Proceedings of Machine Learning Research vol 178:1–37, 2022 35th Annual Conference on Learning Theory

Robust Estimation for Random Graphs

Jayadev Acharya ACHARYA@CORNELL.EDU
School of Electrical and Computer Engineering, Cornell University

Ayush Jain AYJAIN@ENG.UCSD.EDU
School of Electrical and Computer Engineering, UCSD

Gautam Kamath G@CSAIL.MIT.EDU
Cheriton School of Computer Science, University of Waterloo

Ananda Theertha Suresh THEERTHA@GOOGLE.COM
Google Research, New York

Huanyu Zhang HZ388@CORNELL.EDU
Meta

Editors: Po-Ling Loh and Maxim Raginsky

Abstract

We study the problem of robustly estimating the parameter p of an Erdős-Rényi random graph on n nodes, where a γ fraction of nodes may be adversarially corrupted. After showing the deficiencies of canonical estimators, we design a computationally-efficient spectral algorithm which estimates p up to accuracy $\tilde{O}(\sqrt{p(1-p)}/n+\gamma\sqrt{p(1-p)}/\sqrt{n}+\gamma/n)$ for $\gamma < 1/60$. Furthermore, we give an inefficient algorithm with similar accuracy for all $\gamma < 1/2$, the information-theoretic limit. Finally, we prove a nearly-matching statistical lower bound, showing that the error of our algorithms is optimal up to logarithmic factors.

Keywords: random graphs, robust estimation, spectral algorithms

1. Introduction

Finding underlying patterns and structure in data is a central task in machine learning and statistics. Typically, such structures are induced by modelling assumptions on the data generating procedure. While they offer mathematical convenience, real data generally does not match with these idealized models, for reasons ranging from model misspecification to adversarial data poisoning. Thus for learning algorithms to be effective in the wild, we require methods that are *robust* to deviations from the assumed model.

With this motivation, we initiate the study of robust estimation for random graph models. Specifically, we will be concerned with the Erdős-Rényi (ER) random graph model (Gilbert, 1959; Erdős and Rényi, 1959).[1]

Definition 1 (Erdős-Rényi graphs) *The Erdős-Rényi random graph model on n nodes with parameter $p \in [0,1]$, denoted as $G(n,p)$, is the distribution over graphs on n nodes where each edge is present with probability p, independently of the other edges.*

1. This model was introduced by Gilbert (1959), simultaneously with the related $G(n,m)$ model of Erdős and Rényi (1959). Nevertheless, the community refers to both models Erdős-Rényi graphs.

We consider graphs generated according to the Erdős-Rényi random graph model, but which then have a constant fraction of their nodes *corrupted* by an adversary. When a node is corrupted, the adversary can arbitrarily modify its neighborhood. This setting is naturally motivated by social networks, where random graphs are a common modelling assumption (Newman et al., 2002). Even if a fraction of individuals in the network are malicious actors, we still wish to perform inference with respect to the regular users. Apart from adversarial settings, tools for robust analysis of graphs may also assist in addressing deficiencies of existing models, such as in model misspecification. For example, certain random graph models have been criticized for not capturing various statistics of real-world networks (Newman et al., 2002), and some notion of robustness may facilitate better modelling.

1.1. Problem Setup

Let $\gamma \in [0,1]$ denote the fraction of corrupted nodes, and $G \sim G(n,p)$ be a random graph, where p is unknown. Without loss of generality, we assume that the node set is $[n] := \{1,\ldots,n\}$. An adversary $\mathcal{A}$ is then given G, and is allowed to arbitrarily 'rewire' the edges adjacent to a set $B \subseteq [n]$ of nodes of size at most γn, resulting in a graph $\mathcal{A}(G)$. In other words, the adversary can change the status of any edge with at least one end point in B. We call B the set of *corrupted nodes*. We consider two kinds of adversaries.

- γ-*omniscient adversary*: The adversary knows the true value of the edge probability p and observes the realization of the graph $G \sim G(n,p)$. They then choose B and how to rewire its edges.

- γ-*oblivious adversary*: The adversary knows the true value of the edge probability p. They must choose B and the distribution of edges from B without knowing the realization G.

Note that the oblivious adversary is weaker than the omniscient adversary. Given a corrupted graph $\mathcal{A}(G)$, our goal is to output $\hat{p}\left(\mathcal{A}(G)\right)$, an estimate of the true edge probability p.

1.2. Results

We first analyze standard estimators from the robust statistics toolkit, and show that they provide sub-optimal rates. We then propose a computationally-efficient spectral algorithm to estimate p with improved rates. Finally, we prove a lower bound for this problem, showing that our algorithms are optimal up to logarithmic factors. We note that our upper bounds hold for omniscient adversaries, whereas the lower bounds are tight even against the weaker oblivious adversary.

1.2.1. Standard Robust Estimators and Natural Variants

At first glance, the problem appears deceptively simple, as our goal is to estimate a single univariate parameter p. A standard technique is the maximum likelihood estimator, which in this case is the empirical edge density. We call the following the *mean estimator*

$$\hat{p}_{\text{mean}}(\mathcal{A}(G)) = \frac{\text{\# of edges present in } \mathcal{A}(G)}{\binom{n}{2}}. \tag{1}$$

In robust statistics, the median often provides better guarantees than the mean. Let $\deg(i)$ denote the degree of node $i \in [n]$ in $\mathcal{A}(G)$. The *median estimator* is given by

$$\hat{p}_{\mathrm{med}}(\mathcal{A}(G)) = \frac{\mathrm{Median}\{\deg(1), \ldots, \deg(n)\}}{n-1}. \tag{2}$$

Absent corruptions (i.e., $\gamma = 0$), we have $\mathcal{A}(G) = G$. In this simple setting, the mean and median are both very accurate. Specifically, it is not hard to show that $|\hat{p}_{\mathrm{mean}}(G) - p| \leq O\left(\sqrt{p(1-p)}/n\right)$ and $|\hat{p}_{\mathrm{med}}(G) - p| \leq O\left(1/n\right)$ (Lemma 6). However, both estimators perform much worse under even mild corruption. In Lemma 7 we describe and analyze a simple oblivious adversary $\mathcal{A}$ such that both the mean and median estimator have $|\hat{p}(\mathcal{A}(G)) - p| \geq \gamma/2$. Note that if even a single node is corrupted (i.e., $\gamma = 1/n$), the "price of robustness" (informally, the additional error term(s) introduced in the corrupted setting) dominates the baseline $O(1/n)$ error in the uncorrupted setting.

The adversary against the mean and median estimators is easy to describe: either add or remove all edges incident to the nodes in B. This suggests the strategy of first pruning a set of $c\gamma n$ nodes with the largest and smallest degrees and then applying either the mean or median estimator to the resulting graph. These *prune-then-mean/median* algorithms are described in Algorithm 1. Despite this additional step, the pruned estimators are still deficient. We design an oblivious adversary such that the prune-then-median estimate satisfies $|\hat{p}(\mathcal{A}(G)) - p| \geq \Omega(\gamma)$ and the prune-then-mean estimate satisfies $|\hat{p}(\mathcal{A}(G)) - p| \geq \Omega(\gamma^2)$ (Theorem 9). Interestingly, we show the tightness of both these bounds, showing that prune-then-mean improves the error to $O\left(\gamma^2\right)$ (Theorem 8). These results are summarized in the theorem below.

Theorem 2 (Informal) *The price of robustness of the prune-then-mean/median estimators are $\Theta(\gamma^2)$ and $\Theta(\gamma)$, respectively.*

1.2.2. A SPECTRAL ALGORITHM FOR ROBUST ESTIMATION

Given the failings of the approaches described so far, it may appear that a $\mathrm{poly}(\gamma)$ cost for robustness may be unavoidable. Our main result is a computationally-efficient algorithm that bypasses this barrier.

Theorem 3 *Suppose $\gamma < 1/60$ and $p \in [0, 1]$. Let $G \sim G(n, p)$ and $\mathcal{A}(G)$ be a rewiring of G by a γ-omniscient adversary $\mathcal{A}$. There exists a polynomial-time estimator $\hat{p}(\mathcal{A}(G))$ such that with probability at least $1 - 10n^{-2}$,*

$$|\hat{p}(\mathcal{A}(G)) - p| \leq C \cdot \left(\frac{\sqrt{p(1-p)\log n}}{n} + \frac{\gamma\sqrt{p(1-p)\log(1/\gamma)}}{\sqrt{n}} + \frac{\gamma}{n}\log n \right),$$

for some constant C. This estimate can be computed in $\tilde{O}(\gamma n^3 + n^2)$ time.

The first term is the error without corruptions, while the other two terms capture the price of robustness. Except at extreme values of p, the last term will be dominated by one of the other two. In this case, note that the cost of robustness in the second term decreases as the number of nodes n increases. This is in contrast to the previously described approaches, for which the price of robustness did not decrease with n. Observe that the non-robust error will dominate for most regimes when $\gamma \le 1/\sqrt{n}$.

As our lower bounds will establish, our algorithm provides a nearly-tight solution to the problem. Note that while this algorithm requires knowledge of γ, Jain, Orlitsky, and Ravindrakumar (2022) recently proposed a simple argument which using Lepski's method generically removes the need to know the corruption parameter for robust estimation tasks, leading to such an algorithm with the same rates.

Our upper bound requires $\gamma < 1/60$.[2] On the other hand, note that if $\gamma \ge 0.5$, an identifiability argument implies that no estimator can achieve error better than 0.5.[3] This raises the question of whether the above rates are achievable for all $\gamma < 0.5$. We show that this is indeed the case, providing a computationally inefficient algorithm with the following guarantees.

Theorem 4 *Suppose $\gamma < 1/2$. There exists an algorithm such that with probability at least $1-n^{-2}$,*

$$|\hat{p}(\mathcal{A}(G)) - p| \le \frac{C}{1/2-\gamma} \cdot \left(\frac{\sqrt{p(1-p)}}{\sqrt{n}} + \frac{\sqrt{\log n}}{n} \right),$$

for some constant C.

Note that for $\gamma > 1/60$, the error bound above matches that presented in Theorem 3 up to a factor of $1/(0.5-\gamma)$, and therefore extends the error rates of Theorem 3 to the regime $\gamma \in [1/60, 1/2)$ at the cost of computational efficiency.

1.2.3. Information-theoretic Lower Bounds

We provide a lower bound to establish near-optimality of our algorithms. While our upper bounds are against an omniscient adversary, the lower bounds hold for the weaker oblivious adversary.

Theorem 5 *For every $\gamma < 1/2$, $p \in [0,1]$, and $n \ge 0$ and a universal constant C', let*

$$\Delta = C' \cdot \left(\frac{\sqrt{p(1-p)}}{n} + \frac{\gamma\sqrt{p(1-p)}}{\sqrt{n}} + \frac{\gamma}{n} \right).$$

For any $p' \in [p+\Delta, p-\Delta]$ and $G \sim G(n,p)$ and $G' \sim G(n,p')$, there exists an oblivious adversary $\mathcal{A}$ such that no algorithm can distinguish between $\mathcal{A}(G)$ and $\mathcal{A}(G')$ with probability more than 0.65.

1.3. Techniques

Upper bound techniques. Broadly speaking, robust estimation is only possible when samples from the (uncorrupted) distribution enjoy some nice structure. Work in this area generally proceeds by imposing some regularity conditions on the uncorrupted data, which hold with high probability over

2. We have not tried to optimize the value of γ for computationally efficient algorithms, and could likely be made larger than $1/60$ through a more careful analysis.
3. Consider an empty graph $G(n,0)$. An adversary can corrupt half the graph into a clique, making it look like it came from $G(n,1)$. No algorithm can identify which half of the graph was the original.

samples from the distribution. The algorithm subsequently relies solely on these regularity conditions to make progress. For example, for mean estimation problems, it is common to assume that the mean and covariance of the uncorrupted samples are close to the true mean and covariance. However, the appropriate regularity conditions in our setting are far less obvious. We employ conditions which bound the empirical edge density and spectral norm for submatrices of the adjacency matrix, when appropriately centered around the true parameter p (Definition 10), which can be proven using tools from random matrix theory.

With our regularity conditions established, the algorithmic procedure proceeds in two stages: a coarse estimator, followed by a fine estimator.

Stage 1: A coarse estimate. Our regularity conditions are suggestive of the following intuition about how one might estimate the value of p. If one could locate a sufficiently large subgraph S of the uncorrupted nodes, such that their adjacency matrix centered around p has small spectral norm, then the empirical edge density of this subgraph would give a good estimate for the true parameter p. More precisely, we let A be the (corrupted) adjacency matrix of $\mathcal{A}(G)$, let $A_{S\times S}$ be the submatrix of A indexed by the set S, and p_S be the empirical edge density of the subgraph S. The goal is to obtain an S where $\|A_{S\times S} - p\|$ is small,[4] at which point we can output p_S.

There are two clear challenges with this approach. First off, we can not center the adjacency matrix around the unknown parameter p, since estimating that parameter is our goal. However, we demonstrate that it instead suffices to center around p_S (Theorem 13). The other issue is that it is not clear how to identify such a set S of uncorrupted nodes. One (inefficient) approach is to simply inspect all sufficiently large subgraphs. This will be accurate (quantified in Theorem 14), but not computationally tractable.

Instead, our main algorithmic contribution is an efficient algorithm which achieves this same goal. We give an iterative spectral approach, which starts with $S = [n]$. In Lemma 16 we show that if the spectral norm of $A_{S\times S} - p_S$ is large, then the top eigenvector assigns significant weight to the set of corrupted nodes. Normalizing this eigenvector and sampling from the corresponding probability distribution identifies a corrupted node with constant probability. We eliminate this node from S and repeat the process. Finally, using this approach, we obtain a subset $S^* \subset [n]$ of nodes such that p_{S^*} is a coarse estimate of p.

Stage 2: Pruning the coarse estimate. It turns out that the above coarse estimate gives a price of robustness which is roughly $O(1/\sqrt{n})$, rather than the $O(\gamma/\sqrt{n})$ we are trying to achieve. However, a simple pruning step allows us to complete the argument. Specifically, our coarse estimator gave us a set S^* such that the spectral norm of $A_{S^*\times S^*} - p_{S^*}$ is small and p_{S^*} is close to p. We employ this to show that most nodes must have degree close to p (Lemma 33). Thus, we remove $\Theta(\gamma n)$ nodes whose degree (restricted to the subgraph S^*) is furthest from p_{S^*}. Our final estimate is the empirical density of the resulting pruned subgraph.

Lower bound techniques. A strategy for proving lower bounds is the following: Suppose there exists an adversary that with γn corruptions can convert the distribution $G(n, p)$ and $G(n, p + \delta)$ into the same distribution of random graphs, then we cannot estimate p to accuracy better than $\delta/2$. This is akin to couplings between $G(n, p)$ and $G(n, p+\delta)$ by corrupting only a γn nodes. Designing these couplings over Erdős-Rényi graphs can be tricky due to the fact that degrees of nodes are not independent of each other.

4. For clarity: in the expression $A_{S\times S} - p$, p is subtracted entry-wise.

We instead consider directed Erdős-Rényi graphs, where an edge from a node i to j is present independently of all others. Then, the (outgoing) degrees of all the nodes are independent Binomial distributions. Using total variation bounds between Binomial distributions we can design couplings between directed ER graphs with different parameters, thus showing a lower bound on the error of robustly estimating directed ER graphs. Our final argument is a reduction showing that estimating the parameters of undirected of graphs is at least as hard as estimating the parameters of directed ER graphs. Combining these bounds we obtain the lower bounds.

1.4. Related Work

Due to the wealth of study in robust estimation and the page limits, we mention here only a fraction of the most relevant related work. For additional discussion, please see Section A.

Robust statistics is a classic and mature branch of statistics which focuses on precisely this type of setting since at least the 1960s (Tukey, 1960; Huber, 1964). However, since the classic literature typically did not take into account computational considerations, proposed estimators were generally intractable for settings of even moderate dimensionality (Bernholt, 2006). Recently, results by Diakonikolas, Kamath, Kane, Li, Moitra, and Stewart (2016, 2019a) and Lai, Rao, and Vempala (2016) overcame this barrier, producing the first algorithms which are both accurate and computationally efficient for robust estimation in multivariate settings. While they focused primarily on parameter estimation of Gaussian data, a flurry of subsequent works have provided efficient and accurate robust algorithms for a vast array of settings.

A common tool in several of these robust estimation results is to prune suspected outliers from the dataset so that a natural estimator over the remaining points has a small error. We also use this meta technique in this paper. We note that as in the previous works, the main challenge lies in designing efficient schemes to detect and remove corrupted data-points for the particular task at hand. In most prior works, the uncorrupted data-points are unaffected by corruptions. In our setting however, the edges from the good nodes are also affected by corruptions to the corrupted nodes. This presents a new challenge requiring new insights.

2. Notation and Preliminaries

Problem Formulation. Let $G \sim G(n, p)$. An adversary observes G and chooses a subset $B \subseteq [n]$ of nodes with $|B| \le \gamma n$. It can then change the status (i.e., presence or non-presence) of any edge with at least one node in B to get a graph $\mathcal{A}(G)$. Let $F = [n] \setminus B$. We call B the *corrupted* nodes, and F the *uncorrupted* nodes. Let $\tilde{A}$ and A be the $n \times n$ adjacency matrix of the original graph G and the modified graph $\mathcal{A}(G)$ respectively. Then $A_{F\times F} = \tilde{A}_{F\times F}$ and the remaining entries of A can be arbitrary. Given A, the goal is to estimate p, the parameter of the underlying random graph model. The algorithm does not know the set B, though we assume that it knows the value of γ.

Notation. The ℓ_2 norm of a vector $v = [v_1, \ldots, v_n] \in \mathbb{R}^n$ is $\|v\| := \sqrt{\sum_{i=1}^n v_i^2}$. Suppose M is an $m \times n$ real matrix. The spectral norm of M is

$$\|M\| := \max_{u\in\mathbb{R}^m, v\in\mathbb{R}^n : \|u\|=1, \|v\|=1} |u^T M v|. \tag{3}$$

It is easy to check that $\|M\| = \max_{v\in\mathbb{R}^n:\|v\|=1} \|Mv\|$. For a matrix M and real number $a \in \mathbb{R}$, let $M - a$ be the matrix obtained by subtracting a from each entry of M. For $S \subseteq [m]$, $S' \subseteq [n]$, let $M_{S\times S'}$ be the $m \times n$ matrix that agrees with M on $S \times S'$ and is zero elsewhere. Similarly for a vector $v \in \mathbb{R}^n$ and $S \subseteq [n]$, let vector v_S be the vector that agrees with with v on S and has zero entries elsewhere. Our proofs will use several standard properties of the matrix spectral norm, which we state in Appendix B for completeness.

3. Mean- and Median-based Algorithms

To demonstrate the need for our more sophisticated algorithms in Section 4, we first analyze canonical robust estimators for univariate settings – specifically, approaches based on trimming and order statistics (i.e., the median).

Recall the mean and median estimators for p in (1) and (2). The following simple lemma quantifies their guarantees in the setting absent corruptions.

Lemma 6 *Suppose $\gamma = 0$. There exists a constant $C > 0$ such that with probability at least 0.99, $|\hat{p}_{\mathrm{mean}}(G) - p| \le C \cdot \frac{\sqrt{p(1-p)}}{n}$, and $|\hat{p}_{\mathrm{med}}(G) - p| \le C \cdot \frac{1}{n}$.*

The analysis of these estimators is not difficult, but we include them for completeness in Section D.1. Analysis of the median estimator is slightly more involved due to correlations between nodes.

While both estimators are optimal up to constant factors (for constant p) without corruptions, their performance decays rapidly in the presence of an adversary, scaling at least linearly in the corruption fraction γ. In particular, consider an adversary that picks γn nodes at random and either adds all the edges with at least one endpoint in B or removes all of them. In Section D.2 we prove the following lower bound on the performance of the mean and median estimators for such an adversary. Observe that if even one node is corrupted (i.e., $\gamma \ge 1/n$), the error in Lemma 7 dominates the error without corruptions in Lemma 6.

Lemma 7 *There exists an adversary $\mathcal{A}$ such that for $\hat{p} \in \{\hat{p}_{\mathrm{mean}}(\mathcal{A}(G)), \hat{p}_{\mathrm{med}}(\mathcal{A}(G))\}$ with probability at least 0.5, we have $|\hat{p} - p| \ge \gamma/2$.*

A common strategy in robust statistics is to prune or trim the most extreme outliers. Accordingly, in our setting, one may prune the nodes with the most extreme degrees, described in Algorithm 1. This strategy bypasses the adversary which provides the lower bound in Lemma 7.

Algorithm 1 Prune-then-mean/median algorithm

Require: A graph $\mathcal{A}(G)$, corruption parameter γ, a constant $c > 0$

  Remove $c\gamma n$ nodes with largest and smallest degrees from $\mathcal{A}(G)$

  Apply the mean/median estimator from (1)/(2) to the resulting graph on $(1 - 2c\gamma) \cdot n$ nodes

---

However, this strategy can only go so far. Roughly speaking, pruning improves the mean's robust accuracy from $\Theta(\gamma)$ to $\Theta(\gamma^2)$, while pruning does not improve the median's robust accuracy. The upper and lower bounds are described in Theorems 8 and 9, and proved in Sections D.3 and D.4, respectively.

Theorem 8 *For $c \ge 1$ and $0 < \gamma \cdot c < 0.25$, the prune-then-mean and prune-then-median estimators described in Algorithm 1 prune $2c\gamma n$ nodes in total and with probability $1 - n^{-2}$ estimates p to an accuracy $\mathcal{O}\big(c\gamma^2 + \frac{\log n}{n}\big)$ and $\mathcal{O}\big(c\gamma + \sqrt{\frac{\log n}{n}}\big)$, respectively.*

Theorem 9 *Let $p = 0.5$, $\gamma > 100 \cdot \sqrt{\log n/n}$, and $c > 0$ be such that $c\gamma < 0.25$. There exists an adversary such that with probability at least 0.99, the prune-then-median estimate that deletes $c\gamma n$ satisfies $|\hat{p}(\mathcal{A}(G)) - p| \geq C'\gamma$, and the prune-then-mean estimate satisfies $|\hat{p}(\mathcal{A}(G)) - p| \geq C'\gamma^2$.*

To summarize: none of the standard univariate robust estimators we have explored are able to achieve error better than $\Omega(\gamma^2)$. To bypass this barrier, we turn to more intricate techniques in designing our main estimator in Section 4.

4. An Algorithm for Robust Estimation

Non-trivial robust estimation in Erdős-Rényi graphs is possible because even if the set of edges connected to a small set of nodes is changed arbitrarily, the subgraph between the remaining nodes retains a certain structure. In Section 4.1, we formalize this structure as deterministic regularity conditions and show that the subgraph corresponding to the set of uncorrupted nodes satisfy them with high probability. In the following subsections, we use only the fact that the subgraph of uncorrupted nodes satisfy these regularity conditions to derive our robust algorithms for estimating p.

In Section 4.2, we first derive a simple but novel inefficient spectral algorithm for coarse estimation of p. Our efficient algorithm consists of two parts: an efficient version of the spectral algorithm in Section 4.2 that, as its inefficient counterpart, provides a coarse estimate of p, followed by a trimming algorithm which achieves near-optimal error rates for estimating p. We describe and analyze the spectral and trimming components of the algorithm in Sections 4.3 and 4.4, respectively. Finally, in Appendix E.8, we put the pieces together to show that guarantees for these algorithms imply our upper bound in Theorem 3.

4.1. Regularity Conditions

In this section we state a set of three deterministic regularity conditions. We will then show that the set of uncorrupted nodes of a random Erdős-Rényi graph satisfy these regularity conditions with high probability. First, we define the following quantities κ and η, which we use in stating the regularity conditions and in the bounds of several lemmas and theorems. For $p \in [0,1]$ and $n > 0$, let

$$\eta(p, n) := c \cdot \max\left(\sqrt{\frac{p(1-p)}{n}}, \frac{\sqrt{\ln n}}{n}\right). \tag{4}$$

For $\alpha \in (0, 1]$, $p \in [0, 1]$ and $n > 0$, let

$$\kappa(\alpha, p, n) := c_1 \cdot \max\left(\alpha\sqrt{\frac{p}{n}\ln\frac{e}{\alpha}}, \frac{\alpha}{n}\ln\frac{e}{\alpha}, \frac{\sqrt{p\ln n}}{n}\right). \tag{5}$$

In the above definitions c and c_1 are some constants that we determine in Theorem 12.

We employ the following regularity conditions.

Definition 10 *Given $\alpha_1 \in [0, 1/2)$, $\alpha_2 \in [0, 1/2)$, and an $[n] \times [n]$ adjacency matrix A, a set of nodes $F \subseteq [n]$ of the graph corresponding to A satisfy (α_1, α_2, p)-*regularity *if*

1. *$|F^c| \leq \alpha_1 n$.*
2. *For all $F' \subseteq F$,*

$$\|(A - p)_{F' \times F'}\| \leq n \cdot \eta(p, n).$$

3. *For all $F', F'' \subseteq F$ such that $|F'|, |F''| \in [0, \alpha_2 n] \cup [n - \alpha_2 n, n]$, then*

$$\Big|\sum_{i \in F'} \sum_{j \in F''} (A_{i,j} - p)\Big| \leq n^2 \cdot \kappa(\alpha_2, p, n).$$

Item 2 implies that upon subtracting p from each entry of the adjacency matrix A, the spectral norm of the matrix corresponding to all subgraphs of the subgraph $F \times F$ is bounded. Item 3 implies that upon subtracting p from each entry of the adjacency matrix A, the sum of the entries over any of its submatrices $F' \times F'' \subseteq F \times F$ has a small absolute value, as long as each of F' and F'' either leave out or include at most $\alpha_2 n$ nodes. We will informally refer to nodes in the set $F \subseteq [n]$ that satisfy (α_1, α_2, p)-regularity as *good nodes*.

For a subset $S \subseteq [n]$ and adjacency matrix A, we will use $p_S := \frac{\sum_{i,j \in S} A_{i,j}}{|S|^2}$ to denote (approximately) the empirical fraction of edges present in the subgraph induced by a set S. Note that this differs slightly from expression one might anticipate, $\binom{|S|}{2}^{-1} \left(\sum_{i<j:\, i,j \in S} A_{i,j}\right)$. For convenience, our sum double-counts each edge and also includes the $A_{i,i}$ terms (which are always 0 due to the lack of self-loops). The double counting is accounted for since the denominator is scaled by a factor of 2. The inclusion of the diagonal 0's is *not* accounted for, thus leading to p_S being a slight underestimate of the empirical edge parameter for this subgraph, but not big enough to make a significant difference.

The following lemma lists some simple but useful consequences of the regularity conditions that we use in later proofs. We prove it in Appendix E.1.

Lemma 11 *Suppose $0 \leq \alpha_1, \alpha_2 < 1/2$ and adjacency matrix A has a node subset $F \subseteq [n]$ that satisfies (α_1, α_2, p)-regularity, then*

1. *For all $F' \subseteq F$,*

$$\|(A - p_{F'})_{F' \times F'}\| \leq 2n \cdot \eta(p, n). \tag{6}$$

2. *For all $F' \subseteq F$ of size $\geq (1 - \alpha_2)n$,*

$$|p_{F'} - p| \leq 4\kappa(\alpha_2, p, n). \tag{7}$$

Equation (6) implies that if the adjacency matrix of any subset of good nodes is centered around its empirical fraction of the edges, then its spectral norm is bounded. Equation (7) implies that for any subset of good nodes that excludes at most $\alpha_2 n$ nodes, the empirical fraction of edges in the subgraph induced by it estimates p accurately.

The next theorem shows that the set of uncorrupted nodes of a random Erdős-Rényi graph satisfy these regularity conditions with high probability. The proof of the Theorem is in Appendix E.2.

Theorem 12 *For any $\gamma \in [0, 1/2)$, $n > 0$ and $p > 0$, let A be a γ-corrupted adjacency matrix of a sample from $G(n, p)$. There exist universal constants c and c_1 in Equations (4) and (5), respectively, such that with probability at least $1 - 4n^{-2}$ the set of uncorrupted nodes F satisfy (α_1, α_2, p)-regularity for all $\alpha_1 \in [\gamma, 1/2]$ and $\alpha_2 \in [0, 1/2]$.*

4.2. An Inefficient Coarse Estimator

In this section we propose a simple inefficient algorithm to recover a coarse estimate of p, which has an optimal dependence on all parameters other than α_1.

The following theorem serves as the foundation of our coarse estimator. It shows that if, for any subset $S \subseteq [n]$ of size $\geq n/2$ nodes, the spectral norm of its submatrix centered with respect to p_S is small, then p_S is a reasonable estimate of p.

Theorem 13 *Suppose $0 \leq \alpha_1, \alpha_2 < 1/2$, and let A be an adjacency matrix containing a (α_1, α_2, p)-regular subgraph. Then for all $S \subseteq [n]$ such that $|S| \geq n/2$, we have*

$$|p_S - p| \leq \frac{\|(A - p_S)_{S\times S}\| + n \cdot \eta(p, n)}{(1/2 - \alpha_1)n}.$$

Proof Let F be the (α_1, α_2, p)-regular subgraph of A. From the triangle inequality,

$$\|(A - p_S)_{(S\cap F)\times(S\cap F)}\| \geq |p - p_S| \cdot |S \cap F| - \|(A - p)_{(S\cap F)\times(S\cap F)}\|.$$

Then by Lemma 23,

$$|p-p_S|\cdot|S\cap F| \leq \|(A-p_S)_{(S\cap F)\times(S\cap F)}\|+\|(A-p)_{(S\cap F)\times(S\cap F)}\| \leq \|(A-p_S)_{S\times S}\|+\|(A-p)_{F\times F}\|.$$

Finally, noting that $|S \cap F| \geq |S| - |F^c| \geq |S| - \alpha_1 n \geq n/2 - \alpha_1 n$ proves the theorem. ■

With this in hand, it suffices to locate a subset of nodes S such that $\|(A - p_S)_{S\times S}\|$ is small. We provide the accuracy guarantee of our inefficient algorithm in the following theorem.

Theorem 14 *Suppose $0 \leq \alpha_1, \alpha_2 < 1/2$, and let A be an adjacency matrix containing a (α_1, α_2, p)-regular subgraph. Let*

$$\hat{S} = \arg\min_{S\subseteq[n]:|S|\geq n/2} \|(A - p_S)_{S\times S}\|.$$

Then $\|(A - p_{\hat{S}})_{\hat{S}\times\hat{S}}\| \leq 2n \cdot \eta(p, n)$ and $|p_{\hat{S}} - p| \leq \frac{3}{(1/2-\alpha_1)} \cdot \eta(p, n)$.

Proof Let F be the (α_1, α_2, p)-regular subgraph of A. From the definition of $\hat{S}$,

$$\|(A - p_{\hat{S}})_{\hat{S}\times\hat{S}}\| \leq \|(A - p_F)_{F\times F}\| \leq 2n \cdot \eta(p, n),$$

where the last inequality uses Equation (6). The proof follows from Theorem 13. ■

Theorem 14 implies the following simple algorithm to estimate p: compute $\hat{S}$ by iterating over all subsets of $[n]$, and then output $p_{\hat{S}}$. Combining with Theorem 12, this proves Theorem 4. The clear downside of this approach is that it is not computationally efficient, with a running time that depends exponentially on n. Also, as we will later establish, while this algorithm gives near-optimal rates for all constant γ bounded away from $1/2$ by a constant, it may be sub-optimal for smaller γ. In the following sections, we address both of these issues: we provide a computationally efficient algorithm which provides near-optimal rates for $\gamma < 1/60$.

4.3. An Efficient Coarse Spectral Algorithm

In this section, we propose an efficient spectral method (Algorithm 2) which finds a subset $S^* \subseteq [n]$ such that both the set $(S^*)^c$ and the spectral norm $\|(A - p_{S^*})_{S^* \times S^*}\|$ are small. Note that the latter guarantee is comparable to the inefficient algorithm from Section 4.2. Then Theorem 13 implies that p_{S^*} is an accurate estimate of p. We note that this is still a *coarse* estimate of p, which has a sub-optimal dependence on α_1.[5] In the following section, we will post-process the set S^* returned by Algorithm 2 to provide our near-optimal bounds.

Theorem 15 *Suppose* $\alpha_1 \in [\frac{1}{n}, \frac{1}{60}]$, $\alpha_2 \in [0, 1/2]$ *and let* A *be an adjacency matrix containing an* (α_1, α_2, p)*-regular subgraph. With probability at least* $1 - 1/n^2$,[6] *Algorithm 2 returns a subset* S^* *with* $|S^*| \geq (1-9\alpha_1)n$ *such that* $\|(A-p_{S^*})_{S^* \times S^*}\| \leq 20n \cdot \eta(p, n)$. *Furthermore, these conditions on* S^* *imply* $|p_{S^*} - p| \leq 45 \cdot \eta(p, n)$.

Algorithm 2 Spectral algorithm for estimating p

Require: number of nodes n, parameter $\alpha_1 \in [1/n, 1/60]$, adjacency matrix A
  $S \leftarrow [n]$, Candidates $\leftarrow \{\}$
  Candidates $\leftarrow$ Candidates $\cup \{S\}$
  for $t = 1$ to $9\alpha_1 n$ **do**
    Compute a top normalized eigenvector v of the matrix $(A - p_S)_{S \times S}$
    Draw i_t from the distribution where $i \in S$ is selected with probability v_i^2
    $S \leftarrow S \setminus \{i_t\}$
    Candidates $\leftarrow$ Candidates $\cup \{S\}$
  end for
  $S^* \leftarrow \arg\min_{S \in \text{Candidates}} \|(A - p_S)_{S \times S}\|$
  return S^*

In the remainder of this section we will prove that Algorithm 2 indeed outputs a subset S^* with the guarantee in Theorem 15. Let F (unknown) be the (α_1, α_2, p)-regular subgraph of A. The key technical argument is that if the spectral norm of $(A - p_S)_{S \times S}$ is large, the normalized top eigenvector v of $(A - p_S)_{S \times S}$ places constant weight on the subset $S \cap F^c$. Thus, if at a given iteration Algorithm 2 possesses an unsatisfactory set S, it will remove a node from $S \cap F^c$ with a constant probability. We formalize this argument in the following key Lemma 16. The proof of the lemma appears in Appendix E.3.

Lemma 16 *Suppose* $\alpha_1 \in [\frac{1}{n}, \frac{1}{60}]$, $\alpha_2 \in [0, 1/2]$ *and let* A *be an adjacency matrix containing an* (α_1, α_2, p)*-regular subgraph* F*. Let* $S \subseteq [n]$ *be of size* $|S| \geq (1 - 9\alpha_1)n$*, and* v *be the normalized top eigenvector of* $(A - p_S)_{S \times S}$*. If* $\|(A - p_S)_{S \times S}\| \geq 20n \cdot \eta(p, n)$ *then* $\|v_{S \cap F^c}\|^2 \geq 0.15$.

We conclude this section with the proof of Theorem 15.

Proof [Proof of Theorem 15] It suffices to show that at least one of the sets S encountered by Algorithm 2 satisfies the condition $\|(A - p_S)_{S \times S}\| \leq 20n \cdot \eta(p, n)$. From Lemma 16 it follows

5. The guarantees are comparable to Theorem 14, up to constant factors.

6. The probability of success of Algorithm 2 is $\Pr[\text{Bin}(\lfloor 9\alpha_1 n \rfloor, 0.15) \geq \lfloor \alpha_1 n \rfloor] \geq 1 - \exp(-\Omega(\alpha_1 n))$. Note that for all values of $\alpha_1 n$ the success probability is $> 1/2$. When $\alpha_1 n = \Omega(\log n)$ then it gives the probability of success at least $1 - 1/n^2$. When $\alpha_1 n = \mathcal{O}(\log n)$, to get the probability of success $\geq 1 - 1/n^2$ one can run Algorithm 2 $\mathcal{O}(\log n)$ times and choose an S^* for which $\|(A - p_{S^*})_{S^* \times S^*}\|$ is the minimum among all runs.

that until the algorithm finds such a subset S, in each deletion step the probability of deleting a node from F^c is at least 0.15. Since there are $9\alpha_1 n$ steps, a standard Chernoff-style argument implies that either a subset S (including nodes from both F^c and F) satisfying the conditions of the theorem will be created, or with probability at least $\Pr[\text{Bin}(\lfloor 9\alpha_1 n\rfloor, 0.15) \geq \lfloor \alpha_1 n\rfloor] \geq 1-\exp(-\Omega(\alpha_1 n))$, all nodes from F^c will be deleted and thus $S \subseteq F$. In the latter case we apply Equation (6), which implies that $\|(A-p_S)_{S\times S}\| \leq 20n\cdot\eta(p,n)$ and the theorem. ■

Remark 17 *Algorithm 2 runs for $9\alpha_1 n$ rounds, and in each round the algorithm finds the top eigenvector of an $n\times n$ matrix. This may be expensive to compute when the spectral gap is small. However, Lemma 31 shows it suffices to find any unit vector $v \in \mathbb{R}^n$ such that $|v^\intercal(A-p_S)_{S\times S}v| \geq 0.99||(A-p_S)_{S\times S}||$. Note that such a unit vector can be found in $\tilde{O}(n^2)$ time (Musco and Musco, 2015). Therefore, one can implement Algorithm 2 to run in $\tilde{O}(\alpha_1 n^3)$ time.*

4.4. A Fine Trimming Algorithm

In this section, we provide a trimming method (Algorithm 3), which refines the output of Algorithm 2, improving its guarantee (quantified in Theorem 15) by up to a factor of α_1.

The algorithm (Algorithm 3) is easy to describe. For a subset $S^* \subseteq [n]$ and a node $i \in S^*$, we define $p_{S^*}^{(i)} := \frac{\sum_{j\in S^*} A_{i,j}}{|S^*|}$ to be the normalized degree of node i in the subgraph induced by S^*. We remove the $3\alpha_1 n$ nodes for which this normalized degree deviate furthest from the average parameter p_{S^*}. Its guarantees are quantified in Theorem 18, whose proof appears in Appendix E.7.

Theorem 18 *Let $\alpha_1 \in [\frac{1}{n}, \frac{1}{60}]$, and A be an adjacency matrix containing an $(\alpha_1, 13\alpha_1, p)$-regular subgraph. Suppose we have some S^* such that $|S^*| \geq (1-9\alpha_1)n$ and $\|(A-p_{S^*})_{S^*\times S^*}\| \leq 20n\cdot\eta(p,n)$, Algorithm 3 outputs p_{S^f} such that for some universal constants $c_2, c_3 > 0$,*

$$\left|p_{S^f} - p\right| \leq c_2\alpha_1\eta(p,n) + c_3\kappa(13\alpha_1, p, n).$$

Algorithm 3 Trimming Algorithm

Require: number of nodes n, parameter $\alpha_1 \in [1/n, 1/60]$, adjacency matrix A, subset $S^* \subseteq [n]$

Define the score for each node $i \in S^*$ to be $|p_{S^*} - p_{S^*}^{(i)}|$

Remove the $3\alpha_1 n$ nodes in S^* with the highest scores to obtain S^f

return p_{S^f}

At this point, we have all the pieces to prove our main upper bound (Theorem 3). The argument first reasons that a random graph will satisfy certain regularity conditions with high probability. With these guarantees, we feed it into our coarse spectral algorithm (Algorithm 2), followed by our fine trimming algorithm (Algorithm 3). Some (mundane) case analysis is required to achieve the optimal bounds in certain parameter regimes; the full argument is rigorously described in Appendix E.8.

5. Lower Bounds

In this section, we prove our main lower bound for robust parameter estimation in Erdős-Rényi random graphs establishing that our algorithms are tight up to logarithmic factors.

First we consider the problem of parameter estimation for directed version of Erdős-Rényi random graphs. Such graphs have independent (outgoing) degrees across the nodes. We then show a reduction showing that the directed version of the problem is at least as hard as the standard version. We start by describing the directed Erdős-Rényi graphs.

Definition 19 (Directed Erdős-Rényi graphs) *The directed Erdős-Rényi random graph model on n nodes with parameter p, denoted as $DG(n,p)$, is the distribution over directed graphs on n nodes where each edge is present with probability p, independently of the other edges.*

We show the following reduction from the directed problem to standard. We provide the proof in Appendix G.

Lemma 20 *If there exists an algorithm that estimates p in $G(n,p)$ to within $\pm\Delta$ with probability at most $1-\delta$ under γ-corruptions, then there exists an algorithm for estimating p in $DG(n,p)$ to within $\pm\Delta$ with probability at most $1-\delta$ under γ-corruptions.*

Then to prove the lower bound in Theorem 5, we prove its analogue for directed Erdős-Rényi graphs.

Theorem 21 *Let $p \le 0.5$. Then there exists a γ-oblivious adversary such that no algorithm can distinguish between $DG(n,p)$ and $DG\left(n, p + 0.1\max\left(\gamma\sqrt{p/n}, \gamma/n, \sqrt{p}/n\right)\right)$ with probability at least 0.65.*

By symmetry, a similar statement holds for $p > 0.5$, with p replaced by $1-p$. Combining these two statements and Lemma 20 gives the lower bound in Theorem 5.

We prove Theorem 21 formally in Appendix G, and conclude the section with a proof sketch. We consider a weaker γ-oblivious adversary for $DG(n,p)$ that does the following: (a) randomly choose a subset B of γn nodes, (b) for each node $i \in B$, remove all the outgoing edges from i, and draw a number d_i independently from a different distribution over $\{0, 1, \ldots, n\}$, and (c) select d_i nodes from $[n] \setminus \{i\}$ at random and add an edge from i to them. Note that both for the uncorrupted nodes and for nodes corrupted by such an adversary the out-degrees of nodes completely determine its distribution and is therefore a sufficient statistic. For a random directed graph $DG \sim DG(n,p)$, the out-degree of a node is distributed $Bin(n-1,p)$. We can think of observed degrees of n nodes of an uncorrupted directed Erdős-Rényi random graph $DG \sim DG(n,p)$ as independent samples from binomial distribution $Bin(n-1,p)$. Let $\Delta_p = 0.1\max\left(\gamma\sqrt{p/n}, \gamma/n\right)$. We show that for $p \le 0.5$, the TV distance between $Bin(n-1,p)$ and $Bin(n-1,p+\Delta)$ is less than 0.15γ. Then we show that an adversary that chooses a random set of size $Bin(n, 0.15\gamma)$, can choose the distribution of their out-degree in a way that the overall distribution of out-degrees is same for the both graphs $DG \sim DG(n,p)$ and $DG \sim DG(n, p+\Delta_p)$ after corruption. Finally, we show that even without corruption (when $\gamma = 0$), no algorithm can reliably distinguish between $DG(n,p)$ and $DG(n, p + 0.1\sqrt{p}/n)$.

Acknowledgments

The authors would like to thank the anonymous reviewers for their helpful comments. Jayadev acharya is supported by NSF-CCF1846300 (CAREER), NSF-CCF-1815893, and a Google Faculty

Research Award. Gautam kamath is supported by an NSERC Discovery Grant, an unrestricted gift from Google, and a University of Waterloo startup grant. Part of this work was done while the Huanyu Zhang was a graduate student at Cornell University and was supported by NSF-CCF-1815893.

References

José A. Adell and Pedro Jodrá. Exact Kolmogorov and total variation distances between some familiar discrete distributions. *Journal of Inequalities and Applications*, 2006(1):1–8, 2006.

Noga Alon, Michael Krivelevich, and Benny Sudakov. Finding a large hidden clique in a random graph. *Random Structures and Algorithms*, 13(3-4):457–466, 1998.

Pranjal Awasthi, Maria Florina Balcan, and Philip M. Long. The power of localization for efficiently learning linear separators with noise. In *Proceedings of the 46th Annual ACM Symposium on the Theory of Computing*, STOC '14, pages 449–458, New York, NY, USA, 2014. ACM.

Sivaraman Balakrishnan, Simon S. Du, Jerry Li, and Aarti Singh. Computationally efficient robust sparse estimation in high dimensions. In *Proceedings of the 30th Annual Conference on Learning Theory*, COLT '17, pages 169–212, 2017.

Afonso S. Bandeira and Ramon Van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *The Annals of Probability*, 44(4):2479–2506, 2016.

Jess Banks, Sidhanth Mohanty, and Prasad Raghavendra. Local statistics, semidefinite programming, and community detection. In *Proceedings of the 32nd Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '21, pages 1298–1316, Philadelphia, PA, USA, 2021. SIAM.

Boaz Barak, Samuel Hopkins, Jonathan Kelner, Pravesh K Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM Journal on Computing*, 48(2):687–735, 2019.

Thorsten Bernholt. Robust estimators are hard to compute. Technical report, Technische Universität Dortmund, 2006.

Christian Borgs, Jennifer Chayes, and Adam Smith. Private graphon estimation for sparse graphs. In *Advances in Neural Information Processing Systems 28*, NIPS '15, pages 1369–1377. Curran Associates, Inc., 2015.

Christian Borgs, Jennifer Chayes, Adam Smith, and Ilias Zadik. Revealing network structure, confidentially: Improved rates for node-private graphon estimation. In *Proceedings of the 59th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '18, pages 533–543, Washington, DC, USA, 2018. IEEE Computer Society.

T. Tony Cai and Xiaodong Li. Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. *The Annals of Statistics*, 43(3):1027–1059, 2015.

Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM*, 58(3):1–37, 2011.

Venkat Chandrasekaran, Sujay Sanghavi, Pablo A Parrilo, and Alan S Willsky. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21(2):572–596, 2011.

Moses Charikar, Jacob Steinhardt, and Gregory Valiant. Learning from untrusted data. In *Proceedings of the 49th Annual ACM Symposium on the Theory of Computing*, STOC '17, pages 47–60, New York, NY, USA, 2017. ACM.

Sitan Chen, Jerry Li, and Ankur Moitra. Efficiently learning structured distributions from untrusted batches. In *Proceedings of the 52nd Annual ACM Symposium on the Theory of Computing*, STOC '20, pages 960–973, New York, NY, USA, 2020a. ACM.

Sitan Chen, Jerry Li, and Ankur Moitra. Learning structured distributions from untrusted batches: Faster and simpler. In *Advances in Neural Information Processing Systems 33*, NeurIPS '20. Curran Associates, Inc., 2020b.

Yu Cheng, Ilias Diakonikolas, Daniel M. Kane, and Alistair Stewart. Robust learning of fixed-structure Bayesian networks. In *Advances in Neural Information Processing Systems 31*, NeurIPS '18, pages 10304–10316. Curran Associates, Inc., 2018.

Yu Cheng, Ilias Diakonikolas, and Rong Ge. High-dimensional robust mean estimation in nearly-linear time. In *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '19, pages 2755–2771, Philadelphia, PA, USA, 2019a. SIAM.

Yu Cheng, Ilias Diakonikolas, Rong Ge, and David Woodruff. Faster algorithms for high-dimensional robust covariance estimation. In *Proceedings of the 32nd Annual Conference on Learning Theory*, COLT '19, pages 727–757, 2019b.

Yu Cheng, Ilias Diakonikolas, Rong Ge, and Mahdi Soltanolkotabi. High-dimensional robust mean estimation via gradient descent. In *Proceedings of the 37th International Conference on Machine Learning*, ICML '20, pages 1768–1778. JMLR, Inc., 2020.

Yash Deshpande and Andrea Montanari. Improved sum-of-squares lower bounds for hidden clique and hidden submatrix problems. In *Proceedings of the 28th Annual Conference on Learning Theory*, COLT '15, pages 523–562, 2015.

Ilias Diakonikolas and Daniel M. Kane. Recent advances in algorithmic high-dimensional robust statistics. *arXiv preprint arXiv:1911.05911*, 2019.

Ilias Diakonikolas, Gautam Kamath, Daniel M. Kane, Jerry Li, Ankur Moitra, and Alistair Stewart. Robust estimators in high dimensions without the computational intractability. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '16, pages 655–664, Washington, DC, USA, 2016. IEEE Computer Society.

Ilias Diakonikolas, Gautam Kamath, Daniel M. Kane, Jerry Li, Ankur Moitra, and Alistair Stewart. Being robust (in high dimensions) can be practical. In *Proceedings of the 34th International Conference on Machine Learning*, ICML '17, pages 999–1008. JMLR, Inc., 2017.

Ilias Diakonikolas, Gautam Kamath, Daniel M. Kane, Jerry Li, Ankur Moitra, and Alistair Stewart. Robustly learning a Gaussian: Getting optimal error, efficiently. In *Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '18, Philadelphia, PA, USA, 2018a. SIAM.

Ilias Diakonikolas, Daniel M. Kane, and Alistair Stewart. List-decodable robust mean estimation and learning mixtures of spherical Gaussians. In *Proceedings of the 50th Annual ACM Symposium on the Theory of Computing*, STOC '18, pages 1047–1060, New York, NY, USA, 2018b. ACM.

Ilias Diakonikolas, Gautam Kamath, Daniel M. Kane, Jerry Li, Ankur Moitra, and Alistair Stewart. Robust estimators in high-dimensions without the computational intractability. *SIAM Journal on Computing*, 48(2):742–864, 2019a.

Ilias Diakonikolas, Gautam Kamath, Daniel M. Kane, Jerry Li, Jacob Steinhardt, and Alistair Stewart. Sever: A robust meta-algorithm for stochastic optimization. In *Proceedings of the 36th International Conference on Machine Learning*, ICML '19, pages 1596–1606. JMLR, Inc., 2019b.

Yihe Dong, Samuel Hopkins, and Jerry Li. Quantum entropy scoring for fast robust mean estimation and improved outlier detection. In *Advances in Neural Information Processing Systems 32*, NeurIPS '19, pages 6065–6075. Curran Associates, Inc., 2019.

Paul Erdős and Alfréd Rényi. On random graphs i. *Publicationes Mathematicate*, 6:290–297, 1959.

Uriel Feige and Robert Krauthgamer. The probable value of the lovász–schrijver relaxations for maximum independent set. *SIAM Journal on Computing*, 32(2):345–370, 2003.

Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh S Vempala, and Ying Xiao. Statistical algorithms and a lower bound for detecting planted cliques. *Journal of the ACM*, 64(2):1–37, 2017.

Edgar N. Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.

Samuel B. Hopkins and Jerry Li. Mixture models, robustness, and sum of squares proofs. In *Proceedings of the 50th Annual ACM Symposium on the Theory of Computing*, STOC '18, pages 1021–1034, New York, NY, USA, 2018. ACM.

Samuel B Hopkins, Pravesh Kothari, Aaron Henry Potechin, Prasad Raghavendra, and Tselil Schramm. On the integrality gap of degree-4 sum of squares for planted clique. *ACM Transactions on Algorithms*, 14(3):1–31, 2018.

Daniel Hsu, Sham M Kakade, and Tong Zhang. Robust matrix decomposition with sparse corruptions. *IEEE Transactions on Information Theory*, 57(11):7221–7234, 2011.

Peter J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964.

Ayush Jain and Alon Orlitsky. Optimal robust learning of discrete distributions from batches. In *Proceedings of the 37th International Conference on Machine Learning*, ICML '20, pages 4651–4660. JMLR, Inc., 2020a.

Ayush Jain and Alon Orlitsky. A general method for robust learning from batches. In *Advances in Neural Information Processing Systems 33*, NeurIPS '20, pages 21775–21785. Curran Associates, Inc., 2020b.

Ayush Jain and Alon Orlitsky. Robust density estimation from batches: The best things in life are (nearly) free. In *International Conference on Machine Learning*, pages 4698–4708. PMLR, 2021.

Ayush Jain, Alon Orlitsky, and Vaishakh Ravindrakumar. Robust estimation algorithms don't need to know the corruption level. *arXiv preprint arXiv:2202.05453*, 2022.

Mark Jerrum. Large cliques elude the Metropolis process. *Random Structures and Algorithms*, 3 (4):347–359, 1992.

Rob Kaas and Jan M Buhrman. Mean, median and mode in binomial distributions. *Statistica Neerlandica*, 34(1):13–18, 1980.

Richard Karp. Probabilistic analysis of some combinatorial search problems. traub, jf (ed.): Algorithms and complexity: New directions and recent results, 1976.

Adam Klivans, Pravesh K. Kothari, and Raghu Meka. Efficient algorithms for outlier-robust regression. In *Proceedings of the 31st Annual Conference on Learning Theory*, COLT '18, pages 1420–1430, 2018.

Adam R Klivans, Philip M Long, and Rocco A Servedio. Learning halfspaces with malicious noise. *Journal of Machine Learning Research*, 10(12):2715–2740, 2009.

Pravesh Kothari, Jacob Steinhardt, and David Steurer. Robust moment estimation and improved clustering via sum of squares. In *Proceedings of the 50th Annual ACM Symposium on the Theory of Computing*, STOC '18, pages 1035–1046, New York, NY, USA, 2018. ACM.

Luděk Kučera. Expected complexity of graph partitioning problems. *Discrete Applied Mathematics*, 57(2-3):193–212, 1995.

Kevin A. Lai, Anup B. Rao, and Santosh Vempala. Agnostic estimation of mean and covariance. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '16, pages 665–674, Washington, DC, USA, 2016. IEEE Computer Society.

Erik M. Lindgren, Vatsal Shah, Yanyao Shen, Alexandros G. Dimakis, and Adam Klivans. On robust learning of Ising models. In *NeurIPS Workshop on Relational Representation Learning*, 2018.

Liu Liu, Yanyao Shen, Tianyang Li, and Constantine Caramanis. High dimensional robust sparse regression. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics*, AISTATS '20, pages 411–421. JMLR, Inc., 2020.

Konstantin Makarychev, Yury Makarychev, and Aravindan Vijayaraghavan. Learning communities in the presence of errors. In *Proceedings of the 29th Annual Conference on Learning Theory*, COLT '16, pages 1258–1291, 2016.

Raghu Meka, Aaron Potechin, and Avi Wigderson. Sum-of-squares lower bounds for planted clique. In *Proceedings of the 47th Annual ACM Symposium on the Theory of Computing*, STOC '15, pages 87–96, New York, NY, USA, 2015. ACM.

Ankur Moitra, William Perry, and Alexander S Wein. How robust are reconstruction thresholds for community detection? In *Proceedings of the 48th Annual ACM Symposium on the Theory of Computing*, STOC '16, pages 828–841, New York, NY, USA, 2016. ACM.

Cameron Musco and Christopher Musco. Randomized block Krylov methods for stronger and faster approximate singular value decomposition. In *Advances in Neural Information Processing Systems 28*, NeurIPS '15, pages 1396–1404. Curran Associates, Inc., 2015.

Mark E. J. Newman, Duncan J. Watts, and Steven H. Strogatz. Random graph models of social networks. *Proceedings of the National Academy of Sciences*, 99(suppl 1):2566–2572, 2002.

Ankit Pensia, Varun Jog, and Po-Ling Loh. Robust regression with covariate filtering: Heavy tails and adversarial contamination. *arXiv preprint arXiv:2009.12976*, 2021.

Adarsh Prasad, Vishwak Srinivasan, Sivaraman Balakrishnan, and Pradeep Ravikumar. On learning Ising models under Huber's contamination model. In *Advances in Neural Information Processing Systems 33*, NeurIPS '20. Curran Associates, Inc., 2020a.

Adarsh Prasad, Arun Sai Suggala, Sivaraman Balakrishnan, and Pradeep Ravikumar. Robust estimation via robust gradient estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(3):601–627, 2020b.

Mingda Qiao and Gregory Valiant. Learning discrete distributions from untrusted batches. In *Proceedings of the 9th Conference on Innovations in Theoretical Computer Science*, ITCS '18, pages 47:1–47:20, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

Bero Roos. Binomial approximation to the poisson binomial distribution: The krawtchouk expansion. *Theory of Probability & Its Applications*, 45(2):258–272, 2001.

Adam Sealfon and Jonathan Ullman. Efficiently estimating Erdos-Renyi graphs with node differential privacy. In *Advances in Neural Information Processing Systems 32*, NeurIPS '19, pages 3765–3775. Curran Associates, Inc., 2019.

Jacob Steinhardt, Moses Charikar, and Gregory Valiant. Resilience: A criterion for learning in the presence of arbitrary outliers. In *Proceedings of the 9th Conference on Innovations in Theoretical Computer Science*, ITCS '18, pages 45:1–45:21, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

John W. Tukey. A survey of sampling from contaminated distributions. *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, pages 448–485, 1960.

Banghua Zhu, Jiantao Jiao, and Jacob Steinhardt. Generalized resilience and robust statistics. *arXiv preprint arXiv:1909.08755*, 2019.

Appendix A. Additional Related Work

Beyond the aforementioned results on efficient robust estimation (Diakonikolas et al., 2016; Lai et al., 2016), several works have focused on similar estimation tasks in a variety of related settings, including under weaker moment assumptions (Diakonikolas et al., 2017), with a larger fraction of corrupted data (Charikar et al., 2017), under sparsity constraints (Balakrishnan et al., 2017; Liu et al., 2020), for regression or other supervised learning tasks (Klivans et al., 2018; Diakonikolas et al., 2019b; Prasad et al., 2020b; Pensia et al., 2021), under more general robustness conditions (Steinhardt et al., 2018), with alternate perturbation models (Zhu et al., 2019), for mixture models (Hopkins and Li, 2018; Kothari et al., 2018; Diakonikolas et al., 2018b), approaching information-theoretic barriers to accuracy (Diakonikolas et al., 2018a), fast algorithms for robust estimation (Cheng et al., 2019a,b; Dong et al., 2019), and with gradient descent algorithms (Cheng et al., 2020). See (Diakonikolas and Kane, 2019) for a survey.

Our algorithm relies on a spectral outlier-removal technique common to several works in robust estimation. Prior to this line of work, similar approaches were employed for robust supervised learning tasks, namely learning halfspaces with malicious noise (Klivans et al., 2009; Awasthi et al., 2014).

There has been significant work on robust community detection in the presence of adversaries (Moitra et al., 2016; Makarychev et al., 2016; Steinhardt et al., 2018; Banks et al., 2021). Most of this focuses on monotone adversaries (which make only "helpful" changes to the graph) or edge corruptions. It is not clear how to define monotone adversaries for the Erdős-Rényi setting, and for our estimation problem under edge corruptions, the empirical estimator is trivially optimal in the worst case. The work of Cai and Li (2015) also considers a node corruption model similar to ours. However, all of the aforementioned work studies community detection in stochastic block models, which is different from our goal of parameter estimation.

Our corruption model may seem reminiscent of the classic planted clique problem (Karp, 1976; Jerrum, 1992; Kučera, 1995), in which an algorithm must distinguish between a) $G(n, 1/2)$ and b) $G(n, 1/2)$ with the addition of a planted clique of size γn. Our adversary is given much more power (i.e., they can make arbitrary changes to the neighbourhoods of their selected nodes), though the two goals are incomparable. The planted clique problem is known to be information-theoretically solvable for any $\gamma > \frac{2 \log n}{n}$. However, polynomial-time algorithms are only known for $\gamma > 1/\sqrt{n}$ (Alon et al., 1998), and there is strong evidence that efficient algorithms do not exist for smaller values of γ (Feige and Krauthgamer, 2003; Feldman et al., 2017; Meka et al., 2015; Deshpande and Montanari, 2015; Hopkins et al., 2018; Barak et al., 2019). We have not run into issues in our setting related to this intractability, though deeper connections between our model and the planted clique problem would be interesting. Note that our task of parameter estimation is not interesting for the cases of the planted clique problem when $\gamma \le 1/\sqrt{n}$. Simply using the empirical estimator on the two instances would give error $\approx 1/n$ and $\approx 1/n + \gamma^2 = O(1/n)$, which are identical up to constant factors.

Some prior works have studied robust estimation for graphical models, including Ising models (Lindgren et al., 2018; Prasad et al., 2020a) and Bayesian networks (Cheng et al., 2018). Despite the common nomenclature, these works are rather different from our work on random graph models. Graphical models are distributions over vectors, where correlations between coordinates exist based on some latent graph structure. On the other hand, random graph models are distributions over graphs, sampled according to some underlying parameters. While existing work on graphical models necessitates many samples from the same distribution (due to parameters outnumbering the samples), our setting requires a single sample from a random graph model.

Our setting is related to the untrusted batches setting of Qiao and Valiant (2018), in which many batches of samples are drawn from a distribution, but a constant fraction of batches may be adversarially corrupted, see also followup works by Jain and Orlitsky (2020a,b, 2021) and Chen, Li, and Moitra (2020a,b). This is somewhat similar to our setting, where each batch is the set of edges connected to a node. However, the key difference is that in our setting, each edge belongs to both its two endpoint nodes, whereas in the untrusted batches setting, a sample is only associated with a single batch.

Estimation in random graph models has also been studied under the constraint of differential privacy (Borgs et al., 2015, 2018; Sealfon and Ullman, 2019). Despite superficial similarities between the two settings, we are unaware of deeper technical connections.

Our setting bears some conceptual similarity to a line of robustness work focused on decomposing a matrix as a sum of a low rank matrix and a sparse matrix (Chandrasekaran et al., 2011; Candès et al., 2011; Hsu et al., 2011). Our true parameter matrix is the rank-1 matrix pJ, where J is the all-ones matrix. However, the uncorrupted adjacency matrix is a sample from the distribution where each entry is a Bernoulli with the corresponding parameter, which is in general not low rank. Furthermore, our corruption model allows for a bounded number of rows/columns to be changed, whereas this line of work requires that the corruptions satisfy some further sparsity, such as a limited number of changed entries per row/column, or that the corruption positions are chosen randomly.

Appendix B. Spectral norm properties

Matrix Properties. We state some useful properties of matrix spectral norm that will be useful in our proofs.

Lemma 22 *Let $M, M' \in \mathbb{R}^{m\times n}$, then $\|M + M'\| \le \|M\| + \|M'\|$.*

Lemma 23 *For any $M \in \mathbb{R}^{m\times n}$, $S \subseteq [m]$, $S' \subseteq [n]$, $\|M_{S\times S'}\| \le \|M\|$.*

Proof For any unit vectors $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$, let $\tilde{u} = u_S/\|u_S\|$ and $\tilde{v} = v_{S'}/\|v_{S'}\|$. Then

$$|u^\intercal M_{S\times S'} v| = |u_S^\intercal M v_{S'}| = \|u_S\| \cdot \|v_{S'}\| \cdot |\tilde{u}^\intercal M \tilde{v}| \le |\tilde{u}^\intercal M \tilde{v}| \le \|M\|,$$

where the second last inequality used $\|u_S\| \le \|u\| = 1$ and $\|v_{S'}\| \le \|v\| = 1$ and the last inequality used that $\tilde{u}$ and $\tilde{v}$ are unit vectors. Finally, in the above equation taking maximum over all unit vectors u, v completes the proof. ■

Lemma 24 *For any $M \in \mathbb{R}^{m\times n}$, $\|M\| \ge \frac{|\sum_{i,j} M_{i,j}|}{\sqrt{mn}}$.*

Proof Consider $u = \frac{1}{\sqrt{m}}[1, 1, \ldots, 1]$ and $v = \frac{1}{\sqrt{n}}[1, 1, \ldots, 1]^T$, which are unit vectors in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively. Then $|u^T M v| = \frac{|\sum_{i,j} M_{i,j}|}{\sqrt{mn}} \le \|M\|$ by (3). ■

Appendix C. Concentration Inequalities

Lemma 25 (Chernoff bound) *Let $X_1, X_2, ..., X_t \sim Ber(p)$ be t independent Bernoulli random variables. Then for any $\lambda > 0$*

$$\Pr\left[\left| \sum_{i=1}^{t} X_i - tp \right| \geq \lambda \right] \leq 2 \exp\left(-\min\left(\frac{\lambda^2}{3tp}, \frac{\lambda}{3} \right) \right). \tag{8}$$

Appendix D. Proofs for Mean- and Median-Based Algorithms

In this section we provide the proofs for algorithms based on mean and medians. Throughout this section we assume that n is at least 14400 for computational simplifications.

D.1. Upper Bounds for Mean and Median Estimators without Corruptions

Mean estimate. The total number of edges in $G \sim G(n,p)$ is a Binomial distribution with parameters $\binom{n}{2}$ and p. Therefore, its expectation and variance are $\binom{n}{2}p$ and $\binom{n}{2}p(1-p)$, respectively. Thus, $\mathbb{E}\left[\hat{p}_{\text{mean}}(G)\right] = p$ and $\mathrm{Var}(p_{\text{mean}}(G)) = p(1-p)/\binom{n}{2} \leq 4p(1-p)/n^2$. By Chebyshev's inequality,

$$\Pr\left(|\hat{p}_{\text{mean}}(G) - p| \geq 20 \cdot \frac{\sqrt{p(1-p)}}{n} \right) \leq 0.01.$$

Median estimate. We will show that with probability at least 0.995, the median degree of G is at least $(n-1)p - C$ for some constant C. The main hurdle in showing this is the fact that the node degrees $\deg(i)$ are not independent, which requires a careful analysis. For $i \in [n]$, let $Y_i := \mathbb{I}\left(\deg(i) \leq p(n-1) - 121\right)$. Then, $\sum_i Y_i$ is the number of nodes with degree at most $p(n-1) - 121$.

We establish the following bounds for $n \geq 14400$:

$$\mathbb{E}\left[\sum_i Y_i\right] \leq \frac{n}{2} - 15\sqrt{n} \tag{9}$$

$$\mathrm{Var}\left(\sum_i Y_i\right) \leq n \tag{10}$$

With these, we can apply Cantelli's inequality to obtain:

$$\Pr\left(\sum_i Y_i \geq \frac{n}{2} \right) \leq \frac{\mathrm{Var}\left(\sum Y_i\right)}{\mathrm{Var}\left(\sum Y_i\right) + (15\sqrt{n})^2} < 0.005.$$

This shows that with probability at least 0.995 the median degree is at least $(n-1)p - 121$. By symmetry, with probability at least 0.995 the median degree is at most $(n-1)p + 121$. By the union bound, with probability at least 0.99 the error of the median estimate is at most $121/(n-1)$.

We now prove (9) and (10) to complete the proof.

To prove (9), note that $\deg(i) \sim Bin(n-1, p)$ and $\mathbb{E}\left[Y_i\right] = \Pr[Bin(n-1,p) \leq p(n-1) - 121]$.

We show that for any n', $\Pr[Bin(n',p) \le pn' - 121] \le \frac{1}{2} - \frac{15}{\sqrt{n'+1}}$, then (9) follows from the linearity of expectation. If $\Pr(Bin(n',p) \le pn' - 1) \le \frac{1}{2} - \frac{15}{\sqrt{n'+1}}$ then we are done. We prove for the case when $\Pr(Bin(n',p) \le pn' - 1) \ge \frac{1}{2} - \frac{15}{\sqrt{n'+1}}$. By Chebyshev's inequality,

$$\Pr\left(Bin(n',p) \le n'p - \sqrt{n}\right) \le \frac{1}{4}.$$

Then, for $n' \ge 14400$,

$$\begin{aligned}\Pr\left(Bin(n',p) \in [n'p - \sqrt{n'}, pn' - 1)\right) &= \Pr(Bin(n',p) \le pn' - 1) - \Pr\left(Bin(n',p) \le n'p - \sqrt{n}\right) \\ &\ge \frac{1}{2} - \frac{15}{\sqrt{n'+1}} - \frac{1}{4} \ge \frac{1}{8}.\end{aligned}$$

Since the binomial distribution has a unique mode $\ge pn' - 1$, then for any $t \le \sqrt{n'}$,

$$\Pr\left(Bin(n',p) \in [n'p - t, pn' - 1)\right) \ge \frac{t-1}{\sqrt{n'}-1} \cdot \frac{1}{8} \ge \frac{t-1}{\sqrt{n'+1}} \cdot \frac{1}{8}.$$

Since the median of $Bin(n',p)$ is $\ge n'p - 1$, (Kaas and Buhrman, 1980), hence $\Pr[Bin(n',p) \le pn' - 1] \le 1/2$. From it subtracting the above equation for $t - 1 = 15 \cdot 8 = 120$, we get $\Pr[Bin(n',p) \le pn' - 121] \le \frac{1}{2} - \frac{15}{\sqrt{n'+1}}$.

We now prove (10). Since Y_i's are identically distributed indicator random variables,

$$\operatorname{Var}\left(\sum_i Y_i\right) = n \operatorname{Var}(Y_1) + n(n-1)\operatorname{Cov}(Y_1, Y_2) \le \frac{n}{4} + n(n-1)\operatorname{Cov}(Y_1, Y_2). \tag{11}$$

Let $t = (n-1)p - 121$, then $Y_i = \mathbb{I}(deg(i) \le t)$. Let Y_{12} be the number of edges from node 1 to $[n] \setminus \{2\}$ and $\mathbb{I}(E_{1,2})$ be the indicator that edge between 1 and 2 is present. Then $Y_{12} \sim Bin(n-2, p)$. Elementary computations using the observation that $Y_1 = \mathbb{I}(Y_{12} \le t-1) + \mathbb{I}(Y_{12} = t) \cdot (1 - \mathbb{I}(E_{1,2}))$ show that

$$\operatorname{Cov}(Y_1, Y_2) = p(1-p) \cdot \Pr(Y_{12} = t)^2.$$

From Stirling's approximation at $t = np$, we have $\Pr(Y_{12} = t) \le 1/\sqrt{\pi p(1-p)(n-2)}$, and therefore,

$$\operatorname{Cov}(Y_1, Y_2) = p(1-p) \cdot \Pr(Y_{12} = t)^2 \le \frac{1}{\pi(n-2)} \le \frac{1}{3n}$$

for $n > 120^2$. Plugging this in (11) proves (10).

D.2. Lower Bounds for Mean and Median Estimators under Corruptions

We will prove the $\gamma/2$ lower bound for the mean and median estimates. Consider the following oblivious adversary $\mathcal{A}$.

- Pick a random subset $B \subset [n]$ of size γn.
- Let $\mathcal{A}_1(G)$ be the graph obtained by adding all edges (u,v) that have at least one node in B to the graph G, and let $\mathcal{A}_2(G)$ be the graph obtained by removing all edges that have at least one node in B from the graph G.

- Output $\mathcal{A}_1(G)$ or $\mathcal{A}_2(G)$ chosen uniformly at random.

Any node in $\mathcal{A}_1(G)$ has degree at least γn more than the corresponding node in $\mathcal{A}_2(G)$. Therefore, $|\hat{p}_{\text{mean}}(\mathcal{A}_1(G)) - \hat{p}_{\text{mean}}(\mathcal{A}_2(G))| \geq \gamma$, and $|\hat{p}_{\text{med}}(\mathcal{A}_1(G)) - \hat{p}_{\text{med}}(\mathcal{A}_2(G))| \geq \gamma$. Therefore by the triangle inequality, with probability 0.5, $|\hat{p}_{\text{mean}}(\mathcal{A}(G)) - p| \geq \gamma/2$, and $|\hat{p}_{\text{med}}(\mathcal{A}(G)) - p| \geq \gamma/2$.

D.3. Upper Bounds for Prune-then-Mean/Median Algorithms

Recall the prune-then mean/median algorithm in Algorithm 1. We remove $c\gamma$ fraction of nodes with the highest and lowest degrees, and then output the median (or mean) of the remaining subgraphs. We restate the performance bound of the algorithm here.

Theorem 8 *For $c \geq 1$ and $0 < \gamma \cdot c < 0.25$, the prune-then-mean and prune-then-median estimators described in Algorithm 1 prune $2c\gamma n$ nodes in total and with probability $1 - n^{-2}$ estimates p to an accuracy $\mathcal{O}\big(c\gamma^2 + \frac{\log n}{n}\big)$ and $\mathcal{O}\big(c\gamma + \sqrt{\frac{\log n}{n}}\big)$, respectively.*

Proof Let $G \sim G(n,p)$. By Chernoff bound (Lemma 25) and the union bound, with probability $\geq 1 - 1/n^2$,

$$\deg(i) \in \left(np - 100\sqrt{n \log n}, np + 100\sqrt{n\log n}\right)$$

for all nodes $i \in [n]$ of G. We condition on this event.

Suppose an adversary converts G into $\mathcal{A}(G)$ by corrupting nodes in $B \subset [n]$ with $|B| \leq \gamma n$. Note that the degree of a node in $F = [n] \setminus B$ cannot change by more than γn. Therefore, for all nodes $i \in F$ in $\mathcal{A}(G)$,

$$\deg(i) \in \left(np - 100\sqrt{n\log n} - \gamma n, np + 100\sqrt{n\log n} + \gamma n\right). \tag{12}$$

Therefore, at most γn nodes do not satisfy (12). Since we remove $c\gamma n$ nodes with the highest and the lowest degrees for $c \geq 1$ all such nodes are pruned. The degree of any node not pruned decreases by at most $2c\gamma n$, and after pruning all degrees are in the following interval

$$\left(np - 100\sqrt{n\log n} - (2c+1)\gamma n, np + 100\sqrt{n\log n} + \gamma n\right). \tag{13}$$

We can rewrite this interval as follows

$$\left(n(1-2c\gamma)p - 100\sqrt{n\log n} + (2cp - 2c - 1)\gamma n, n(1-2c\gamma)p + 100\sqrt{n\log n} + (2cp+1)\gamma n\right).$$

The prune-then-median estimator outputs one of these degrees (normalized), and its error is at most

$$\left(\frac{100\sqrt{n\log n} + (4c+1)\gamma n}{(1-2c\gamma)n}\right) = \mathcal{O}\left(\sqrt{\frac{\log n}{n}} + c\gamma\right).$$

We now bound the performance of prune-then-mean estimator. Let $V' \subseteq [n]$ be the nodes that are not pruned, so $|V'| = (1-2c\gamma)n$. Let $F^p := V' \cap F$ and $B^p := V' \cap B$ be the uncorrupted and corrupted nodes that remain after pruning. We have $|B^p| \leq |B| \leq \gamma n$ and $|F^p| \geq (1-(2c+1)\gamma)n$.

There are three types of edges among the nodes in V': (i) $\mathcal{E}_1$: edges whose both end points are good nodes (in F^p), (ii) $\mathcal{E}_2$: edges with at least one end point in B^p. The mean estimator outputs

$$\frac{|\mathcal{E}_1| + |\mathcal{E}_2|}{\binom{|V'|}{2}}.$$

Its error is at most

$$\begin{aligned}\left|\frac{|\mathcal{E}_1| + |\mathcal{E}_2|}{\binom{|V'|}{2}} - p\right| &= \left|\frac{|\mathcal{E}_1| - \binom{|F^p|}{2}p}{\binom{|V'|}{2}}\right| + \left|\frac{|\mathcal{E}_2| - (|V'| - |F^p|)((|V'| + |F^p| - 1)/2)p}{\binom{|V'|}{2}}\right| \\ &= \left|\frac{|\mathcal{E}_1| - \binom{|F^p|}{2}p}{\binom{|V'|}{2}}\right| + \left|\frac{|\mathcal{E}_2| - |B^p|((|V'| + |F^p| - 1)/2)p}{\binom{|V'|}{2}}\right|\end{aligned}$$

We will bound each term individually. Since the subgraph $F^p \times F^p$ between the uncorrupted nodes remains unaffected from the original graph G, then Theorem 28 implies that, with probability $\geq 1 - 3n^{-2}$,

$$\left|\frac{|\mathcal{E}_1|}{\binom{|F^p|}{2}} - p\right| = \mathcal{O}\left(\max\left\{c\gamma\sqrt{\frac{\ln(e/c\gamma)}{n}}, \frac{c\gamma\log n}{n}, \frac{1}{n}\right\}\right) \leq \mathcal{O}\left(c\gamma^2 + \frac{\log n}{n}\right).$$

Therefore,

$$\left||\mathcal{E}_1| - \binom{|F^p|}{2}p\right| = \binom{|F^p|}{2} \cdot O\left(c\gamma^2 + \frac{\log n}{n}\right) \leq \binom{|V'|}{2} \cdot O\left(c\gamma^2 + \frac{\log n}{n}\right).$$

This shows that the first error term is at most $O\left(c\gamma^2 + \frac{\log n}{n}\right)$.

We now consider the second term. Note that $|n - (|V'| + |F^p| - 1)/2| \leq 3c\gamma n$. By the triangle inequality,

$$\left||\mathcal{E}_2| - \frac{1}{2} \cdot |B^p|(|V'| + |F^p| - 1)p\right| \leq ||\mathcal{E}_2| - |B^p| \cdot np| + 3\gamma np \cdot |B^p|. \tag{14}$$

Let $\deg(i)$ be the degree of node i after pruning. By the triangle inequality adding and subtracting $\sum_{i \in B^p} \deg(i)$ to the first term we obtain,

$$||\mathcal{E}_2| - |B^p| \cdot np| \leq \left||\mathcal{E}_2| - \sum_{i \in B^p} \deg(i)\right| + \sum_{i \in B^p} |\deg(i) - np|.$$

Now note that $|\mathcal{E}_2|$ is the number of edges with at least one endpoint in B^p. Therefore $\left||\mathcal{E}_2| - \sum_{i \in B^p} \deg(i)\right|$ is the number of edges inside $B^p \times B^p$ and is at most $|B^p|^2$. For the second term we use the fact that each node in B^p satisfies (13), and $|B^p| \leq \gamma n$. This gives

$$||\mathcal{E}_2| - |B^p| \cdot np| \leq \left||\mathcal{E}_2| - \sum_{i \in B^p} \deg(i)\right| + \sum_{i \in B^p} |\deg(i) - np| \leq |B^p| \cdot \left(100\sqrt{n \log n} + (2c+2)\gamma n\right).$$

Plugging this along with the fact that $|B^p| \leq \gamma n$ in (14), we obtain

$$\left| |\mathcal{E}_2| - \frac{1}{2} \cdot |B^p|(|V'| + |F^p| - 1)p \right| \leq \gamma n \cdot \left(\left(100\sqrt{n \log n} + (5c+2)\gamma n \right) \right).$$

Since $\binom{|V'|}{2} > (n/2)^2$, the second term can be bounded by

$$\mathcal{O}\left(4\gamma \cdot \left(\sqrt{\frac{\log n}{n}} + (5c+2)\gamma\right)\right) = \mathcal{O}\left(c\gamma^2 + \frac{\log n}{n}\right),$$

thus proving the result.

■

D.4. Lower Bounds for Prune-then-Mean/Median Algorithms

We will prove the following result showing the tight dependence of the upper bounds on γ.

Theorem 9 *Let $p = 0.5$, $\gamma > 100 \cdot \sqrt{\log n/n}$, and $c > 0$ be such that $c\gamma < 0.25$. There exists an adversary such that with probability at least 0.99, the prune-then-median estimate that deletes $c\gamma n$ satisfies $|\hat{p}(\mathcal{A}(G)) - p| \geq C'\gamma$, and the prune-then-mean estimate satisfies $|\hat{p}(\mathcal{A}(G)) - p| \geq C'\gamma^2$.*

Let $G \sim G(n, 0.5)$. The oblivious adversary $\mathcal{A}$ operates as follows. It partitions G into five random sets B, S_0, S_1, S_2, and S_3 with $|B| = \gamma n$, $|S_0| = c\gamma n$, $|S_1| = c\gamma n$, $|S_2| = \frac{2}{3}(1-(2c+1)\gamma)n$, $|S_3| = \frac{1}{3}(1-(2c+1)\gamma)n$.

- Remove all edges with at least one endpoint in B.
- Remove all edges between S_0 and B.
- Add all edges between S_1 and B.
- Connect each node in B to each node in S_2 independently with probability $3/5$.
- Connect each node in B to each node in S_3 independently with probability $3/10$.
- Connect nodes within B to each other with probability $3/5$.

By the Chernoff bound (Lemma 25) and the union bound, we obtain the following bounds on the node degrees in $\mathcal{A}(G)$.

Lemma 26 *In $\mathcal{A}(G)$, the following hold with probability at least $1 - 3n^{-3}$*

$$\deg(u) = n\left(\frac{1}{2} + \frac{\gamma}{10}\right) \pm 4\sqrt{n \log n} \qquad \textit{for } u \in B,$$

$$\deg(u) = n \cdot \left(\frac{1}{2} - \frac{\gamma}{2}\right) \pm 4\sqrt{n \log n} \qquad \textit{for } u \in S_0,$$

$$\deg(u) = n \cdot \left(\frac{1}{2} + \frac{\gamma}{2}\right) \pm 4\sqrt{n \log n} \qquad \textit{for } u \in S_1,$$

$$\deg(u) = n \cdot \left(\frac{1}{2} + \frac{\gamma}{10}\right) \pm 4\sqrt{n \log n} \qquad \textit{for } u \in S_2,$$

$$\deg(u) = n \cdot \left(\frac{1}{2} - \frac{\gamma}{5}\right) \pm 4\sqrt{n \log n} \qquad \textit{for } u \in S_3.$$

Since $\gamma > 100\sqrt{\log n/n}$, the nodes in S_0 are the $c\gamma n$ nodes with the lowest degrees and the nodes in S_1 are the $c\gamma n$ nodes with the highest degrees, and they are pruned by the algorithm. Now since the sets S_0 and S_1 were randomly chosen ahead of time, in the pruned graph, once again by the Chernoff bound (Lemma 25) and the union bound, the following holds with probability at least $1 - 3n^{-3}$

$$\deg(u) = n\left(\frac{1-2c\gamma}{2} + \frac{\gamma}{10}\right) \pm 8\sqrt{n\log n} \qquad \text{for } u \in B,$$
$$\deg(u) = n \cdot \left(\frac{1-2c\gamma}{2} + \frac{\gamma}{10}\right) \pm 8\sqrt{n\log n} \qquad \text{for } u \in S_2,$$
$$\deg(u) = n \cdot \left(\frac{1-2c\gamma}{2} - \frac{\gamma}{5}\right) \pm 8\sqrt{n\log n} \qquad \text{for } u \in S_3.$$

Since we assume that $c\gamma < 0.25$, there are more nodes in S_3 than in $S_2 \cup B$ and every node in $S_2 \cup B$ had a higher degree than any node in S_3. Therefore a node in S_3 is chosen as the median node, thus deviating from the median degree by at least $\gamma/5 \pm 8\sqrt{\log n/n} > \gamma/10$ for $\gamma > 100\sqrt{\log n/n}$. This proves the lower bound for prune-then-median estimate.

Now for the prune-then-mean estimate, note that each edge that remains after pruning is chosen at random, independent of all other edges. The total expected number of edges after pruning is $\frac{1}{2} \cdot \frac{n^2(1-2c\gamma)^2}{2} + \frac{n^2\gamma^2}{20}$ and the variance is at most $n^2/4$. Therefore, the total error of the prune-then-mean estimate is at least $\gamma^2/20 \pm O(1/n)$, and since $\gamma > 100\sqrt{\log n/n}$, the error is at least $\gamma^2/40$.

Appendix E. Upper Bound Proofs

E.1. Proof of Lemma 11

Proof We first prove Equation (6). From the triangle inequality

$$\|(A - p_{F'})_{F' \times F'}\| \le \|(A - p)_{F' \times F'}\| + |p - p_{F'}| \cdot F'.$$

From Lemma 24 we have

$$\|(A - p)_{F' \times F'}\| \ge |F'| \cdot |p_{F'} - p|.$$

Combining the above two equations with regularity proves Equation (6),

$$\|(A - p_{F'})_{F' \times F'}\| \le 2\|(A - p)_{F' \times F'}\| \le 2n \cdot \eta(p, n),$$

where the last inequality follows from regularity condition 2.

Next, Equation (7) is obtained by using $F' = F''$ in regularity condition 3 and $|F'| \ge n/2$. ∎

E.2. Proof of Theorem 12

Proof In a γ-corrupted graph the set of uncorrupted nodes F has size $\ge (1-\gamma)n$, which proves regularity condition 1.

We use the following bound on the spectral norm of a centered version of $\tilde{A}$, which follows from Remark 3.13 of (Bandeira and Van Handel, 2016).

Lemma 27 *Let $\tilde{A}$ be the adjacency matrix of a sample from $G(n,p)$ and I be the $n \times n$ identity matrix. There exist a universal constant c such that with probability at least $1-n^{-2}$, $\|\tilde{A}-p+pI\| \leq c\sqrt{np(1-p)+\ln n}$.*

To establish regularity condition 2, note that A and $\tilde{A}$ agree on $(i,j) \in F \times F$, and therefore by Lemma 23 and Lemma 27, $\|(A-p)_{F' \times F'}\| = \|(\tilde{A}-p)_{F' \times F'}\| \leq \|(\tilde{A}-p)\| \leq \|(\tilde{A}-p+pI)\| + p\|I\| \leq c\sqrt{np(1-p)+\ln n}+1$.

The following theorem implies regularity condition 3. The proof uses a Chernoff and union bound style argument, and is provided in Section F.

Theorem 28 *Let $\tilde{A}$ be the adjacency matrix of a sample from $G(n,p)$. With probability at least $1-3n^{-2}$, simultaneously for all $\alpha \in [0, \frac{1}{2}]$, we have*

$$\max_{|S|,|S'| \in C_\alpha} \left| \sum_{i \in S,\, j \in S'} (\tilde{A}_{i,j} - p) \right| \leq 6 \max \left\{ 16\alpha n \sqrt{pn \ln \frac{e}{\alpha}}, 60\alpha n \ln \frac{e}{\alpha}, 5n\sqrt{p \ln(en)} \right\},$$

where we define $C_\alpha := [0, \alpha n] \cup [n-\alpha n, n]$.

■

E.3. Proofs for Lemma 16

Proof We first require the following lemma, which lower bounds the spectral norm of a matrix $(A-p_S)_{S\times S}$ primarily in terms of the empirical estimates of p corresponding to the submatrices induced by S and $S \cap F$. The proof appears in Section E.4.

Lemma 29 *Given any symmetric matrix A, and subsets $S, F \subseteq [n]$*

$$\|(A-p_S)_{S\times S}\| \geq \frac{|p_{S\cap F} - p_S| \cdot |S \cap F|}{3} \cdot \min\left\{ \sqrt{\frac{|S\cap F|}{|S \cap F^c|}}, \frac{|S\cap F|}{|S\cap F^c|} \right\}.$$

For $\alpha_1 \leq 1/60$ and $|S| \geq (1-9\alpha_1)n$, we can deduce that $|S \cap F| \geq n(1-10\alpha_1) \geq 5n/6$ and $|S \cap F^c| \leq |F^c| \leq \alpha_1 n \leq n/60$. Therefore, $|S\cap F^c|/|S \cap F| \leq 1/50$. By Lemma 29,

$$|S\cap F| \cdot |p_{S\cap F} - p_S| \leq 3\|(A-p_S)_{S\times S}\| \max\left\{ \sqrt{\frac{|S\cap F^c|}{|S\cap F|}}, \frac{|S\cap F^c|}{|S\cap F|} \right\} \leq \frac{3}{\sqrt{50}} \cdot \|(A-p_S)_{S\times S}\|.$$

Applying Equation (6) with $F' = S \cap F$, we have

$$\|(A-p_{S\cap F})_{(S\cap F)\times(S\cap F)}\| \leq 2n \cdot \eta(p,n).$$

This implies $\|(A-p_{S\cap F})_{(S\cap F)\times(S\cap F)}\| \leq 0.1\|(A-p_S)_{S\times S}\|$. Next, by the triangle inequality,

$$\begin{aligned}\|(A-p_S)_{(S\cap F)\times(S\cap F)}\| &\leq \|(A-p_{S\cap F})_{(S\cap F)\times(S\cap F)}\| + |S\cap F|\cdot|p_{S\cap F}-p_S| \\ &\leq \left(\frac{1}{10} + \frac{3}{\sqrt{50}}\right)\|(A-p_S)_{S\times S}\|.\end{aligned}$$

To interpret the derivation above: we have reasoned that if the spectral norm of $(A - p_S)_{S\times S}$ is large, the contribution due to $S \cap F$ (i.e., the submatrix induced by the intersection with the good nodes) is relatively small. This suggests that any top eigenvector must place a constant mass on $S \cap F^c$. Indeed, the following theorem formalizes this reasoning, showing that the normalized top eigenvector contains significant weight in this complementary subset of indices. The proof appears in Section E.5.

Theorem 30 *Let M be a non-zero $n \times n$ real symmetric matrix such that for some set $S \subseteq [n]$ and $0 \le \rho \le 1$ we have $\|M_{S\times S}\| \le \rho \|M\|$. Let v be any normalized top eigenvector of M. Then $\|v_{S^c}\|^2 \ge \frac{(1-\rho)^2}{1+(1-\rho)^2}$.*

Applying Theorem 30 with $\rho = \frac{1}{10} + \frac{3}{\sqrt{50}}$ implies that $\|v_{S\setminus(S\cap F)}\|^2 = \|v_{S\cap F^c}\|^2 \ge \frac{(1-\rho)^2}{1+(1-\rho)^2} > 0.15$. ■

E.4. Proof of Lemma 29

First note that

$$0 = \sum_{i,j\in S}(A_{i,j} - p_S) = \sum_{i,j\in S\cap F}(A_{i,j} - p_S) + \sum_{i,j\in S\cap F^c}(A_{i,j} - p_S) + 2\sum_{i\in S\cap F,\, j\in S\cap F^c}(A_{i,j} - p_S).$$

Therefore,

$$\left|\sum_{i,j\in S\cap F}(A_{i,j} - p_S)\right| \le \left|\sum_{i,j\in S\cap F^c}(A_{i,j} - p_S)\right| + 2\left|\sum_{i\in S\cap F,\, j\in S\cap F^c}(A_{i,j} - p_S)\right|.$$

Hence,

$$\frac{|\sum_{i,j\in S\cap F}(A_{i,j} - p_S)|}{3} \le \max\left\{\left|\sum_{i,j\in S\cap F^c}(A_{i,j} - p_S)\right|, \left|\sum_{i\in S\cap F,\, j\in S\cap F^c}(A_{i,j} - p_S)\right|\right\}. \tag{15}$$

From Lemma 23 , Lemma 24 and the above inequality, it follows that

$$\begin{aligned}
\|(A - p_S)_{S\times S}\| &\ge \max\left\{\|(A - p_S)_{(S\cap F^c)\times(S\cap F^c)}\|, \|(A - p_S)_{(S\cap F^c)\times(S\cap F)}\|\right\} && (16)\\
&\ge \max\left\{\frac{|\sum_{i,j\in S\cap F^c}(A_{i,j} - p_S)|}{|S\cap F^c|}, \frac{|\sum_{i\in S\cap F^c,\, j\in S\cap F}(A_{i,j} - p_S)|}{\sqrt{|S\cap F|\cdot|S\cap F^c|}}\right\} && (17)\\
&\ge \min\left\{\frac{|\sum_{i,j\in S\cap F}(A_{i,j} - p_S)|}{3|S\cap F^c|}, \frac{|\sum_{i,j\in S\cap F}(A_{i,j} - p_S)|}{3\sqrt{|S\cap F|\cdot|S\cap F^c|}}\right\} && (18)\\
&= \frac{|\sum_{i,j\in S\cap F}(A_{i,j} - p_S)|}{3\sqrt{|S\cap F|\cdot|S\cap F^c|}}\cdot\min\left\{\sqrt{\frac{|S\cap F|}{|S\cap F^c|}}, 1\right\}\\
&= \frac{|p_{S\cap F} - p_S||S\cap F|}{3}\cdot\min\left\{\frac{|S\cap F|}{|S\cap F^c|}, \sqrt{\frac{|S\cap F|}{|S\cap F^c|}}\right\},
\end{aligned}$$

where (16) is from Lemma 23, (17) follows from Lemma 24, (18) from (15).

E.5. Proof of Theorem 30

Since eigenvalues of symmetric matrices are real, let $v \in \mathbb{R}^n$ be the normalized top eigenvector of M with eigenvalue $\lambda \in \mathbb{R}$ such that $Mv = \lambda v$ and $\|M\| = |\lambda|$. Since $Mv = \lambda v$, we have $M_{S\times[n]}\, v \; = \lambda v_S$, and

$$M_{S\times[n]}\, v \; = M_{S\times S}\, v_S + M_{S\times S^c}\, v_{S^c} \tag{19}$$

By Lemma 22 on (19),

$$\begin{aligned} & \|M_{S\times[n]}\, v\| \leq \|M_{S\times S}\, v_S\| + \|M_{S\times S^c}\, v_{S^c}\| \\ \Rightarrow\quad & |\lambda| \cdot \|v_S\| \leq \rho|\lambda| \cdot \|v_S\| + |\lambda| \cdot \|v_{S^c}\| \qquad (20) \\ \Rightarrow\quad & (1-\rho)\|v_S\| \leq \|v_{S^c}\| \\ \Rightarrow\quad & (1-\rho)^2\|v_S\|^2 \leq \|v_{S^c}\|^2 \end{aligned}$$

where (20) uses the assumption of the lemma. Finally using $\|v_S\|^2 + \|v_{S^c}\|^2 = 1$ gives the bound.

E.6. An Approximate Top Eigenvector Suffices

As discussed in Remark 17, computing an exact top eigenvector in Algorithm 2 may be costly. The guarantees associated with this top eigenvector are quantified in Lemma 16, which relies upon Theorem 30. In this section, we prove a variant of Theorem 30, which works with an approximate rather than an exact top eigenvector. By repeating the proof of Lemma 16 with Lemma 31 swapped in place of Theorem 30, we can instead use approximate top eigenvector procedures, reducing the runtime.

Lemma 31 *Let M be a nonzero $n \times n$ real matrix such that for some set $S \subset [n]$ we have $\|M_{S\times S}\| \leq 0.53\|M\|$. Let $v \in \mathbb{R}^n$ be a unit vector such that $\|Mv\| \geq 0.99\|M\|$, then $\|v_{S^c}\|^2 \geq \frac{1}{8}$.*

Proof Let $u = Mv$. Note that $M_{S\times[n]}\, v = u_S$ and $M_{S^c\times[n]}\, v = u_{S^c}$, therefore

$$v^T\, M\, v = v^T\, (M_{S\times[n]} + M_{S^c\times[n]})\, v = v^T(u_S + u_{S^c}) = v_S^T\, u_S + v_{S^c}^T\, u_{S^c}.$$

Then by the triangle inequality,

$$\begin{aligned} & |v^T\, M\, v| \leq \|v_S\| \cdot \|u_S\| + \|v_{S^c}\| \cdot \|u_{S^c}\| \\ \Rightarrow\quad & 0.99\|M\| \leq \|v_S\| \cdot \|u_S\| + \|v_{S^c}\| \cdot \|u_{S^c}\| \\ \Rightarrow\quad & 0.99\|M\| \leq \sqrt{1 - \|v_{S^c}\|^2} \cdot \|u_S\| + \|v_{S^c}\| \cdot \sqrt{\|M\|^2 - \|u_S\|^2}. \end{aligned}$$

In the last line, we used the fact that $\|u\| \leq \|M\| \cdot \|v\| = \|M\|$ and $\|u\|^2 = \|u_S\|^2 + \|u_{S^c}\|^2$. Rearranging this expression, it is easy to show that in the case $\|u_S\|^2 \leq \frac{3\|M\|^2}{4}$, the inequality is violated if $\|v_{S^c}\|^2 \leq \frac{1}{8}$. Therefore, $\|u_S\|^2 \leq \frac{3\|M\|^2}{4}$ implies $\|v_{S^c}\|^2 \geq \frac{1}{8}$.

To prove the lemma, we must handle the remaining case: we show that if $\|u_S\|^2 \geq \frac{3\|M\|^2}{4}$, then $\|v_{S^c}\|^2 \geq \frac{1}{8}$.

Note that

$$M_{S\times[n]}\, v \; = M_{S\times S}\, v_S + M_{S\times S^c}\, v_{S^c}.$$

Then

$$
\begin{aligned}
& \|M_{S\times[n]}\, v\| \le \|M_{S\times S}\, v_S\| + \|M_{S\times S^c}\, v_{S^c}\| \\
\Rightarrow \quad & \|u_S\| \le 0.53\|M\|\cdot\|v_S\| + \|M\|\cdot\|v_{S^c}\| \\
\Rightarrow \quad & \|u_S\|^2 \le 2(0.53^2)\|M\|^2\cdot\|v_S\|^2 + 2\|M\|^2\cdot\|v_{S^c}\|^2 \\
\Rightarrow \quad & \|u_S\|^2 \le 0.5618\|M\|^2(1-\|v_{S^c}\|^2) + 2\|M\|^2\cdot\|v_{S^c}\|^2 \\
\Rightarrow \quad & \|u_S\|^2 \le 0.5618\|M\|^2 + 1.4382\|M\|^2\cdot\|v_{S^c}\|^2.
\end{aligned}
$$

When $\|u_S\|^2 \ge 3\|M\|^2/4$, the above equation implies $\|v_{S^c}\|^2 \ge 1/8$, which completes the proof of the lemma. ■

E.7. Proofs for Theorem 18

Before proving the Theorem we state and prove two auxiliary lemmas. The first lemma shows that the average of entries of all *small submatrices* of $S^* \times S^*$ are close to p_{S^*}.

Lemma 32 *Assume the conditions of Theorem 18 hold. For all $S_1, S_2 \subseteq S^*$ with $|S_1|, |S_2| \le 3\alpha_1 n$ we have*

$$\left|\sum_{i\in S_1,\, j\in S_2} (A_{i,j} - p_{S^*})\right| \le 60\alpha_1 n^2 \cdot \eta(p,n).$$

Proof Since $\|(A - p_{S^*})_{S^*\times S^*}\| \le 20n\cdot\eta(p,n)$, and using Lemma 23 and Lemma 24, we get

$$\left|\sum_{(i,j)\in S_1\times S_2} (A_{i,j} - p_{S^*})\right| \le \sqrt{|S_1|\cdot|S_2|}\cdot\|(A-p_{S^*})_{S_1\times S_2}\| \le 3\alpha_1 n\|(A-p_{S^*})_{S^*\times S^*}\| \le 60\alpha_1 n^2\cdot\eta(p,n).$$

■

We now show that all the nodes in S^f have normalized degree close to p_{S^*}.

Lemma 33 *Assume the conditions of Theorem 18 hold, and let S^f be the output of Algorithm 3, then for every node $i \in S^f$,*

$$|p_{S^*}^{(i)} - p_{S^*}| \le \left(\frac{2\kappa(13\alpha_1, p, n)}{\alpha_1} + 210\eta(p,n)\right).$$

Proof Suppose to the contrary that after $3\alpha_1 n$ nodes are deleted by Algorithm 3, there is a node $i \in S^f$ such that $|p_{S^*}^{(i)} - p_{S^*}| > \left(\frac{2\kappa(13\alpha_1,p,n)}{\alpha_1} + 210\eta(p,n)\right)$. Therefore, all the nodes deleted by Algorithm 3 are such that $|p_{S^*}^{(i)} - p_{S^*}| > \left(\frac{2\kappa(13\alpha_1,p,n)}{\alpha_1} + 210\eta(p,n)\right)$. Let D^+ be the set of nodes deleted by Algorithm 3 such that $p_{S^*}^{(i)} > p_{S^*}$ for $i \in D^+$ and D^- be the set of deleted nodes i such that $p_{S^*}^{(i)} < p_{S^*}$ for $i \in D^-$. Since $|D^+| + |D^-| = 3\alpha_1 n$ and $|(D^+ \cup D^-) \setminus F| \le |F^c| \le \alpha_1 n$, we have that $|D^+ \cap F| \geq \alpha_1 n$ or $|D^- \cap F| \geq \alpha_1 n$. Suppose $|D^+ \cap F| \geq \alpha_1 n$. Let $F' = D^+ \cap F$. Then, using $|F'| \geq \alpha_1 n$ and $|S^*| > n/2$, we have

$$\sum_{i \in F', j \in S^*} (A_{i,j} - p_{S^*}) = \sum_{i \in F'} |S^*|(p_{S^*}^{(i)} - p_{S^*}) > |F'| \cdot |S^*| \cdot \Big(\frac{2\kappa(13\alpha_1, p, n)}{\alpha_1} + 210\eta(p,n)\Big)$$
$$\geq n^2 \kappa(13\alpha_1, p, n) + 105|F'|n \cdot \eta(p,n).$$

Now, note that

$$\sum_{i \in F', j \in S^*} (A_{i,j} - p_{S^*}) = \sum_{i \in F', j \in S^* \cap F} (A_{i,j} - p) + |F'| \cdot |S^* \cap F| \cdot (p - p_{S^*}) + \sum_{i \in F', j \in S^* \cap F^c} (A_{i,j} - p_{S^*})$$

By Lemma 32 with $S_1 = F'$ and $S_2 = S^* \cap F^c$ the last term in the expression above is at most $60\ \alpha_1\ n^2 \cdot \eta(p,n)$. For the second term note that $|p - p_{S^*}| < 45 \cdot \eta(p,n)$ and therefore, the second term is at most $45|F'|n \cdot \eta(p,n)$. Finally using regularity condition 3 with F' and $F'' = S^* \cap F$ and $\alpha_2 = 13\alpha_1$ bounds the first term by $n^2 \cdot \kappa(13\alpha_1, p, n)$. Combining the three bounds and using $|F'| \geq \alpha_1 n$,

$$\sum_{i \in F', j \in S^*} (A_{i,j} - p_{S^*}) \leq n^2 \kappa(13\alpha_1, p, n) + (45|F'| + 60\alpha_1 n) n \cdot \eta(p,n) \leq n^2 \kappa(13\alpha_1, p, n) + 105|F'| n \cdot \eta(p,n),$$

This shows the contradiction and completes the proof for the case $|D^+ \cap F| > \alpha_1 n$. The case when $|D^- \cap F| > \alpha_1 n$ has a similar argument and is omitted. ■

Combining these lemmas appropriately allows us to conclude our main result on the guarantees of Algorithm 3.

Proof of Theorem 18 We will partition $S^f \times S^f$ into the following groups and bound each term separately.

$$\sum_{i,j \in S^f} A_{i,j} = \sum_{i,j \in S^f \cap F} A_{i,j} + 2 \sum_{i \in S^f, j \in S^f \cap F^c} A_{i,j} - \sum_{i,j \in S^f \cap F^c} A_{i,j}.$$

Since $p_{S^f} = \sum_{i,j \in S^f} A_{i,j}/|S^f|^2$, by the triangle inequality,

$$|p_{S^f} - p| \leq \left|\frac{\sum_{i,j \in S^f \cap F}(A_{i,j} - p)}{|S^f|^2}\right| + 2\left|\frac{\sum_{i \in S^f\, j \in S^f \cap F^c}(A_{i,j} - p)}{|S^f|^2}\right| + \left|\frac{\sum_{i,j \in S^f \cap F^c}(A_{i,j} - p)}{|S^f|^2}\right|.$$

For the first term, $|S^f \cap F| \geq (1 - 13\alpha_1)n \geq n/2$. Using Equation (7) with $F' = S^f \cap F$ and $\alpha_2 = 13\alpha_1$,

$$\left|\frac{\sum_{i,j \in S^f \cap F}(A_{i,j} - p)}{|S^f|^2}\right| \leq \left|\frac{\sum_{i,j \in S^f \cap F}(A_{i,j} - p)}{|S^f \cap F|^2}\right| = |p_{S^f \cap F} - p| \leq 4\kappa(13\alpha_1, p, n).$$

Since $|S^f \times (S^f \cap F^c)| \le \alpha_1 n^2$, and $|S^f| \ge n/2$, by the triangle inequality, the second term is bounded by

$$2\left|\frac{\sum_{i\in S^f\, j\in S^f\cap F^c}(A_{i,j}-p)}{|S^f|^2}\right| \le 2\left|\frac{\sum_{i\in S^f\, j\in S^f\cap F^c}(A_{i,j}-p_{S^*})}{n^2/4}\right| + \frac{2\alpha_1 n^2}{n^2/4}|p_{S^*}-p|$$
$$\le 8\left|\frac{\sum_{i\in S^*\, j\in S^f\cap F^c}(A_{i,j}-p_{S^*})}{n^2}\right| + 8\left|\frac{\sum_{i\in S^*\setminus S^f\, j\in S^f\cap F^c}(A_{i,j}-p_{S^*})}{n^2}\right|$$
$$+ 8\alpha_1 \cdot 45 \cdot \eta(p,n).$$

Since $|S^* \setminus S^f| \le 3\alpha_1 n$ and $|S^f \cap F^c| \le \alpha_1 n$, by taking $S_1 = S^* \setminus S^f$ and $S_2 = S^f \cap F^c$ in Lemma 32 bounds the second term above by $8(60\alpha_1 \cdot \eta(p,n))$. For the first term,

$$\left|\frac{\sum_{i\in S^*\, j\in S^f\cap F^c}(A_{i,j}-p_{S^*})}{n^2}\right| \le \sum_{j\in S^f\cap F^c}\left|\frac{\sum_{i\in S^*}(A_{i,j}-p_{S^*})}{n^2}\right|$$
$$\le \sum_{j\in S^f\cap F^c}\frac{|S^*|}{n^2}\left|\frac{\sum_{i\in S^*}(A_{i,j}-p_{S^*})}{|S^*|}\right|$$
$$\le \sum_{j\in S^f\cap F^c}\frac{1}{n}|p_{S^*}^{(j)}-p_{S^*}|$$
$$\le \alpha_1 \cdot \Big(\frac{2\kappa(13\alpha_1,p,n)}{\alpha_1} + 210\eta(p,n)\Big),$$

where we use Lemma 33 and $|S^f \cap F^c| \le \alpha_1 n$.

For the final term, since $|(S^f \cap F^c) \times (S^f \cap F^c)| \le \alpha_1^2 n^2$,

$$\left|\frac{\sum_{i,j\in S^f\cap F^c}(A_{i,j}-p)}{|S^f|^2}\right| \le \left|\frac{\sum_{i,j\in S^f\cap F^c}(A_{i,j}-p_{S^*})}{|S^f|^2}\right| + |p_{S^*}-p| \cdot \frac{|S^f\cap F^c|^2}{|S^f|^2},$$

which can be bounded again by taking $S_1 = S_2 = S^f \cap F^c$ in Lemma 32. ■

E.8. Putting Things Together: Proof of Theorem 3

We now combine our methods from previous sections to prove our main upper bound. This primarily consists of running Algorithm 2 followed by Algorithm 3, as described by Algorithm 4 and quantified by Theorem 34. For technical reasons, to get the correct scaling of the error with respect to the parameter p, we run this procedure on both the graph and its complement, and output the appropriate of the two estimates. This is described in Algorithm 5, and quantified in Theorem 35. This theorem implies our upper bound (Theorem 3).

Theorem 34 *Suppose $\alpha_1 \in [\frac{1}{n}, \frac{1}{60}]$ and let A be an adjacency matrix containing an $(\alpha_1, 13\alpha_2, p)$-regular subgraph. With probability at least $1-n^{-2}$, Algorithm 4 outputs p_{S^f} such that for some universal constants $c_2, c_3 > 0$,*

$$\big|p_{S^f} - p\big| \le c_2\alpha_1\eta(p,n) + c_3\kappa(13\alpha_1,p,n).$$

The running time of this algorithm is $\tilde{O}(\alpha_1 n^3)$.

Proof The estimation guarantees in Theorem 34 follows by combining the guarantees of Theorems 15, and 18. We conclude the proof by analyzing the running time. As discussed in Remark 17, Algorithm 2 can be implemented in $\tilde{\mathcal{O}}(\alpha_1 n^3)$ time. Algorithm 3 takes $\mathcal{O}(n^2)$ time. Hence, Algorithm 4 runs in $\tilde{\mathcal{O}}(\alpha_1 n^3)$ time. ■

Algorithm 4 Algorithm for estimating p

Require: number of nodes n, parameter $\alpha_1 \in [1/n, 1/60]$, adjacency matrix A
  $S^* \leftarrow$ run the spectral algorithm (Algorithm 2) with inputs n, α_1, A
  $p_{Sf} \leftarrow$ run the trimming algorithm (Algorithm 3) with inputs n, α_1, A, S^*
  return p_{Sf}

Observe that the $\kappa(13\alpha_1, p, n)$ error term in Theorem 34 scales proportional to $\sqrt{p}$, which gives improved error when p is close to 0. To enjoy the same improvement for p close to 1, we can run the algorithm on the complement of the graph. Theorem 35 describes the resulting guarantees, and the procedure appears as Algorithm 5. Note that we apply Theorem 12 to convert from adjacency matrices containing regular subgraphs (which we have considered up to this point) back to our original problem.

Theorem 35 *Suppose $\gamma \in [\frac{1}{n}, \frac{1}{60}]$ and $p \in [0,1]$. Let $G \sim G(n,p)$, and A be the adjacency matrix of a rewiring of G by a γ-omniscient adversary. With probability at least $1 - 10n^{-2}$, running Algorithm 5 will output a $\hat{p}$ such that*

$$|\hat{p} - p| \le C \cdot \left(\frac{\sqrt{p(1-p)\log n}}{n} + \frac{\gamma\sqrt{p(1-p)\log(1/\gamma)}}{\sqrt{n}} + \frac{\gamma}{n}\log n \right),$$

for some universal constant C. The running time of this algorithm is $\tilde{O}(\gamma n^3)$.

Proof Theorem 34 and Theorem 12 imply that with probability $\ge 1 - 5n^{-2}$, p^* in Algorithm 5 satisfies:

$$\left|p^* - p\right| \le c_2 \gamma \eta(p,n) + c_3 \kappa(13\gamma, p, n). \tag{21}$$

By symmetry, with probability $\ge 1 - 5n^{-2}$, q^* in Algorithm 5 satisfies:

$$\left|q^* - (1-p)\right| \le c_2 \gamma \eta(1-p,n) + c_3 \kappa(13\gamma, 1-p, n). \tag{22}$$

When $p \le 0.1$, equation (21) implies $p^* \le 0.5$, and hence $\hat{p} = p^*$ and $|\hat{p} - p| = |p^* - p|$. Similarly, when $p \ge 0.9$, (21) implies $p^* > 0.5$, and hence $\hat{p} = 1 - q^*$ and $|\hat{p} - p| = |(1-q^*) - p| = |(1-p) - q^*|$. Finally, for $0.1 \le p \le 0.9$, we have $|\hat{p} - p| \le \max\{|p^* - p|, |q^* - (1-p)|\}$. Combining the bound for the three cases completes the proof. ■

Algorithm 5 Algorithm for Robust Erdős-Rényi parameter estimation

Require: number of nodes n, parameter $\gamma \in [1/n, 1/60]$, adjacency matrix A
  $p^* \leftarrow$ run Algorithm 4 with inputs n, γ, A
  $q^* \leftarrow$ run Algorithm 4 with inputs n, γ, $(1 - I - A)$ (1 and I are the $n \times n$ all-ones and identity matrix)
  if $p^* \leq 0.5$ **then**
    $\hat{p} \leftarrow p^*$
  else
    $\hat{p} \leftarrow 1 - q^*$
  end if
  return $\hat{p}$

Appendix F. Proof of Theorem 28

Throughout this proof, let $\beta = \max\left\{16\alpha n\sqrt{pn\ln\frac{e}{\alpha}}, 60\alpha n\ln\frac{e}{\alpha}, 5n\sqrt{p\ln(en)}\right\}$. First fix $\alpha \in [0, 1/2]$.

We first consider the entire matrix $\tilde{A}$, namely $S = S' = [n]$. Recall that the diagonal entries of $\tilde{A}$ are zero. Then, note that $\sum_{(i,j)\in[n]\times[n]}(\tilde{A}_{i,j} - p) = 2\cdot\sum_{(i,j)\in[n]\times[n]:\, i>j}(\tilde{A}_{i,j} - p) - np$. Now since all the entries $\tilde{A}_{ij}$ are independent for $i > j$, we can apply the Chernoff bound (Equation (8)) with $\lambda = \beta$ over these entries and with probability at least $1 - n^{-3}$,

$$\left|\sum_{(i,j)\in[n]\times[n]:\, i>j}(\tilde{A}_{i,j} - p)\right| \leq \beta. \tag{23}$$

Since $np \leq n\sqrt{p} \leq \beta$, then from the above equation we get $|\sum_{(i,j)\in[n]\times[n]}(\tilde{A}_{i,j} - p)| \leq 3\beta$, with probability at least $1 - n^{-3}$. Note that for $\alpha < 1/n$ the statement only applies to $S = S' = [n]$, and thus this case is handled. In the remaining proof $\alpha \in [1/n, 1/2]$.

Conditioned on the event $|\sum_{(i,j)\in[n]\times[n]}(\tilde{A}_{i,j} - p)| \leq 3\beta$, note that for all $T \subset [n] \times [n]$,

$$\left|\sum_{(i,j)\in T}(\tilde{A}_{i,j} - p)\right| > 6\beta \Rightarrow \left|\sum_{(i,j)\in T^c}(\tilde{A}_{i,j} - p)\right| > 3\beta, \tag{24}$$

where $T^c = [n] \times [n] \setminus T$. In particular, if $T = S \times S'$ with $|S| \geq n - \alpha n$ and $|S'| \geq n - \alpha n$, then $|T^c| < 2\alpha n^2$ and if $\min\{|S|, |S'|\} \leq \alpha n$, then $|T| \leq \alpha n^2$. Therefore, for $T = S \times S'$ with $|S|$, $|S'| \in C_\alpha$, either $|T|$ or $|T^c|$ is smaller than $2\alpha n^2$. With this in hand, the theorem will follow from the following lemmas.

Lemma 36 *Let $T \subset [n] \times [n]$ be a given subset of size at most $2\alpha n^2$, then*

$$\Pr\left[\left|\sum_{(i,j)\in T}(\tilde{A}_{i,j} - p)\right| \geq 3\beta\right] \leq 4\exp\left(-20\alpha n \ln e/\alpha\right).$$

We now bound the number of subsets of interest.

Lemma 37 *For a given $\alpha \in [1/n, 1/2]$, the number of sets S, S' with $|S|, |S'| \in C_\alpha$ is at most* $4\exp(4\alpha n \ln(e/\alpha))$.

For a given $\alpha \in [1/n, 1/2]$ and $T = S \times S'$ such that $|S|, |S'| \in C_\alpha$, since either of T or T^c have size $\leq 2\alpha n^2$, therefore, combining the two lemmas implies that with probability $\geq 1 - 16\exp(-16\alpha n \ln e/\alpha) \geq 1 - n^3$,

$$\min\left\{\left|\sum_{(i,j)\in T}(\tilde{A}_{i,j} - p)\right|, \left|\sum_{(i,j)\in T^c}(\tilde{A}_{i,j} - p)\right|\right\} \leq 3\beta.$$

Then from Equation (24), with probability $\geq 1 - n^3 - n^3$, $\left|\sum_{(i,j)\in T}(\tilde{A}_{i,j} - p)\right| \leq 6\beta$. This completes the proof for a given value of α. To extend it to all $\alpha \in [1/n, 1/2]$ first note that it suffices to prove the theorem for $\alpha \in \{\frac{1}{n}, \frac{2}{n}, ..., \frac{\lfloor 0.5n \rfloor}{n}\}$, and then upon taking the union bound over these values of α completes the proof.

We now prove Lemma 36. Note that

$$\sum_{(i,j)\in T}(\tilde{A}_{i,j} - p) = \sum_{(i,j)\in T: i>j}(\tilde{A}_{i,j} - p) + \sum_{(i,j)\in T: i<j}(\tilde{A}_{i,j} - p) - \sum_{(i,i)\in T} p. \tag{25}$$

Then using the triangle inequality, $\{(i,i) \in T\} \leq n$ and $np \leq \beta$ to disregard the third term (as done before),

$$\Pr\left[\left|\sum_{(i,j)\in T}(\tilde{A}_{i,j}-p)\right| \geq 2\beta\right] \leq \Pr\left[\left|\sum_{(i,j)\in T: i>j}(\tilde{A}_{i,j}-p)\right| \geq \beta\right] + \Pr\left[\left|\sum_{(i,j)\in T: i<j}(\tilde{A}_{i,j}-p)\right| \geq \beta\right].$$

The two events on the right hand side are for sums of independent mean-centered Bernoulli random variables. We will now apply the Chernoff bound (Equation (8)). Note that for a fixed λ the right hand side of (8) is a non-decreasing function of t. Further note that $|\{(i,j) \in T : i > j\}|, |\{(i,j) \in T : i < j\}| \leq |T| < 2\alpha n^2$. Therefore,

$$\Pr\left[\left|\sum_{(i,j)\in T: i>j}(\tilde{A}_{i,j} - p)\right| \geq \beta\right] \leq 2\exp\left(-\min\left(\frac{\beta^2}{6\alpha n^2 p}, \frac{\beta}{3}\right)\right) \leq 2\exp\left(-20\alpha n \ln\frac{e}{\alpha}\right).$$

Similarly,

$$\Pr\left[\left|\sum_{(i,j)\in T: i<j}(\tilde{A}_{i,j} - p)\right| \geq \beta\right] \leq 2\exp\left(-20\alpha n \ln\frac{e}{\alpha}\right).$$

Combining the two bounds completes the proof of Lemma 36.

We finally prove Lemma 37. The number of such sets can be upper bounded by $4\cdot\left(\sum_{j=0}^{\lfloor \alpha n \rfloor}\binom{n}{j}\right)^2$, where

$$\sum_{j=0}^{\lfloor \alpha n \rfloor}\binom{n}{j} \leq (\alpha n + 1)\cdot\binom{n}{\lfloor \alpha n \rfloor} \leq (\alpha n + 1)\cdot\left(\frac{e}{\alpha}\right)^{\alpha n} \leq e^{\alpha n \ln\left(\frac{e}{\alpha}\right) + \ln(\alpha n + 1)} \leq e^{\alpha n \ln\left(\frac{e}{\alpha}\right) + \alpha n} \leq e^{2\alpha n \ln(e/\alpha)}.$$

Appendix G. Lower bound proofs

Proof of Lemma 20 We prove this lemma by converting a γ-corrupted graph from $DG(n, p)$ to a γ-corrupted graph from $G(n, p)$. Then one can run the algorithm for the undirected setting to obtain an estimate of p, which implies the same error guarantees for the directed instance.

Suppose there exists a random directed graph $DG \sim DG(n, p)$ which is γ-corrupted by an adversary. Assume there exists some lexicographic ordering of the nodes (e.g., they are numbered from 1 to n). We define a corresponding undirected graph G as follows: let there be an edge between nodes i and j in G if there exists an edge from i to j in DG and $i < j$. Sans corruptions, this converts $DG(n, p)$ into $G(n, p)$ since the edges are still independent and the probability of each edge existing is p. Furthermore, when at most γn nodes in the original directed graph are modified, at most γn nodes are changed in the corresponding undirected graph. ■

Proof of Theorem 21 Our γ-oblivious adversary for the directed graph model works as follows. The adversary picks a set B of size $Bin(n, 0.15\gamma)$ to corrupt, by independently picking each node in $[n]$ with probability 0.15γ. Note that it is possible that the size of the set of corrupted nodes B may exceed γn with small probability. We will address this issue later.

The adversary will corrupt outgoing edges of the nodes in B. The adversary's strategy to corrupt the neighborhood of node $i \in B$ is as follows. They first choose node i's new out-degree $\deg(i)$ independently from some distribution P over $\{0, \ldots, n-1\}$. Then, they select an independent random subset S_i of nodes $[n] \setminus \{i\}$ of size $\deg(i)$. Finally, they introduce the directed edge (i, j) for each $j \in S_i$, and remove the directed edge (i, j) for each $j \notin S_i$. The distribution of the degree of corrupted nodes, P, depends on the parameter p of the Erdős-Rényi graph and will be specified later.

By this construction, all graphs with a given outgoing degree distribution $d_1, d_2, \ldots, d_n$ have the same probability, and they form a sufficient statistic for estimating p. The out-degree of any uncorrupted node is distributed as $Bin(n-1, p)$ and the out-degree of any corrupted node has distribution P. Since each node is corrupted with probability 0.15γ, the out-degree of each node is an i.i.d. sample from the mixture distribution $(1 - 0.15\gamma) \cdot Bin(n-1, p) + 0.15\gamma \cdot P$.

Next, we show that for any $p_1 \leq 1/2$ and $p_2 = p_1 + 0.1 \max\left(\gamma\sqrt{p/n}, 0.1\gamma/n\right)$ there exist distributions P_1 and P_2 such that

$$(1 - 0.15\gamma) \cdot Bin(n-1, p_1) + 0.15\gamma \cdot P_1 = (1 - 0.15\gamma) \cdot Bin(n-1, p_2) + 0.15\gamma \cdot P_2. \quad (26)$$

This will imply that, with the aforementioned adversary, any estimator that distinguishes between the two cases will be correct with probability at most $1/2$. At this point, we account for the probability that the adversary selects a set B of size $> \gamma n$, which is not allowed according to the corruption model. By Markov's inequality, this occurs with probability at most 0.15. Therefore, even counting such violations as a success at distinguishing the two cases, it still succeeds with probability at most $0.5 + 0.15 = 0.65$.

To prove the existence of P_1 and P_2 satisfying (26) we use the following folklore fact: given any two distributions D_1 and D_2 and $\varepsilon > 0$, if $d_{\mathrm{TV}}(D_1, D_2) \leq \varepsilon$, then there exist distributions Q_1 and Q_2 such that $(1-\varepsilon)D_1 + \varepsilon Q_1 = (1-\varepsilon)D_2 + \varepsilon Q_2$.

Hence, it suffices to show that

$$d_{\mathrm{TV}}\left(Bin(n-1, p_1), Bin(n-1, p_2)\right) \leq 0.15\gamma. \quad (27)$$

The total variation distance between two binomials can be bounded as (Roos, 2001), (Adell and Jodrá, 2006, Eq (2.16)).

$$d_{\mathrm{TV}}\left(Bin(n',p), Bin(n',p+x)\right) \leq \sqrt{\frac{e}{2}}\frac{\tau(x)}{(1-\tau(x))^2}, \tag{28}$$

where $\tau(x) = x\cdot\sqrt{\frac{n'+2}{2p(1-p)}}$. We also use the trivial upper bound $d_{\mathrm{TV}}\left(Bin(n',p), Bin(n',p+x)\right) \leq n'x$.

For the case when $\gamma\sqrt{p/n} \geq 0.1\gamma/n$, applying the first bound for $x = 0.1\gamma\sqrt{p/n}$ and $n' = n-1$ we get

$$\tau(x) = 0.1\gamma\sqrt{\frac{p}{n}}\sqrt{\frac{n+1}{2p(1-p)}} \leq 0.1\cdot 1.1\gamma = 0.11\gamma.$$

For this case, using (28) gives

$$d_{\mathrm{TV}}\left((Bin(n-1,p_1), Bin(n-1,p_2)\right) \leq 0.15\gamma.$$

For the other case when $\gamma\sqrt{p/n} < 0.1\gamma/n$, applying the trivial bound gives

$$d_{\mathrm{TV}}\left((Bin(n-1,p_1), Bin(n-1,p_2)\right) \leq 0.1\gamma(n-1)/n < 0.1\gamma.$$

This proves (27) and shows the existence of P_1 and P_2, which completes the proof of the first two terms in Δ_p.

Finally, we show that the third term in the max in Δ_p holds even when there is no corruption. To show this we first note that in absence of corruption the sufficient statistics for estimating p is the total number of edges in the directed graph, which has a distribution $Bin((n-1)^2, p)$. Then to show that for $p \leq 0.5$ no algorithm can distinguish between between $DG(n,p)$ and $DG(n, p+0.1\sqrt{p}/n)$ with probability ≥ 0.6 it suffices to show that $d_{\mathrm{TV}}(Bin((n-1)^2,p), Bin((n-1)^2, p+0.1\sqrt{p}/n)) < 0.2$, which can be verified using (28) for $x = 0.1\sqrt{p}/n$, $n' = (n-1)^2$, and any $p < 1/2$. ■