi}(\epsilon) \right) \frac{\lambda d_{\text{avg}}}{2} \left( 1 - e^{-D^*} \right) \right\}
\]
where (a) follows from Lemma 3, (b) arises from Fact 4, and (c) is a consequence of (20) together with the assumption that $d_{\text{avg}} \gtrsim \log^2 n$.

So far we have demonstrated that the fraction of incorrect estimates for $\mathcal{V}_i$ is vanishingly small with probability exceeding $1 - O \left( n^{-2} \right)$. This establishes the claim concerning $\mathcal{A}_{i+1}$ when conditional on $\mathcal{A}_0 \cap \cdots \cap \mathcal{A}_i$, which in turn justifies approximate recovery for $\mathcal{V}$.

\[\square\]

4.2.2 Beyond lines / rings

The preceding analysis only relies on very few properties of lines / rings, and can be readily applied to many other graphs. In fact, all arguments continue to hold as long as the following assumptions are satisfied:

1. Each vertex $v$ ($v > |\mathcal{V}_c|$) is connected with at least $\Theta(d_{\text{avg}})$ vertices in $\{1, \cdots, i-1\}$;

2. For any $v \in \mathcal{V}_i$ ($i \geq 1$), its backward neighborhood $\mathcal{N}(v) \cap \mathcal{V}_{\rightarrow i}$ is covered by at most $O \left( \frac{d_{\text{avg}}}{|\mathcal{V}_{\rightarrow i}|} \right) = O \left( \frac{d_{\text{avg}}}{\epsilon \log^2 n} \right)$ distinct groups among $\mathcal{V}_1, \cdots, \mathcal{V}_{i-1}$.

In short, the first condition ensures that the information of a diverse range of prior vertices can be propagated to each $v$, whereas the second condition guarantees that the estimation errors are fairly spread out within the backward neighborhood of each node.
We are now in position to look at grids as well as lines / rings with nonuniform weights.

(a) It is straightforward to verify that the choices of $V_c$ and the ordering of $V$ isolated in Section ?? satisfy Conditions 1-2, which in turn establishes approximate recovery results for grids.

(b) Suppose that $\frac{\max_{(i,j) \in E} w_{i,j}}{\min_{(i,j) \in E} w_{i,j}}$ is bounded. Define the weighted degree as

$$d_w^v := \sum_{i : (i,v) \in E} w_{i,v}$$

and let the average weighted degree be $d_{\text{avg}}^w := \frac{1}{n} \sum d_w^v$. Then all arguments continue to hold if $d_{\text{avg}}^w$ is replaced by $d_w^v$. This reveals approximate recovery for lines (with $r = n^\beta$ for some constant $0 < \beta < 1$) / rings / grids under nonuniform sampling schemes.

(c) For the case with multi-linked samples, we redefine several metrics as follows:

- $B_u$: the set of backward samples $\{Y_e^{(l)} \mid u \in e, j < u \text{ for all other } j \in e, 1 \leq l \leq N_c\}$, where $e$ represents the hyper-edge;

- $B_{\text{good}}^u$: the set of samples $Y_e^{(l)}$ in $B_u$ such that (i) $X^{(0)}_j = X_j$ for all $j \in e$ and $j \neq u$, and (ii) $j \in V \rightarrow (i+1)$ for all $j \in e$ and $j \neq u$;

- $B_{\text{bad}}^u$: the remaining samples $B_u \setminus B_{\text{good}}^u$.

We also need to re-define the score $\text{score}_u^{\text{good}}$ as

$$\text{score}_u^{\text{good}} := \sum_{Y_e^{(l)} \in B_{\text{good}}^u} \log \frac{\mathbb{P} \{ Y_e^{(l)} \mid X_u = 1, X_i = X^{(0)}_i (1 \leq i < u) \}}{\mathbb{P} \{ Y_e^{(l)} \mid X_u = 0, X_i = X^{(0)}_i (1 \leq i < u) \}}$$

with the decision boundary being 0.

With these metrics in place, all proof arguments for the basic setup carry over to the multi-linked sample case.

### 4.3 Stage 3 achieves exact recovery

We now turn to the last stage, and the goal is to prove that $X^{(l)}$ converges to $X$ within $O(\log n)$ iterations. Before proceeding, we introduce a few more notations that will be used throughout.

- For any vertex $v$, denote by $N(v)$ the neighborhood of $v$ in $G$, and let $S(v)$ be the set of samples that involve $v$;

- For any vector $Z = [Z_1, \cdots, Z_n]^{\top}$ and any set $\mathcal{I} \subseteq \{1, \cdots, n\}$, define the $\ell_0$ norm restricted to $\mathcal{I}$ as follows

$$\|Z\|_{0,\mathcal{I}} := \sum_{i \in \mathcal{I}} 1_{\{Z_i \neq 0\}}.$$

- Generalize the definition of the majority vote operator such that

$$\text{majority}_v (Z) := [\text{majority}_1(Z_1), \cdots, \text{majority}_n(Z_n)]^{\top}$$

obtained by applying $\text{majority}_v (\cdot)$ component-wise, where

$$\text{majority}_v (Z_v) := \begin{cases} 1, & \text{if } Z_v \geq \frac{1}{2} |S(v)|; \\ 0, & \text{else}. \end{cases}$$

- Let $V_Z$ (resp. $V_X$) denote the local voting scores using $Z = [Z_i]_{1 \leq i \leq n}$ (resp. $X = [X_i]_{1 \leq i \leq n} = 0$) as the current estimates, i.e. for any $1 \leq u \leq n$,

$$V_Z(u) = \sum_{y_i^{(l)} \in S(u)} y_i^{(l)} \oplus Z_i; \quad V_X(u) = \sum_{y_i^{(l)} \in S(u)} y_i^{(l)} \oplus X_i = \sum_{y_i^{(l)} \in S(u)} y_i^{(l)}.$$  

(21)
With these notations in place, the iterative procedure can be succinctly written as
\[ X^{(t+1)} = \text{majority} (V_{X^{(t)}}). \]

The main subject of this section is to prove the following theorem.

**Theorem 6.** Consider any \( 0 < \epsilon \leq \epsilon_0 \), where \( \epsilon_0 \) is some sufficiently small constant. Define
\[
Z_{\epsilon} := \left\{ Z \in \{0,1\}^n \mid \forall v : \|Z - X\|_{0,N(v)} \leq \epsilon d_v \right\}.
\] (23)

Then with probability approaching one,
\[
\text{majority} (V_Z) \in Z_{\frac{1}{2} \epsilon}, \quad \forall Z \in Z_{\epsilon} \text{ and } \forall \epsilon \geq \frac{1}{d_{\text{max}}}. \]

**Remark 2.** When the iterate falls within the set \( Z_{\epsilon} \) (cf. (23)), the neighborhood of each vertex suffers from a small number of errors. This essentially implies that (i) the fraction of estimation errors is low; (ii) the estimation errors are fairly spread out instead of clustering within a few nodes’ neighborhoods.

**Remark 3.** This is a uniform result: it holds regardless of whether \( Z \) is statistically independent of the samples \( Y \) or not.

Note that the subscript of \( Z_{\epsilon} \) indicates the fraction of estimation errors in an iterate. From the preceding analyses, we know that Stage 3 is seeded with some initial guess \( X^{(0)} \in Z_{\epsilon} \) for some arbitrarily small constant \( \epsilon > 0 \). This taken collectively with Theorem 6 gives rise to the following error contraction result: for any \( t \geq 0 \),
\[
\|X^{(t+1)} - X\|_{0,N(v)} = \|\text{majority} (V_{X^{(t)}}) - X\|_{0,N(v)} \leq \frac{1}{2} \|X^{(t)} - X\|_{0,N(v)}, \quad 1 \leq v \leq n. \] (24)

This reveals the geometric convergence rate of \( X^{(t)} \), namely, \( X^{(t)} \) converges to the truth within \( O(\log n) \) iterations.

The rest of this section is devoted to proving Theorem 6. We will start by proving the result any fixed candidate \( Z \in Z_{\epsilon} \) independent of the samples, and then generalize it to all \( Z \in Z_{\epsilon} \). Our strategy is to first quantify \( V_X \) (which corresponds to the score we obtain when only a single vertex is uncertain), and then control the difference between \( V_X \) and \( V_Z \). The first observation is that all entries of \( V_X \) are strictly below the voting boundary, as asserted by the following lemma.

**Lemma 9.** For any small constant \( \delta > 0 \), one has
\[
(V_x) u < \frac{1}{2} |S(u)| - \delta \log n, \quad 1 \leq u \leq n
\]

with probability approaching one, provided that the following conditions are satisfied:

1. **Rings with \( r \geq \log^2 n \):**
   \[
m > (1 + \xi (\delta)) \frac{n \log n}{2 (1 - e^{-KL(0.5 \| \theta)})};
   \]

2. **Lines with \( r = n^\beta \) for some constant \( 0 < \beta < 1 \):**
   \[
m > (1 + \xi (\delta)) \max \left\{ \beta, \frac{1}{2} \right\} \frac{n \log n}{1 - e^{-KL(0.5 \| \theta)}};
   \]

3. **Lines with \( r = \gamma n \) for some constant \( 0 < \gamma \leq 1 \):**
   \[
m > (1 + \xi (\delta)) \left( 1 - \frac{1}{2} \gamma \right) \frac{n \log n}{1 - e^{-KL(0.5 \| \theta)}};
   \]

4. **Grids with \( r = n^\beta \) for some constant \( 0 < \beta < 1/2 \):**
   \[
m > (1 + \xi (\delta)) \max \left\{ 4 \beta, \frac{1}{2} \right\} \frac{n \log n}{1 - e^{-KL(0.5 \| \theta)}};
   \]
(5) Small-world graphs described in Section 4.1:

\[ m > (1 + \bar{\xi}(\delta)) \frac{n \log n}{2 (1 - e^{-\text{KL}(0.5\|\theta)})}; \]

(6) Multi-linked reads described in Section 4.2:

\[ m > (1 + \bar{\xi}(\delta)) \frac{n \log n}{2 (1 - e^{-D^*})}. \]

In all these cases, \( \bar{\xi}(\cdot) \) is some function independent of \( n \) satisfying \( \bar{\xi}(\delta) \to 0 \) as \( \delta \to 0 \). Here, we allow Cases (1), (2) and (4) to have nonuniform sampling rate over different edges, as long as \( \max_{i,j \in E} \frac{\log w_{i,j}}{\min_{i,j \in E} w_{i,j}} \) is bounded.

Proof. See Appendix E.

It remains to control the difference between \( V_X \) and \( V_Z \):

\[ \Delta_Z := V_Z - V_X. \]

Specifically, we would like to demonstrate that most entries of \( \Delta_Z \) are bounded in magnitude by \( \delta \log n \), so that most of the perturbations are absolutely controlled. To facilitate analysis, we decouple the statistical dependency by writing

\[ V_Z = F_Z + B_Z, \]

where \( F_Z \) represents the votes using only forward samples, namely,

\[ (F_Z)_u = \sum_{i > u, y^{(l)}_{i,u} \in S(u)} y^{(l)}_{i,u} \oplus Z_i, \quad 1 \leq u \leq n. \]

This is more convenient to work with since the entries of \( F_Z \) (or \( B_Z \)) are jointly independent. In what follows, we will focus on bounding \( F_Z \), but all arguments immediately apply to \( B_Z \). To simplify presentation, we also decompose \( V_X \) into two parts \( V_X = F_X + B_X \) in the same manner.

Note that the \( v \)th entry of the difference

\[ \Delta^F := F_Z - F_X \]

is generated by those entries from indices in \( \mathcal{N}(v) \) satisfying \( Z_v \neq X_v \). From the assumption \( (23) \), each \( \Delta^F_v \) (1 ≤ \( v \) ≤ \( n \)) is dependent on at most \( O(\epsilon_d v) \) non-zero entries of \( Z - X \), and hence on average each \( \Delta^F_v \) is only affected by \( O(\lambda \cdot d_{\text{avg}}) \) samples. Moreover, each non-zero entry of \( Z - X \) is bounded in magnitude by 2. This together with Lemma 3 yields that: for any sufficiently large constant \( c_1 > 0 \),

\[ \mathbb{P} \left\{ |\Delta^F| \geq 2c_1 \frac{\lambda d_{\text{avg}}}{\log \frac{1}{\epsilon}} \right\} \leq 2 \exp \left\{ -\Theta(c_1 \lambda d_{\text{avg}}) \right\} \leq 2n^{-c_2}, \]

provided that \( \lambda d_{\text{avg}} \geq \log n \) (which is the regime of interest), where \( c_2 = \Theta(c_1) \) is some absolute positive constant.

The preceding bound only concerns a single component. In order to obtain overall control, we introduce a set of independent indicator variables \( \{\eta_i(Z)\} \):

\[ \eta_i(Z) := \begin{cases} 1, & \text{if } |\Delta^F| \geq \frac{c_1 \lambda d_{\text{avg}}}{\log(1/\epsilon)}, \\ 0, & \text{else}. \end{cases} \]

For any 1 ≤ \( v \) ≤ \( n \), applying Lemma 4 gives

\[ \mathbb{P} \left\{ \frac{1}{d_v} \sum_{i \in \mathcal{N}(v)} \eta_i(Z) \geq \tau \right\} \leq \exp \left\{ -d_v \text{KL} \left( \tau \| \max_i \mathbb{E}[\xi_i] \right) \right\} \]

\[ \leq \exp \left\{ -d_v \left( \tau \log \frac{\tau}{2n^{-c_2}} - \tau \right) \right\}. \]
where the last line follows from Fact 1 as well as (26). For any $\tau \geq 1/n$,

$$
\tau \log \frac{\tau}{2n - c_2} - \tau \gtrsim \tau \log n,
$$

indicating that

$$
P \left\{ \frac{1}{d_v} \sum_{i \in \mathcal{N}(v)} \eta_i(Z) \geq \tau \right\} \leq \exp \{-c_3 \tau d_{\text{avg}} \log n\}
$$

for some universal constant $c_3 > 0$.

We are now in position to derive the more uniform type of results. Suppose that $d_{\text{max}} = Kd_{\text{avg}}$. When restricted to $Z_{\epsilon}$, the neighborhood of each $v$ has at most $(Kd_{\text{avg}}/\epsilon Kd_{\text{avg}})$ different values. If we set $\tau = 1/\epsilon$, then in view of the union bound,

$$
P \left\{ \exists Z \in Z_{\epsilon} \text{ s.t. } \frac{1}{d_v} \sum_{i \in \mathcal{N}(v)} \eta_i(Z) \geq \tau \right\} \leq \left( \frac{Kd_{\text{avg}}}{\epsilon Kd_{\text{avg}}} \right) \exp \{-c_3 \tau d_{\text{avg}} \log n\}
$$

$$
\leq \left( Kd_{\text{avg}} \right)^{\epsilon Kd_{\text{avg}}} \exp \{-c_3 \tau d_{\text{avg}} \log n\}
$$

$$
\leq \exp \left\{ (1 + o(1)) \epsilon Kd_{\text{avg}} \log n \exp \{-o(1) \epsilon d_{\text{avg}} \log n\} \right\}
$$

$$
\leq \exp \left\{ - \frac{1}{4} c_3 - (1 + o(1)) K \epsilon d_{\text{avg}} \log n \right\}.
$$

Since $Z, X \in \{0,1\}^n$, it suffices to consider the case where $\epsilon \in \left\{ \frac{1}{d_v} | 1 \leq v \leq n, 1 \leq i \leq d_v \right\}$, which has at most $O(n^2)$ distinct values. Set $c_3$ to be sufficiently large and apply the union bound (over both $v$ and $\epsilon$) to deduce that: with probability exceeding $1 - \exp \left( -\Theta(\epsilon d_{\text{avg}} \log n) \right) = 1 - n^{-O(1)}$,

$$
\text{card} \left\{ i \in \mathcal{N}(v) : |\Delta_i^F| \geq \frac{c_1 \lambda d_{\text{avg}}}{\log \frac{1}{\epsilon}} \right\} \leq \frac{1}{4} \epsilon d_v, \quad 1 \leq v \leq n, \quad (27)
$$

holds simultaneously for all $Z \in Z_{\epsilon}$ and all $\epsilon \geq \frac{1}{d_{\text{max}}} \geq \frac{1}{d_{\text{avg}}}.$

The uniform bound (27) continues to hold if $\Delta_i^F$ is replaced by $\Delta_i^B$. Putting these together suggests that with probability exceeding $1 - \exp \left( -\Theta(\epsilon d \log n) \right)$,

$$
\text{card} \left\{ i \in \mathcal{N}(v) : |(\Delta Z)_i| \geq \frac{c_1 \lambda d_{\text{avg}}}{\log \frac{1}{\epsilon}} \right\} \leq \frac{1}{2} \epsilon d_v, \quad 1 \leq v \leq n
$$

holds simultaneously for all $Z \in Z_{\epsilon}$ and all $\epsilon \geq \frac{1}{d_{\text{max}}}.$

Taking $\delta$ to be $2c_1/\log \frac{1}{\epsilon}$ in (43), we see that all but $\frac{1}{2} \epsilon d_v$ entries of $V_Z = V_X + \Delta Z$ at indices from $\mathcal{N}(v)$ exceed the voting boundary. Consequently, after the majority voting one has

$$
||\text{majority}(V_Z) - X||_{0,\mathcal{N}(v)} \leq \frac{1}{2} \epsilon d_v, \quad 1 \leq v \leq n
$$

or, equivalently,

$$
\text{majority}(V_Z) \in Z_{\frac{1}{2} \epsilon}, \quad \forall Z \in Z_{\epsilon}
$$

as claimed.

The proof also extends with minor modification to the multi-linked sample case with $L$ being a constant, as long as the voting scores are replaced with appropriate local log likelihood ratio score. We omit the details for conciseness.
Algorithm 2

1. Break each $L$-wise sample $Y_e = (Y_{i_1}, \cdots, Y_{i_L})$ into $\binom{L}{2}$ pairwise samples of the form $Y_{i_j} \oplus Y_{i_l}$ (for all $j \neq l$), and run spectral method (Algorithm 3 of [1]) on a core subgraph induced by $\mathcal{V}_e$ using these parity samples. This yields estimates $X^{(0)}_e, 1 \leq j \leq |\mathcal{V}_e|.$

2. **Progressive estimation**: for $k = |\mathcal{V}_e| + 1, \cdots, n,$

$$X^{(0)}_k \leftarrow \text{local} - \text{ML}_{\{X^{(0)}_{i} \mid 1 \leq i < k\}} \left\{ Y_e^{(l)} \mid e = (i_1, \cdots, i_L) \text{ with } i_L = k, \ 1 \leq l \leq N_e \right\}.$$ 

3. **Successive local refinement**: for $t = 0, \cdots, T - 1,$

$$X^{(t+1)}_k \leftarrow \text{local} - \text{ML}_{\{X^{(0)}_{i} \mid i \neq k\}} \left\{ Y_e^{(l)} \mid k \in e, \ 1 \leq l \leq N_e \right\}, \ 1 \leq k \leq n.$$ 

4. **Output** $X^{(T)}_k, 1 \leq k \leq n.$

Here, $\text{local} - \text{ML}_{\{ \cdot \}}$ represents the local maximum likelihood rule: for any sequence $s_1, \cdots, s_N \in \{0,1\},$

$$\text{local} - \text{ML}_{\{X^{(0)}_{i} \mid 1 \leq i < k\}} \{ s_1, \cdots, s_N \} = \begin{cases} 1, \text{ if } \sum_{j=1}^{N} \log \frac{P\{s_j \mid X_k = 1, X_i = X^{(0)}_{i} (1 \leq i < k)\}}{P\{s_j \mid X_k = 0, X_i = X^{(0)}_{i} (1 \leq i < k)\}} \geq 0, \\ 0, \text{ else}, \end{cases}$$

and

$$\text{local} - \text{ML}_{\{X^{(0)}_{i} \mid i \neq k\}} \{ s_1, \cdots, s_N \} = \begin{cases} 1, \text{ if } \sum_{j=1}^{N} \log \frac{P\{s_j \mid X_k = 1, X_i = X^{(0)}_{i} (i \neq k)\}}{P\{s_j \mid X_k = 0, X_i = X^{(0)}_{i} (i \neq k)\}} \geq 0, \\ 0, \text{ else}. \end{cases}$$

5 Multi-Linked Samples

5.1 A Modification of Algorithm 1

Recall that the inputs of Algorithm [1] are parity samples, while in the multi-linked sampling case the measurements are $L$-wise samples of the form

$$Y_e = (Y_{i_1}, \cdots, Y_{i_L}),$$

where

$$Y_i = \begin{cases} X_i, \quad \text{with probability } 1 - p, \\ X_i \oplus 1, \quad \text{else} \end{cases}$$

for some crossover probability $0 < p < 0.5.$

In order to apply the first step of Algorithm 1, we convert each $L$-wise sample $Y_e = (Y_{i_1}, \cdots, Y_{i_L})$ into $\binom{L}{2}$ pairwise samples of the form $Y_{i_j} \oplus Y_{i_l}$ (for all $j \neq l$), and then apply the spectral method over these parity samples. For Step 2 and Step 3, we replace the local majority vote procedure by the local maximum likelihood rule. The modified algorithm is summarized in Algorithm [2].

5.2 Chernoff Information for Multi-Linked Samples

Suppose now that each vertex $v$ is involved in $N_v$ multi-linked samples or, equivalently, $N_v (L - 1)$ pairwise samples. Careful readers will note that these parity samples are not independent. The key step in dealing with such dependency is not to treat them as $N_v (L - 1)$ independent samples, but instead $N_v$ independent groups. Thus, it suffices to compute the Chernoff information associated with each group, as detailed below.
Without loss of generality, suppose only \( X_1 \) is uncertain and \( X_2 = \cdots = X_n = 0 \). Consider a multi-linked sample that covers \( X_1, \cdots, X_L \). According to our model, each \( L \)-wise sample is an independent copy of the following random vector:

\[
Y_c = \begin{cases} 
(Z_1, \cdots, Z_L), & \text{with probability } \frac{1}{2}, \\
(Z_1 \oplus 1, \cdots, Z_L \oplus 1), & \text{else},
\end{cases}
\]

where

\[ Z_i \overset{\text{ind.}}{=} \begin{cases} 
X_i, & \text{with probability } 1 - p, \\
X_i \oplus 1, & \text{else}.
\end{cases} \]

Since we never observe the global phase, a sufficient statistic for \( Y_c \) is given by

\[
\hat{Y}_c = (Z_1 \oplus Z_2, Z_1 \oplus Z_3, \cdots, Z_1 \oplus Z_L).
\]

From (11), the Chernoff information \( D^* \) is the large-deviation exponent when distinguishing between the conditional distributions of

\[
\hat{Y}_c \mid (X_1, \cdots, X_L) = (0, \cdots, 0) \quad \text{and} \quad \hat{Y}_c \mid (X_1, \cdots, X_L) = (1, \cdots, 0),
\]

which we discuss as follows.

- **When \( X_1 = \cdots = X_L = 0 \):**
  - if \( Z_1 = 0 \) (which occurs with probability \( 1 - p \)), then \( \hat{Y}_c \sim \text{Binomial}(L - 1, p) \);
  - if \( Z_1 = 1 \) (which occurs with probability \( p \)), then \( \hat{Y}_c \sim \text{Binomial}(L - 1, 1 - p) \);

- **When \( X_1 = 1 \) and \( X_2 = \cdots = X_L = 0 \):**
  - if \( Z_1 = 0 \) (which occurs with probability \( p \)), then \( \hat{Y}_c \sim \text{Binomial}(L - 1, p) \);
  - if \( Z_1 = 1 \) (which occurs with probability \( 1 - p \)), then \( \hat{Y}_c \sim \text{Binomial}(L - 1, 1 - p) \).

To summarize, one has

\[
\hat{Y}_c \mid (X_1, \cdots, X_L) = (0, 0, \cdots, 0) \sim (1 - p) \text{Binomial}(L - 1, p) + p \text{Binomial}(L - 1, 1 - p) := P_0;
\]
\[
\hat{Y}_c \mid (X_1, \cdots, X_L) = (1, 0, \cdots, 0) \sim p \text{Binomial}(L - 1, p) + (1 - p) \text{Binomial}(L - 1, 1 - p) := P_1.
\]

To derive a closed-form expression, we note that a random variable \( W_0 \sim P_0 \) obeys

\[
P_0(W_0 = i) = (1 - p) \binom{L - 1}{i} p^i (1 - p)^{L - i - 1} + p \binom{L - 1}{i - 1} (1 - p)^i p^{L - i - 1} = \binom{L - 1}{i} \left\{ p^i (1 - p)^{L - i} + (1 - p)^i p^{L - 1} \right\}.
\]

Similarly, if \( W_1 \sim P_1 \), then

\[
P_1(W_1 = i) = \binom{L - 1}{i} \left\{ p^{i + 1} (1 - p)^{L - i - 1} + (1 - p)^{i + 1} p^{L - i - 1} \right\}.
\]

Making use of the symmetry between \( P_0 \) and \( P_1 \) (i.e. \( P_0(W_0 = i) = P_1(W_1 = L - 1 - i) \)), one can easily verify that

\[
D^* = - \inf_{0 \leq \tau \leq 1} \log \left\{ \sum_z P_0^{1-\tau}(z) P_1^\tau(z) \right\}
\]

is achieved when \( \tau = 1/2 \), which gives

\[
D(P_0, P_1) = - \log \left\{ \sum_{z=0}^{L-1} \sqrt{P_0(W_0 = i) P_1(W_1 = i)} \right\}
\]

\[
= - \log \left\{ \sum_{i=0}^{L-1} \binom{L - 1}{i} \left\{ p^i (1 - p)^{L - i} + (1 - p)^i p^{L - 1} \right\} \left\{ p^{i + 1} (1 - p)^{L - i - 1} + (1 - p)^{i + 1} p^{L - i - 1} \right\} \right\}.
\]

Interestingly, this expression enjoys a very simple asymptotic limit, as stated below.
**Lemma 10.** Fix any $0 < p < 1/2$. The Chernoff information \( (32) \) satisfies
\[
\lim_{L \to \infty} D(P_0, P_1) = \text{KL} (0.5\|p) .
\] (33)

**Proof.** See Appendix F. □

**Remark 4.** The asymptotic limit [33] admits a simple interpretation. Consider the typical event where only \( X_1 \) is uncertain and \( X_2 = \cdots = X_n = 0 \). Conditional on \( Z_1 \), the \( L-1 \) parity samples \( (Z_1 \oplus Z_2, \cdots, Z_1 \oplus Z_L) \) are i.i.d., which reveals accurate information about \( Z_1 \oplus 0 \) in the regime where \( L \to \infty \) (by the law of large number). As a result, the uncertainty arises only because \( Z_1 \) is a noisy version of \( X_1 \), which behaves like passing \( X_1 \) through a binary symmetric channel with crossover probability \( p \). This essentially boils down to distinguishing between Bernoulli \( p \) (when \( X_1 = 0 \)) and Bernoulli \( (1-p) \) (when \( X_1 = 1 \)), for which the associated Chernoff information is known to be \( \text{KL} (0.5\|p) \).

## A Proof of Lemma 1

Let \( M \) be the alphabet size for \( Z_i \). The standard method of types result (e.g. [5] Chapter 2 and [2] Section 11.7-11.9) reveals that
\[
\frac{1}{(N_z + 1)^M} \exp \left\{ - \left(1 + \frac{\epsilon}{2} \right) N_z D^* \right\} \leq P_0 \left( \frac{P_1(Z)}{P_0(Z)} \geq 1 \bigg| N_z \right) \leq \exp \left\{ - N_z D^* \right\} ;
\] (34)
here, the left-hand side holds for sufficiently large \( N_z \), while the right-hand side holds for arbitrary \( N_z \) (see [4] Exercise 2.12 or [3] Theorem 1 and recognize the convexity of the set of types under consideration). Moreover, since \( D^* > 0 \) and \( M \) is fixed, one has \( \frac{1}{(N_z + 1)^M} = \exp (-M \log (N_z + 1)) \geq \exp \left\{ - \frac{1}{2} \epsilon N_z D^* \right\} \) for any sufficiently large \( N_z \), thus indicating that
\[
P_0 \left( \frac{P_1(Z)}{P_0(Z)} \geq 1 \bigg| N_z \right) \geq \exp \left\{ - (1 + \epsilon) N_z D^* \right\}
\] (35)
as claimed.

We now move on to the case where \( N_z \sim \text{Poisson} (N) \). Employing (35) we arrive at
\[
P_0 \left( \frac{P_1(Z)}{P_0(Z)} \geq 1 \right) = \sum_{l=0}^{\infty} P(N_z = l) P_0 \left( \frac{P_1(Z)}{P_0(Z)} \geq 1 \bigg| N_z = l \right)
\] (36)
\[
\geq \sum_{l=0}^{\infty} \frac{N^l e^{-N}}{l!} \exp \left\{ - (1 + \epsilon) l D^* \right\}
\] (37)
\[
= e^{-(N - N_0)} \sum_{l=\tilde{N}}^{\infty} \frac{N_0^l \exp (-N_0)}{l!}
\] (38)
for any sufficiently large \( \tilde{N} \), where we have introduced \( N_0 := Ne^{-(1+\epsilon)D^*} \). Furthermore, taking \( \tilde{N} = \log N_0 \) we obtain
\[
\sum_{l=\tilde{N}}^{\infty} \frac{N_0^l}{l!} \exp (-N_0) = 1 - \sum_{l=0}^{\tilde{N}} \frac{N_0^l}{l!} \exp (-N_0) \geq 1 - \sum_{l=0}^{\tilde{N}} N_0^l \exp (-N_0)
\]
\[
\geq 1 - (\tilde{N} + 1) N_0^{\tilde{N}} \exp (-N_0)
\]
\[
= 1 - (\log N_0 + 1) N_0^{\log N_0} \exp (-N_0) = 1 - o_N (1)
\]
\[
\geq 0.5
\]
as long as \( N \) is sufficiently large. Substitution into (38) yields
\[
P_0 \left( \frac{P_1(Z)}{P_0(Z)} \geq 1 \right) \geq 0.5 e^{-(N - N_0)} \geq \exp \left\{ - (1 + \epsilon) N \left( 1 - e^{-(1+\epsilon)D^*} \right) \right\} .
\] (39)
This finishes the proof of the lower bound in (8) since $\epsilon > 0$ can be arbitrary.

Additionally, applying the upper bound (34) we derive

$$
\sum_{l=0}^{\infty} Ne^{-N} \cdot e^{-lD^*} = \exp \left( -N \left( 1 - e^{-D^*} \right) \right),
$$

establishing the upper bound (8).

**B Proof of Lemma 2**

We start with the general case, and suppose that the Chernoff information (31) is attained by $\tau = \tau^*$. It follows from the Chernoff bound that

$$
P_0 \left\{ \sum_{i=1}^{N} \log \frac{P_1(Z_i)}{P_0(Z_i)} \geq -\epsilon \lambda \mid N = k \right\} = P_0 \left\{ \tau^* \sum_{i=1}^{k} \log \frac{P_1(Z_i)}{P_0(Z_i)} \geq -\epsilon \lambda \right\} \leq \frac{\prod_{i=1}^{k} \mathbb{E}_{P_0} \left[ \left( \frac{P_1(Z_i)}{P_0(Z_i)} \right)^{\tau^*} \right]}{\exp \left( -\epsilon \lambda \right)}
$$

$$
= \exp \left( \epsilon \lambda \right) \left( \mathbb{E}_{P_0} \left[ \left( \frac{P_1(Z_i)}{P_0(Z_i)} \right)^{\tau^*} \right] \right)^k
$$

$$
= \exp \left( \epsilon \lambda \right) \left( \sum_{z} P_1^{1-\tau^*}(z) P_0^{1-\tau^*}(z) \right)^k
$$

$$
\leq \exp \left( \epsilon \lambda \right) \exp \left( -kD^* \right).
$$

This suggests that

$$
P_0 \left\{ \sum_{i=1}^{N} \log \frac{P_1(Z_i)}{P_0(Z_i)} \geq -\epsilon \lambda \right\} = P_0 \left\{ \sum_{i=1}^{N} \log \frac{P_1(Z_i)}{P_0(Z_i)} \geq -\epsilon \lambda \mid N = k \right\} \mathbb{P} \{ N = k \}
$$

$$
\leq \exp \left( \epsilon \lambda \right) \mathbb{E}_{N \sim \text{Poisson} \left( \lambda \right)} \left[ \exp \left( -ND^* \right) \right]
$$

$$
= \exp \left( \epsilon \lambda \right) \exp \left\{ -\lambda \left( 1 - e^{-D^*} \right) \right\},
$$

where the last identity follows from the moment generating function of Poisson variables. This establishes the claim for the general case.

When specialized to the Bernoulli case, one has

$$
\log \frac{P_1(Z_i)}{P_0(Z_i)} = I \{ Z_i = 0 \} \log \frac{\theta}{1 - \theta} + I \{ Z_i = 1 \} \log \frac{1 - \theta}{\theta}
$$

$$
= \{ 2I \{ Z_i = 1 \} - 1 \} \log \frac{1 - \theta}{\theta}.
$$

This leads to the following equivalence

$$
\sum_{i=1}^{N} \log \frac{P_1(Z_i)}{P_0(Z_i)} \geq -\epsilon \lambda \quad \iff \quad \sum_{i=1}^{N} I \{ Z_i = 1 \} \geq \frac{1}{2} N - \frac{\epsilon \lambda}{2 \log \frac{1 - \theta}{\theta}}.
$$

Recognizing that $\sum_{i=1}^{N} Z_i = \sum_{i=1}^{N} I \{ Z_i = 1 \}$ and replacing $\epsilon$ with $\epsilon \cdot 2 \log \frac{1 - \theta}{\theta}$, we complete the proof.
C Proof of Lemma 3

For any constant $c_1 \geq 2e,$

$$
P\left\{ N \geq \frac{c_1\lambda}{\log \left( \frac{1}{\epsilon} \right)} \right\} = \sum_{k \geq \frac{c_1\lambda}{\log \left( \frac{1}{\epsilon} \right)}} P\{ N = k \} = \sum_{k \geq \frac{c_1\lambda}{\log \left( \frac{1}{\epsilon} \right)}} \frac{(e\lambda)^k}{k!} \exp \left( -e\lambda \right)
$$

\[\leq (i) \sum_{k \geq \frac{c_1\lambda}{\log \left( \frac{1}{\epsilon} \right)}} \frac{(e\lambda)^k}{(k/e)^k} \leq \sum_{k \geq \frac{c_1\lambda}{\log \left( \frac{1}{\epsilon} \right)}} \left( \frac{e\lambda}{k} \right)^k \leq \sum_{k \geq \frac{c_1\lambda}{\log \left( \frac{1}{\epsilon} \right)}} \left( \frac{e\lambda}{c_1} \right)^k \leq 2 \exp \left( -\log \left( \frac{c_1}{e} \right) \cdot \frac{1}{\sqrt{\epsilon}} \right) \leq 2 \exp \left\{ -\log \left( \frac{1}{\sqrt{\epsilon}} \right) \cdot \frac{c_1\lambda}{\log \left( \frac{1}{\epsilon} \right)} \right\} = 2 \exp \left\{ -\frac{c_1\lambda}{2} \right\},
\]

where (i) arises from the elementary inequality $a! \geq \left( \frac{a}{2} \right)^a$, (ii) holds because $e \log \left( \frac{1}{\epsilon} \right) \leq \sqrt{\epsilon}$ holds for any $0 < \epsilon \leq 1$, and (iii) follows due to the inequality $\sum_{k \geq K} a^k \leq \frac{a^K}{1-a} \leq 2a^K$ as long as $0 < a \leq 1/2$.

D Proof of Lemmas 5, 6, 7

(1) We start by proving Lemma 5, which contains all ingredients for proving Lemmas 5, 6, 7. First of all, we demonstrate that there are many vertices in $V_1$ that are isolated in the subgraph induced by $V_1$. In fact, let $\bar{V}_2$ be a random subset of $V_1$ of size $|\bar{V}_2| = \frac{|V_1|}{\log^4 n}$. By Markov’s inequality, the number of samples with both endpoints lying in $\bar{V}_2$, denoted by $N_{\bar{V}_2}$, is bounded above by

$$N_{\bar{V}_2} \leq \log n \cdot \mathbb{E}[|\mathcal{E}(\bar{V}_2, \bar{V}_2)|] \leq \lambda \left( \frac{1}{\log^6 n} \cdot |\mathcal{E}(V_1, V_1)| \right) \log n \leq \lambda \left( \frac{1}{\log^6 n} \cdot |V_1| \cdot \tilde{d} \right) \log n$$

\[\overset{(i)}{=} O \left( \frac{|V_1|}{\log^4 n} \right) = o(|V_2|) \]

with high probability, where (i) follows from the assumption $|V_2| = \frac{|V_1|}{\log^4 n}$. As a consequence, one can find $(1 - o(1)) |V_2|$ vertices in $\bar{V}_2$ that are involved in absolutely no sample falling within $\mathcal{E}(\bar{V}_2, \bar{V}_2)$. Let $\bar{V}_3$ be the set of these vertices, which obeys

$$\bar{V}_3 = (1 - o(1)) |V_2| = \frac{(1 - o(1)) |V_1|}{\log^4 n} \geq \frac{|V_1|}{2 \log^4 n}. \quad (40)$$

We emphasize that the discussion so far only concerns $\mathcal{E}(\bar{V}_2, \bar{V}_2)$, which is independent of samples taken over $\mathcal{E}(\bar{V}_2, \bar{V}_2)$.

Suppose the ground truth is $X_i = 0$ ($1 \leq i \leq n$). For each vertex $v \in \bar{V}_3$, construct a representative singleton hypothesis $X^v$ such that $X_i^v = 1$ if $i = v$. Let $P_0$ (resp. $P_\epsilon$) denote the output distribution given $X = 0$ (resp. $X = X^\epsilon$). Assuming a uniform prior over all candidates, it suffices to study the ML rule which achieves the best error exponent. For each $v \in \bar{V}_3$, since it is isolated in $\bar{V}_2$, all information useful for differentiating $X = X^v$ and $X = 0$ falls within the positions $v \times \bar{V}_2$, which in total account for at most $\tilde{d}$ entries. The main point is that for any $v, u \in \bar{V}_3$, the corresponding samples over $v \times \bar{V}_2$ and $u \times \bar{V}_2$ are statistically independent, and hence conditional on $\bar{V}_3$, the events $\left\{ \frac{P_v(Y)}{P_0(Y)} \geq 1 \right\}$ are independent across all $v \in \bar{V}_3$.

Conditional on $\bar{V}_3$, Lemma 1 suggests that

$$P_0 \left( \frac{P_v(Y)}{P_0(Y)} \geq 1 \right) \geq \exp \left\{ -\lambda \tilde{d} \left( 1 - e^{-D^\epsilon} \right) \right\}.$$
The conditional independence of events \( \{ \frac{dP_x}{dP_0} \geq 1 \} \) gives

\[
\Pr_e(\psi_{ml}) \geq 1 - \prod_{v \in V_3} \left\{ 1 - P_0 \left( \frac{P_v(Y)}{P_0(Y)} \geq 1 \right) \right\}^{\left| V_3 \right|}
\]

\[
\geq 1 - \left\{ 1 - \exp \left[ - (1 + o(1)) \lambda \tilde{d} \left( 1 - e^{-D^*} \right) \right] \right\}^{\left| V_3 \right| / 2 \log^3 n} \quad (41)
\]

\[
\geq 1 - \exp \left\{ - (1 + o(1)) \lambda \tilde{d} \left( 1 - e^{-D^*} \right) \right\} \frac{\left| V_1 \right|}{2 \log^3 n} \quad (42)
\]

where (42) comes from (40), and the last inequality follows since \( 1 - x \leq \exp(-x) \).

To summarize, ML fails with probability approaching one if one has

\[
\exp \left\{ - (1 + o(1)) \lambda \tilde{d} \left( 1 - e^{-D^*} \right) \right\} \frac{\left| V_1 \right|}{\log^3 n} \to \infty,
\]

which would hold if

\[
\lambda \tilde{d} \leq (1 - \epsilon) \frac{\log \left| V_1 \right|}{1 - e^{-D^*}}.
\]

(2) Now we turn to Lemma 6. The preceding argument immediately carries over to the nonuniform sampling case, as long as all vertex degrees are replaced with corresponding weighted degrees (cf. (15)).

(3) Finally, the preceding argument is still valid for proving Lemma 7, provided that the degree of each node \( v \) is replaced by the number of hyperedges (those linking \( L \) vertices) that involve \( v \).

### E Proof of Lemma 9

It follows from Lemma 2 that for any small constant \( \delta > 0 \)

\[
\Pr \left\{ \left( V_X \right)_v \geq \frac{1}{2} |S(v)| - \delta \lambda d_v \right\} \leq \exp \left\{ - (1 - o(1)) (1 - \xi(\delta)) \lambda d_v \left( 1 - e^{-D^*} \right) \right\},
\]

where \( D^* \) represents the Chernoff information.

Recall that \( \lambda d_v \geq \log n \). This together with the union bound and a little manipulation gives

\[
\Pr \left\{ \exists 1 \leq v \leq n : \left( V_X \right)_v \geq \frac{1}{2} |S(v)| - \delta \log n \right\} \leq \sum_{v=1}^{n} \exp \left\{ - (1 - o(1)) (1 - \tilde{\xi}(\delta)) \lambda d_v \left( 1 - e^{-D^*} \right) \right\} \quad (43)
\]

for some function \( \tilde{\xi}(\delta) \) that vanishes as \( \delta \to 0 \). We can now analyze different sampling models on a case-by-case basis.

(1) **Rings.** All vertices have the same degree \( d_{avg} \), and hence

\[
\text{(43)} \leq n \exp \left\{ - (1 - o(1)) d_{avg} \cdot (1 - \tilde{\xi}(\delta)) \lambda \left( 1 - e^{-D^*} \right) \right\},
\]

which tends to zero as long as

\[
\lambda d_{avg} \geq \frac{1 + \delta}{1 - \xi(\delta)} \cdot \log n \cdot \frac{1}{1 - e^{-D^*}}.
\]

Since the sample complexity is \( m = \frac{1}{2} \lambda nd_{avg} \), we arrive at

\[
m > \frac{1 + \delta}{1 - \xi(\delta)} \cdot \frac{n \log n}{2 (1 - e^{-D^*})}.
\]
(2) **Lines with** $r = n^\beta$ **for some constant** $0 < \beta < 1$. The first and the last $r$ vertices have degrees at least $(1 - o(1)) \frac{1}{2} d_{avg}$, while all remaining $n - 2r$ vertices have degrees equal to $(1 - o(1)) d_{avg}$. This gives
\[
\frac{1}{2} d_{avg} \geq 2r \cdot \exp \left\{ - \left(1 - o(1)\right) \left(1 - \tilde{\xi} \left(\delta\right)\right) \lambda \cdot \frac{1}{2} d_{avg} \left(1 - e^{-D^*}\right) \right\}
\]
\[
+ (n - 2r) \exp \left\{ - \left(1 - o(1)\right) \left(1 - \tilde{\xi} \left(\delta\right)\right) \lambda d_{avg} \left(1 - e^{-D^*}\right) \right\}
\]
\[
\leq 2 \exp \left\{ \beta \log n - \left(1 - o(1)\right) \left(1 - \tilde{\xi} \left(\delta\right)\right) \frac{1}{2} \lambda d_{avg} \left(1 - e^{-D^*}\right) \right\}
\]
\[
+ \exp \left\{ \log n - \left(1 - o(1)\right) \left(1 - \tilde{\xi} \left(\delta\right)\right) \lambda d_{avg} \left(1 - e^{-D^*}\right) \right\},
\]
which converges to zero as long as
\[
\lambda d_{avg} > \left(1 + o(1)\right) \frac{1 + \delta}{1 - \xi \left(\delta\right)} \cdot \frac{2\beta \log n}{1 - e^{-D^*}}.
\]
This taken collectively with the fact $m = \frac{1}{2} n \lambda d_{avg}$ establishes the claim for this case.

(3) **Lines with** $r = \gamma n$ **for some constant** $0 < \gamma \leq 1$. Each vertex has degree exceeding $(1 - o(1)) r$, indicating that
\[
\frac{1}{2} d_{avg} \geq n \exp \left\{ - \left(1 - o(1)\right) \left(1 - \tilde{\xi} \left(\delta\right)\right) \lambda r \left(1 - e^{-D^*}\right) \right\}
\]
\[
\leq \exp \left\{ \log n - \left(1 - o(1)\right) \left(1 - \tilde{\xi} \left(\delta\right)\right) \lambda r \left(1 - e^{-D^*}\right) \right\},
\]
which converges to zero as long as
\[
\lambda r > \frac{1 + \delta}{1 - \xi \left(\delta\right)} \cdot \frac{\log n}{1 - e^{-D^*}}.
\]
Plugging it into (43) and $m = \frac{1}{2} n \lambda d_{avg}$ we arrive at
\[
m > \frac{1 + \delta}{1 - \xi \left(\delta\right)} \left(1 - \frac{1}{2}\right) \cdot \frac{n \log n}{1 - e^{-D^*}}.
\]

(4) **Grids with** $r = n^\beta$ **for some constant** $0 < \beta < 1$. Note that $d_{avg} \propto r^2 = n^{2\beta}$. There are at least $n - \pi r^2$ vertices with degrees equal to $(1 - o(1)) d_{avg}$, while the remaining vertices have degree at least $(1 - o(1)) d_{avg}/4$. This gives
\[
\frac{1}{4} d_{avg} \geq \pi r^2 \cdot \exp \left\{ - \left(1 - o(1)\right) \left(1 - \tilde{\xi} \left(\delta\right)\right) \lambda \frac{d_{avg}}{4} \left(1 - e^{-D^*}\right) \right\}
\]
\[
+ (n - \pi r^2) \exp \left\{ - \left(1 - o(1)\right) \left(1 - \tilde{\xi} \left(\delta\right)\right) \lambda d_{avg} \left(1 - e^{-D^*}\right) \right\}
\]
\[
\leq 4 \exp \left\{ 2\beta \log n - \left(1 - o(1)\right) \left(1 - \tilde{\xi} \left(\delta\right)\right) \lambda \cdot \frac{d_{avg}}{4} \left(1 - e^{-D^*}\right) \right\}
\]
\[
+ \exp \left\{ \log n - \left(1 - o(1)\right) \left(1 - \tilde{\xi} \left(\delta\right)\right) \lambda d_{avg} \left(1 - e^{-D^*}\right) \right\},
\]
which vanishes as long as
\[
\lambda d_{avg} > \left(1 + o(1)\right) \frac{1 + \delta}{1 - \xi \left(\delta\right)} \cdot \frac{8\beta \log n}{1 - e^{-D^*}};
\]
\[
\lambda d_{avg} > \left(1 + o(1)\right) \frac{1 + \delta}{1 - \xi \left(\delta\right)} \cdot \frac{\log n}{1 - e^{-D^*}}.
\]
This together with the fact that $m = \frac{1}{4} n \lambda d_{avg}$ establishes the proof for this case.

Finally, for the cases of lines (with $r = n^\beta$ for some constant $0 < \beta < 1$) / rings / grids with non-uniform sampling rates, it suffices to replace $d_{avg}$ with the average weighted degree (see Section 4.2.2). The case of small-world graphs follows exactly the same argument as in the case of rings with nonuniform weights.
F Proof of Lemma 10

For notational convenience, set

\[ b_i := \sqrt{\left\{ p^i (1 - p)^{L-i} + (1-p)^i p^{L-i} \right\} \left\{ p^{i+1} (1 - p)^{L-i-1} + (1-p)^{i+1} p^{L-i-1} \right\}}. \tag{44} \]

For any \( i < \frac{L}{2} - \log L \), one can verify that

\[ p^i (1 - p)^{L-i} + (1-p)^i p^{L-i} = p^i (1 - p)^{L-i} \left\{ 1 + \left( \frac{p}{1-p} \right)^{L-2i} \right\} \]

\[ = (1 + o_L(1)) p^i (1 - p)^{L-i} \]

and

\[ p^{i+1} (1 - p)^{L-i-1} + (1-p)^{i+1} p^{L-i-1} = p^{i+1} (1 - p)^{L-i-1} \left\{ 1 + \left( \frac{p}{1-p} \right)^{L-2i-2} \right\} \]

\[ = (1 + o_L(1)) p^{i+1} (1 - p)^{L-i-1}. \]

These identities suggest that

\[
\sum_{i=0}^{L/2-\log L} \binom{L-1}{i} b_i = (1 + o_L(1)) \sum_{i=0}^{L/2-\log L} \binom{L-1}{i} \sqrt{\left\{ p^i (1 - p)^{L-i} \right\} \left\{ p^{i+1} (1 - p)^{L-i-1} \right\}} \\
= (1 + o_L(1)) \sqrt{p(1-p)} \sum_{i=0}^{L/2-\log L} \binom{L-1}{i} p^i (1 - p)^{L-i-1} \\
= (1 + o_L(1)) \sqrt{p(1-p)},
\]

where the last line makes use of the following fact.

**Fact 2.** Fix any \( 0 < p < 1/2 \). Then one has

\[
\sum_{i=0}^{L/2-\log L} \binom{L-1}{i} p^i (1 - p)^{L-i-1} = 1 - o_L(1).
\]

**Proof.** To simplify writing, we concentrate on the case where \( L \) is even. From the binomial theorem, we see that

\[
\sum_{i=0}^{L-1} \binom{L-1}{i} p^i (1 - p)^{L-i-1} = 1.
\]

(45)

Hence, it suffices to control \( \sum_{i=L/2-\log L+1}^{L-1} \binom{L-1}{i} p^i (1 - p)^{L-i-1} \). To this end, we first make the observation that

\[
\sum_{i=L/2-\log L+1}^{L/2-\log L-2} \binom{L-1}{i} p^i (1 - p)^{L-i-1} \leq (2 \log L) \max_{i \geq \frac{L}{2} - \log L + 1} \binom{L-1}{i} \left( \frac{p}{1-p} \right)^i (1-p)^{L-1} \\
\leq (2 \log L) \cdot \binom{L-1}{L/2} \left( \frac{p}{1-p} \right)^{\frac{L}{2}-\log L+1} (1-p)^{L-1} \\
\leq \left\{ \left( 2 \log L \right) \left( 1-p \right)^{2 \log L-3} \right\} \cdot 2L \left( p(1-p) \right)^{\frac{L}{2}-\log L+1} \\
\leq o_L(1) \cdot \left[ 2 \left( p(1-p) \right)^{\frac{L}{2}-\log L+1} \right]^L 
\]

\[ = o_L(1), \] (46)
where (i) comes from the inequalities \( \frac{(L - 1)}{L/2} \leq 2^{L - 1} \leq 2^L \), and (ii) holds because \( \log L (1 - p)^{2 \log L - 3} = o_L(1) \). The last identity is a consequence of the inequality

\[\sqrt{p(1 - p)} < 1/2 \quad (\forall p < 1/2),\]

as well as the fact that \( (p(1 - p))^{-\log L} \to 1 (L \to \infty) \) and hence

\[(p(1 - p))^{1/2 - \log L} = \sqrt{p(1 - p)} (p(1 - p))^{-\log L} < 1/2.\]

On the other hand, the remaining terms can be bounded as

\[
\sum_{i=\frac{L}{4} + \log L - 1}^{L-1} \binom{L - 1}{i} p^i (1 - p)^{L - i - 1} = \sum_{i=\frac{L}{4} + \log L - 1}^{L-1} \binom{L - 1}{L - i - 1} p^i (1 - p)^{L - i - 1} = \sum_{i=0}^{\frac{L}{4} - \log L} \binom{L - 1}{i} p^i (1 - p)^{L - i - 1} \cdot \left( \frac{p}{1-p} \right)^{L - 2i - 1}
\]

\[= o_L(1) \cdot \sum_{i=0}^{\frac{L}{4} - \log L} \binom{L - 1}{i} p^i (1 - p)^{L - i - 1}.
\]

Putting the above results together yields

\[
1 = \left( \sum_{i=0}^{\frac{L}{2} - \log L} + \sum_{i=\frac{L}{4} + \log L - 1}^{L-1} + \sum_{i=L/2 - \log L + 1}^{L/2 - \log L} \right) \binom{L - 1}{i} p^i (1 - p)^{L - i - 1} = (1 + o_L(1)) \sum_{i=0}^{\frac{L}{2} - \log L} \binom{L - 1}{i} p^i (1 - p)^{L - i - 1} + o_L(1),
\]

which in turn gives

\[
\sum_{i=0}^{\frac{L}{2} - \log L} \binom{L - 1}{i} p^i (1 - p)^{L - i - 1} = 1 - o_L(1)
\]

as claimed. \(\square\)

Following the same arguments, we arrive at

\[
\sum_{i=L/2 + \log L}^{L-1} \binom{L - 1}{i} b_i = (1 + o_L(1)) \sqrt{p(1 - p)}.
\]

Moreover,

\[
\sum_{i=L/2 - \log L + 1}^{L/2 + \log L - 1} \binom{L - 1}{i} b_i \leq \sum_{i=L/2 - \log L + 1}^{L/2 + \log L - 1} \binom{L - 1}{i} \left\{ p^i (1 - p)^{L - i} + (1 - p)^i p^{L - i} \right\}
\]

\[
+ \sum_{i=L/2 - \log L + 1}^{L/2 + \log L - 1} \binom{L - 1}{i} \left\{ p^{i+1} (1 - p)^{L - i - 1} + (1 - p)^{i+1} p^{L - i - 1} \right\}
\]

\[= O \left\{ \sum_{i=L/2 - \log L + 1}^{L/2 + \log L - 1} \binom{L - 1}{i} \left\{ p^i (1 - p)^{L - i - 1} \right\} \right\}
\]

\[= o_L(1),
\]
where the last line follows the same step as in the proof of Fact 2 (cf. (46)). Taken together these results lead to

$$\sum_{i=0}^{L-1} \binom{L-1}{i} b_i = \left\{ \sum_{i=0}^{L/2-\log L-1} + \sum_{i=L/2+\log L}^{L-1} \sum_{i=0}^{L-1} \binom{L-1}{i} b_i \right\} \left( \binom{L-1}{i} b_i \right)$$

$$= 2 \left( 1 + o_L(1) \right) \sqrt{p(1-p)},$$

thus demonstrating that

$$D(P_0, P_1) = -\log \left\{ 2 \left( 1 + o_L(1) \right) \sqrt{p(1-p)} \right\} = (1 + o_L(1)) \text{KL}(0.5\|p).$$

References


