Proceedings of Machine Learning Research vol 336:1–48, 2026 39th Annual Conference on Learning Theory

Spectral Recovery of a Planted Triangle-Dense Subgraph

Sam van der Poel SAMVANDERPOEL@GATECH.EDU
Cheng Mao CHENG.MAO@MATH.GATECH.EDU
Benjamin M^c Kenna MCKENNA@MATH.GATECH.EDU
School of Mathematics, Georgia Institute of Technology, Atlanta, GA

Editors: Steve Hanneke and Tor Lattimore

Abstract

Given a simple graph on n vertices and a parameter k, the triangle-densest-k-subgraph problem is known to be computationally hard in the worst case. To circumvent the computational hardness, we study an average-case model where a triangle-dense subgraph on k vertices is planted in an Erdős–Rényi random graph on n vertices. For the recovery of the planted subgraph, we propose a simple spectral algorithm and a semidefinite program, both of which use a graph matrix whose entries are local signed triangle counts. Theoretical guarantees for these algorithms are established through spectral analysis of the graph matrix. Finally, we provide evidence showing a statistical-to-computational gap analogous to that for the planted clique problem. The computational threshold in terms of the subgraph size k is at least $\sqrt{n}$ in the framework of low-degree polynomial algorithms, while the information-theoretic threshold is at most logarithmic in n.

Keywords: Planted models, random graphs, triangle-dense subgraphs, spectral methods, graph matrices

1. Introduction

Planted models have achieved significant success in the study of average-case computational hardness for inference on random graphs. A canonical example is the densest k-subgraph problem: although it is NP-hard in the worst case, its planted variants on random graphs—most notably the iconic planted clique problem (Jerrum, 1992; Alon et al., 1998) and the planted dense subgraph problem (Feige and Seltser, 1997; Bhaskara et al., 2010)—have generated a substantial body of work on computational thresholds and the development of efficient algorithms.

In this work, we propose a planted model for a triangle-dense subgraph in a random graph and study the recovery of the planted subgraph. Our first motivation comes from the *triangle-densest k-subgraph problem*, which refers to finding a subset of k vertices that induces the maximum number of triangles in a graph on n vertices. This problem, as well as the approximation of the optimal value, is known to be computationally hard in the worst case (Konar and Sidiropoulos, 2022). Our work proposes an average-case model and efficient algorithms for this problem, analogous to the study of the planted dense subgraph model for the densest k-subgraph problem.

The model we propose plants a k-subgraph in an ambient Erdős–Rényi graph on n vertices, where the subgraph is a *random geometric graph* (RGG) generated from a linear inner product kernel. RGGs are latent space models in which connections between nodes depend on the geometric proximity of corresponding node features in a latent space. The geometric structure of RGGs makes them useful in applications including social networks (Hoff et al., 2002). Further, one of the hallmark properties of RGGs is their elevated *triangle density*, which has been described as one of the defining characteristics of social networks (Easley and Kleinberg, 2010; Sala et al., 2010; Ugander et al., 2013; Gupta et al., 2014). This phenomenon has also played a significant role in recent theoretical studies of RGGs, as the *signed triangle count* has been identified as the optimal statistic for distinguishing an RGG from an Erdős–Rényi graph in various settings (Bubeck et al., 2016; Liu et al., 2022; Liu and Rácz, 2023; Bangachev and Bresler, 2025; Mao et al., 2026). From this perspective, our model is also an average-case model for planted geometry in a random graph.

1.1. Main contributions

We now briefly introduce our model and results. Suppose that we plant a subset $\mathcal{S}$ of size k in the vertex set $[n] := \{1, \dots, n\}$. The observed graph, identified with its adjacency matrix $\mathbf{A}$, follows the Erdős–Rényi model $G(n, p)$ for $p \in (0, 1/2]$, except that the planted subgraph $\mathbf{A}[\mathcal{S}]$ induced by $\mathcal{S}$ has independent edges

$$\mathbf{A}[\mathcal{S}]_{ij} \sim \mathrm{Ber}(p(X_i^\top X_j + 1)) \tag{1}$$

where X_i's are i.i.d. uniformly random vectors on the unit sphere $\mathbb{S}^{d-1}$ in $\mathbb{R}^d$ for some $d \geq 3$. It is easily seen that the planted subgraph has edge density p, same as $G(n, p)$, but triangle density $p^3(1 + 1/d^2)$, higher than p^3 in the ambient Erdős–Rényi graph (see Lemma 2.3). Our goal is to recover the planted subgraph, i.e., to estimate $\mathcal{S}$, by leveraging the elevated triangle density.

Towards this end, we consider the matrix

$$\mathbf{M} := \bar{\mathbf{A}}^2 \circ \bar{\mathbf{A}}, \tag{2}$$

where $\circ$ denotes the Hadamard product and $\bar{\mathbf{A}}$ denotes the centered adjacency matrix defined by $\bar{\mathbf{A}}_{ij} = \mathbf{A}_{ij} - p$ for $i \neq j$ and $\bar{\mathbf{A}}_{ii} = 0$. The reason for considering this matrix is that each entry $\mathbf{M}_{ij} = \sum_{\ell=1}^{n} \bar{\mathbf{A}}_{ij}\bar{\mathbf{A}}_{i\ell}\bar{\mathbf{A}}_{\ell j}$ is the *local, signed triangle count* at (i, j). As a result, $\mathbf{M}$ has an elevated mean over entries in the planted part $\mathcal{S}$ (see Lemma 2.4 for a precise statement). Given this crucial property, we can then recover $\mathcal{S}$ from $\mathbf{M}$ by running either (i) a spectral method that computes the leading eigenvector of $\mathbf{M}$ and takes the indices of its k largest entries, or (ii) a semidefinite program with $\mathbf{M}$ as the input. Our main results show that the above two efficient methods achieve almost exact recovery and exact recovery of $\mathcal{S}$, respectively, under the primary condition $kp^{3/4} \geq Cn^{1/2}d$ for a sufficiently large constant $C > 0$.

When p and d are both constants, the above condition reduces to $k \gtrsim \sqrt{n}$, which is the same as the celebrated $\sqrt{n}$ computational threshold for the planted clique problem. This motivates us to ask whether an analogous statistical-to-computational gap exists for the planted triangle-dense subgraph model. We provide evidence supporting this gap: if $k = o(\sqrt{n})$, then no low-degree polynomial can estimate $\mathcal{S}$ in the sense of Schramm and Wein (2022); however, if k is at least logarithmic in n, then an exponential-time algorithm finds the planted triangle-dense subgraph.

Finally, the main technical challenge in proving theoretical guarantees for our algorithms lies in the spectral analysis of the matrix $\mathbf{M}$ defined in (2). Such matrices are known as *graph matrices* (Medarametla and Potechin, 2016) or *motif adjacency matrices* (Benson et al., 2016), which are special cases of *polynomial random matrices* (Rajendran and Tulsiani, 2023) or *matrix chaoses* (Bandeira et al., 2025). Standard matrix concentration inequalities do not apply directly to such graph matrices because their entries are polynomials of the entries of the adjacency matrix. In addition, the planted subgraph in our case is defined through latent points on a sphere as in (1), which further complicates the analysis. To tackle these challenges, our proofs combine decoupling inequalities (de la Peña and Giné, 2012) with iterated matrix concentration inequalities (similar to Bandeira et al. (2025)) for the noise part, and use tools from spherical harmonics (Dai and Xu, 2013) for the signal part.

1.2. Related work

Planted dense subgraphs: There is an extensive literature on community detection (Fortunato, 2010; Arias-Castro and Verzelen, 2014; Abbe et al., 2016; Mossel et al., 2015) and the dense subgraph problem (Feige and Seltser, 1997; Bhaskara et al., 2010; Ames, 2015). Particularly, the computational thresholds for both detection and recovery have received considerable attention in recent years (Hajek et al., 2015; Bresler and Jiang, 2023; Schramm and Wein, 2022; Dhawan et al., 2025). Moving beyond the assumption of inhomogeneous edge density in the observed graph, one may ask how to recover a planted subgraph that has the same expected edge density as the ambient graph, but a different expected density of another template graph. We initiate this study by considering the case of elevated triangle density. Both the spectral method and the semidefinite program we use are analogous to their counterparts for planted dense subgraph recovery (Hajek et al., 2016a), with the centered adjacency matrix replaced by (2).

Triangle-densest k-subgraph problem: The triangle-densest k-subgraph problem refers to finding a subset of k vertices in a given graph that induces the maximum number of triangles. Konar and Sidiropoulos (2022) proved that this problem is NP-hard and cannot be approximated efficiently in the worst case. This is also related to a few other optimization problems concerning triangle-densest subgraphs in the literature (Tsourakakis, 2015; Samusevich et al., 2016; Wang et al., 2020). It is worth noting that the semidefinite program we use for the triangle-densest k-subgraph problem is quite different from the convex relaxation proposed by Konar and Sidiropoulos (2022) and is more in line with the programs in Goemans and Williamson (1995); Abbe et al. (2016); Hajek et al. (2016a) that are often used for planted models.

Random graphs with elevated triangle density: Triangle density is a central concept in social network analysis. For example, the well-known *clustering coefficient* is defined to be the triangle density in the neighborhood of a vertex (Watts and Strogatz, 1998), and the related concept of the number of edges in an *egonet* also concerns the triangle count in a neighborhood (Bhadra and Sengupta, 2018). To model a subgraph with an elevated triangle density, we use a random geometric graph, which belongs to the general class of latent space models that has long been studied in network analysis (Hoff et al., 2002). The node feature vectors in a latent space give rise to a clustering effect, and thus to a higher triangle density.

The RGG model we consider is particularly motivated by a recent line of work on the detection of latent geometry in random graphs using the signed triangle count (Bubeck et al., 2016; Liu et al., 2022; Bangachev and Bresler, 2025; Mao et al., 2026). The linear kernel we assume is a special case of general smooth kernels considered by (Liu and Rácz, 2023). As noted above, the excess triangle density of the planted subgraph compared to the ambient Erdős–Rényi graph is p^3/d^2, so the dimension d is a modeling parameter that determines the triangle density.

There are other models for random graphs with more triangles. One direct way is to plant extra triangles in an Erdős–Rényi graph (Bresler et al., 2023). It is also possible to consider Erdős–Rényi graphs conditional on having a higher triangle density (Chatterjee and Varadhan, 2011). Moreover, a random graph with desired edge and triangle densities can be generated from the exponential random graph model (Chatterjee and Diaconis, 2013) or from a suitable choice of a graphon (Kenyon et al., 2017). It is interesting and challenging to develop a general theory for triangle-dense subgraphs that applies to a range of models.

Graph matrices and matrix chaos: An important part of our work is a spectral analysis of the graph matrix $\mathbf{M}$ defined in (2) and is therefore related to the literature on graph matrices and matrix chaos. Also known as motif adjacency matrices, they have been used in higher-order spectral methods in network analysis (Benson et al., 2016; Paul et al., 2023). Our work extends this line of research. One step in our proofs, Theorem B.1, is bounding the spectral norm of $\mathbf{Z}^2 \circ \mathbf{Z}$ where $\mathbf{Z}$ is a Wigner matrix. The spectral norm of such a graph matrix has been studied by Medarametla and Potechin (2016); Ahn et al. (2016); Rajendran and Tulsiani (2023), but their bounds are not sufficiently sharp to yield the desired condition without extra logarithmic factors. More recently, a remarkable work by Bandeira et al. (2025) provides a general way to obtain sharp bounds on the expected norms of matrix chaos. Our proof follows a similar recipe, which consists of matrix decoupling and iterated concentration inequalities, and establishes a high-probability bound.

2. Main results

2.1. Planted triangle-dense subgraph

For a positive integer n and an edge density parameter $p = p_n \in (0,1)$, let $G(n,p)$ denote the Erdős–Rényi random graph model. For a graph $\mathbf{A}$ on n vertices, we identify $\mathbf{A}$ with its adjacency matrix in $\mathbb{R}^{n\times n}$. We assume the graph is simple, so that $\mathbf{A}_{ii} = 0$ for $i \in [n]$. For a subset of vertices $\mathcal{S} \subset [n]$, we let $\mathbf{A}[\mathcal{S}]$ denote the subgraph of $\mathbf{A}$ induced by $\mathcal{S}$. Let us first define a model for planting a fixed subgraph in an Erdős–Rényi graph.

Definition 2.1 (Planting a subgraph in an Erdős–Rényi graph) *For $k = k_n \in [n]$, fix a graph $\mathbf{H}$ on k vertices. Let $\mathcal{S}$ be a uniformly random subset of $[n]$ of size $|\mathcal{S}| = k$. Given $\mathbf{A}_0 \sim G(n,p)$, we replace $\mathbf{A}_0[\mathcal{S}]$ by $\mathbf{H}$ by assigning the vertices of $\mathbf{H}$ uniformly randomly to the vertices in $\mathcal{S}$. We denote the resulting graph by $\mathbf{A}$ and write $\mathbf{A} \sim G(n,p;\mathbf{H})$.*

In this work, we focus on a subgraph $\mathbf{H}$ whose expected edge density is also p, matching that of the ambient Erdős–Rényi graph, and whose triangle density is higher than the Erdős–Rényi graph. To this end, we model $\mathbf{H}$ as a random geometric graph generated from a linear kernel of inner products, or simply a random inner product graph, defined as follows.

Definition 2.2 (Random inner product graph) *For positive integers $k, d \geq 3$, let $X_1, \ldots, X_k$ be i.i.d. random vectors from the uniform distribution on $\mathbb{S}^{d-1}$, the unit sphere in $\mathbb{R}^d$. For $p \in (0, 1/2]$, suppose that the random graph $\mathbf{H}$ on k vertices has independent edges $\mathbf{H}_{ij} \sim \mathrm{Ber}(p(X_i^\top X_j + 1))$ up to symmetry $\mathbf{H}_{ij} = \mathbf{H}_{ji}$. We also zero the diagonal, $\mathbf{H}_{ii} = 0$. We write $\mathbf{H} \sim G_{\mathrm{IP}}(k,p,d)$.*

One can potentially consider a different kernel $\kappa(X_i^\top X_j)$ that takes values in $[0,1]$ in the above definition of the random inner product graph (see, e.g., Liu and Rácz (2023)), but we focus on the linear kernel $\kappa(t) = p(t+1)$ for simplicity.

With the above definitions, our observation is a graph $\mathbf{A} \sim G(n,p;\mathbf{H})$, where here $\mathbf{H} \sim G_{\mathrm{IP}}(k,p,d)$. In other words, we observe a random graph $\mathbf{A}$ on n vertices which contains a subgraph

$\mathbf{H}$ on k vertices planted in an unknown location $\mathcal{S}$. Note that the planted subgraph $\mathbf{H}$ has edge density p (averaged over the randomness in the X_i's). To see that it has a higher triangle density, let us note the following.

Lemma 2.3 *For distinct $i, j, \ell \in [n]$, we have $\mathbb{E}[\mathbf{H}_{ij}\mathbf{H}_{j\ell}\mathbf{H}_{\ell i}] = p^3(1 + 1/d^2)$.*

Proof Since X_i, X_j, X_ℓ are i.i.d. uniform over $\mathbb{S}^{d-1}$, we have $\mathbb{E}[X_i X_i^\top] = \mathbf{I}_d/d$ and $\mathbb{E}[X_i^\top X_j] = 0$, and similarly for other indices. Thus all the cross-terms vanish in the following expansion:

$$\begin{aligned}
\mathbb{E}[\mathbf{H}_{ij}\mathbf{H}_{j\ell}\mathbf{H}_{\ell i}] &= p^3\mathbb{E}[(X_i^\top X_j + 1)(X_j^\top X_\ell + 1)(X_\ell^\top X_i + 1)] \\
&= p^3\mathbb{E}[X_i^\top X_j X_j^\top X_\ell X_\ell^\top X_i] + p^3 \\
&= p^3\mathbb{E}\Big[X_i^\top \Big(\frac{1}{d}\mathbf{I}_d\Big)\Big(\frac{1}{d}\mathbf{I}_d\Big) X_i\Big] + p^3 \\
&= p^3/d^2 + p^3\,.
\end{aligned}$$

■

Since $\mathbf{H}_{ij}\mathbf{H}_{j\ell}\mathbf{H}_{\ell i}$ indicates the presence of the triangle with vertices i, j, ℓ in $\mathbf{H}$, we see that $\mathbf{H}$ indeed has a higher triangle density than the rest of $\mathbf{A}$ (which has triangle density p^3).

2.2. Spectral method

Our spectral method for recovering the planted subgraph, i.e., estimating $\mathcal{S}$, is as follows. Let $\mathbf{M}$ be the matrix defined by (2). Let $\hat{u} = \hat{u}(\mathbf{M})$ denote the leading eigenvector of $\mathbf{M}$, i.e., the eigenvector associated with the largest eigenvalue $\hat{\lambda}$ of $\mathbf{M}$. By convention $\hat{u}$ is a unit vector chosen with arbitrary sign; compare to (4) below. We estimate the planted location $\mathcal{S}$ by the set $\hat{\mathcal{S}}$ of indices $i \in [n]$ such that $|\hat{u}_i|$ is among the largest k entries of $\hat{u}$ in absolute value (with ties broken arbitrarily).

Before presenting our theoretical guarantees for $\hat{u}$ and $\hat{\mathcal{S}}$, let us first explain the intuition behind the spectral method. To see that the matrix $\mathbf{M}$ captures the the triangle density of the graph, it is helpful to consider the non-centered version $\mathbf{A}^2 \circ \mathbf{A}$ whose entry

$$(\mathbf{A}^2 \circ \mathbf{A})_{ij} = \sum_{\ell=1}^{n} \mathbf{A}_{i\ell}\mathbf{A}_{\ell j}\mathbf{A}_{ij}$$

is precisely the number of triangles in $\mathbf{A}$ that contain the edge (i, j). Therefore, the entries of $\mathbf{A}^2 \circ \mathbf{A}$ are local triangle counts, while the entries of $\mathbf{M} = \bar{\mathbf{A}}^2 \circ \bar{\mathbf{A}}$ are local signed triangle counts, where the centering has proved to be effective for variance reduction in various settings (Bubeck et al., 2016). More formally, it is not hard to see the following.

Lemma 2.4 *Let $\mathbf{M}$ be as defined above. Then we have*

$$\mathbb{E}[\mathbf{M}] = \frac{p^3(k-2)}{d^2}(\mathbf{1}_{\mathcal{S}}\mathbf{1}_{\mathcal{S}}^\top - \mathbf{I}_{\mathcal{S}}),$$

where $\mathbf{1}_{\mathcal{S}} \in \mathbb{R}^n$ is the indicator vector of the set $\mathcal{S} \subset [n]$, and $\mathbf{I}_{\mathcal{S}} \in \mathbb{R}^{n\times n}$ denotes the matrix with a $k \times k$ identity principal minor indexed by $\mathcal{S}$ and zeros elsewhere.

Proof The diagonal of $\mathbf{M}$ is zero by definition. Consider $i \neq j$. If i or j is not in $\mathcal{S}$, then it is easily seen that $\mathbb{E}[\bar{\mathbf{A}}_{ij}] = 0$, $\mathbb{E}[\bar{\mathbf{A}}_{i\ell}\bar{\mathbf{A}}_{\ell j}\bar{\mathbf{A}}_{ij}] = 0$ for all $\ell \in [n]$, and so $\mathbb{E}[\mathbf{M}_{ij}] = \mathbb{E}[(\bar{\mathbf{A}}^2 \circ \bar{\mathbf{A}})_{ij}] = 0$. If $i, j \in \mathcal{S}$, then

$$\mathbb{E}[\mathbf{M}_{ij}] = \sum_{\ell=1}^{n} \mathbb{E}[\bar{\mathbf{A}}_{i\ell}\bar{\mathbf{A}}_{\ell j}\bar{\mathbf{A}}_{ij}] = \sum_{\ell \in \mathcal{S}\setminus\{i,j\}} p^3 \mathbb{E}[X_i^\top X_j X_j^\top X_\ell X_\ell^\top X_i] = (k-2)p^3/d^2$$

where the last step has been shown in the proof of Lemma 2.3. ■

Thus the spectrum of $\mathbb{E}[\mathbf{M}]$ is explicit: it consists of an eigenvalue of $\frac{p^3(k-2)}{d^2}(|\mathcal{S}|-1) = \frac{p^3(k-2)(k-1)}{d^2}$ with corresponding eigenvector $\mathbf{1}_{\mathcal{S}}$, an eigenvalue of $-\frac{p^3(k-2)}{d^2}$ with multiplicity $|\mathcal{S}| - 1 = k-1$, and a kernel of dimension $n - |\mathcal{S}| = n-k$. If $k \to \infty$, then the first eigenvalue is the leading one, meaning that $\mathbb{E}[\mathbf{M}]$ has leading eigenvector $\mathbf{1}_{\mathcal{S}}$. In other words, one can recover $\mathcal{S}$ from the top eigenvector of $\mathbb{E}[\mathbf{M}]$. If $\mathbf{M}$ is close to its expectation, then one might hope that the top eigenvector of $\mathbf{M}$ would give approximate recovery of $\mathcal{S}$. Our first main theorem, below, shows that this heuristic is correct, even though its proof (deferred to Section 3) shows that $\mathbf{M}$ is *not* just "$\mathbb{E}[\mathbf{M}]$ plus small noise."

Theorem 2.5 *Let $\mathbf{A} \sim G(n, p; \mathbf{H})$ be given, where $\mathbf{H} \sim G_{\mathrm{IP}}(k, p, d)$. Recall that $\hat{u}$ is the leading eigenvector of $\mathbf{M}$ defined by* (2). *For any $\epsilon > 0$, there is $C > 0$ depending only on ε with the following property. If*

$$kp^{3/4} \geq Cn^{1/2}d \tag{3}$$

and $np^{3/2} \geq (\log n)^{3/2}$, then the following holds with probability at least $1 - n^{-10}$: We have

$$\min_{\zeta \in \{\pm 1\}} \left\| \zeta \hat{u} - \frac{1}{\sqrt{k}} \mathbf{1}_{\mathcal{S}} \right\| \leq \epsilon, \tag{4}$$

and for any set[1] $\hat{\mathcal{S}}$ consisting of exactly k indices $i \in [n]$ such that $|\hat{u}_i|$ is among the largest k entries of $\hat{u}$ in absolute value, we have

$$|\hat{\mathcal{S}} \triangle \mathcal{S}| \leq 8\varepsilon^2 k. \tag{5}$$

We note that the secondary assumption $np^{3/2} \geq (\log n)^{3/2}$ is mild. Moreover, it is, up to a logarithmic factor, subsumed by the main condition $kp^{3/4} \geq Cn^{1/2}d$ in (3). Indeed, if we have $n \geq k \geq Cn^{1/2}d/p^{3/4}$, then $np^{3/2} \geq Cd^2 \geq C$.

To provide heuristics[2] for the main condition (3), suppose that vertex 1 is planted, i.e., $1 \in \mathcal{S}$, and consider the statistic

$$T := \sum_{j,\ell=1}^{n} \bar{\mathbf{A}}_{1j}\bar{\mathbf{A}}_{1\ell}\bar{\mathbf{A}}_{\ell j} = \sum_{j=1}^{n} \mathbf{M}_{1j},$$

1. There can be more than one such set, because of possible ties when $\hat{u}$ has multiple entries with the same magnitude.
2. This simple heuristic is analogous to the case of planted clique recovery: the celebrated $k = \sqrt{n}$ threshold can be seen by comparing the number of additional neighbors of a planted vertex (order k) to the standard deviation of the total number of neighbors of that vertex in $G(n, 1/2)$ (order $\sqrt{n}$). In our case, we consider (signed) triangles instead of neighbors (i.e., incident edges).

which is twice the signed count of triangles containing vertex 1. By Lemma 2.4, we have $\mathbb{E}[T] = \frac{p^3(k-2)}{d^2}(k-1) \asymp \frac{p^3k^2}{d^2}$. On the other hand, the expectation of T under $G(n,p)$ is zero, and the variance of T under $G(n,p)$ is

$$\begin{aligned}\mathrm{Var}_{G(n,p)}(T) &= \sum_{j,\ell=1}^{n}\sum_{j',\ell'=1}^{n}\mathrm{Cov}_{G(n,p)}(\bar{\mathbf{A}}_{1j}\bar{\mathbf{A}}_{1\ell}\bar{\mathbf{A}}_{\ell j}, \bar{\mathbf{A}}_{1j'}\bar{\mathbf{A}}_{1\ell'}\bar{\mathbf{A}}_{\ell' j'}) \\ &= 2\sum_{j,\ell=1}^{n}\mathbb{E}_{G(n,p)}[\bar{\mathbf{A}}_{1j}^2\bar{\mathbf{A}}_{1\ell}^2\bar{\mathbf{A}}_{\ell j}^2] \asymp n^2p^3.\end{aligned}$$

Comparing $\frac{p^3k^2}{d^2}$ to $\sqrt{n^2p^3}$ yields the threshold $kp^{3/4} = n^{1/2}d$.

This heuristic can potentially be turned into a rigorous analysis of the algorithm that computes the local signed triangle count at each vertex (which would then be similar to the egonet method in Bhadra and Sengupta (2018), except that the triangle count there is not signed). However, this method is unlikely to achieve the tight condition (3), because the analysis of such a local algorithm typically entails a union bound which incurs an additional logarithmic factor. On the other hand, the spectral method aggregates information globally and is thus able to achieve the threshold (3).

The above spectral method, theoretical guarantees, and heuristics are all analogous to those for the planted clique problem. However, the analysis of the spectral method is significantly more challenging in this case, because we need to analyze the matrix $\mathbf{M}$ whose entries are polynomials of random variables.

2.3. Semidefinite programming

Theorem 2.5 shows that the spectral method achieves $(1-\varepsilon)$-recovery of the planted part $\mathcal{S}$ for any small $\epsilon > 0$ if $kp^{3/4} \geq Cn^{1/2}d$ for a large enough C. One further direction is to design a rounding step for the spectral estimator to obtain exact recovery of $\mathcal{S}$, or to develop an iterative procedure that uses the spectral method as a module to achieve recovery under an even weaker condition, $kp^{3/4} \geq cn^{1/2}d$ for any small $c > 0$, analogous to Alon et al. (1998). Instead of pursuing these directions, we focus in this work on global, one-shot algorithms and show that a semidefinite program based on the matrix $\mathbf{M}$ defined in (2) achieves exact recovery, assuming the same main condition as the spectral method.

Let $\mathbf{J}$ be the $n \times n$ matrix of all ones. Consider the following semidefinite program:

$$\begin{aligned} \max \quad & \langle \mathbf{M}, \mathbf{X}\rangle \\ \text{s.t.} \quad & \mathbf{X} \succeq 0\,, \\ & \mathbf{X}_{ij} \geq 0\,, \qquad i,j = 1,\dots,n\,, \\ & \mathbf{X}_{ii} \leq 1\,, \qquad i = 1,\dots,n\,, \\ & \mathrm{Tr}(\mathbf{X}) = k\,, \\ & \langle \mathbf{J}, \mathbf{X}\rangle = k^2\,. \end{aligned} \tag{6}$$

This program, with $\mathbf{A}$ in place of $\mathbf{M}$, was proposed by Hajek et al. (2016a) to solve the planted dense subgraph problem. The following theorem shows that the same program with $\mathbf{M}$ as the data matrix succeeds at exact recovery of a planted triangle-dense subgraph.

Theorem 2.6 *Let $\mathbf{A} \sim G(n,p;\mathbf{H})$ be given, where $\mathbf{H} \sim G_{\mathrm{IP}}(k,p,d)$. Let $\mathbf{M}$ be defined by* (2)*. If* (3) *holds for a sufficiently large constant $C > 0$ and $np^{3/2} = \omega((d\log n)^2)$, then $\mathbf{1}_{\mathcal{S}}\mathbf{1}_{\mathcal{S}}^\top$ is the unique optimizer of* (6) *with probability at least $1 - n^{-10}$.*

Similar to the discussion after Theorem 2.5, the condition $np^{3/2} = \omega((d\log n)^2)$ is a mild assumption which does not involve the subgraph size k, and it is subsumed by the main condition $kp^{3/4} \geq Cn^{1/2}d$ in (3) up to a logarithmic factor.

2.4. Statistical-to-computational gap

The main condition (3) reduces to $k \gtrsim \sqrt{n}$ if p and d are both constants. Given the analogy to the $k = \sqrt{n}$ computational threshold for the planted clique problem, we provide evidence demonstrating that this is also the computational threshold for the recovery of a planted triangle-dense subgraph, and that there is a statistical-to-computational gap in our case, too.

For the computational hardness result, we follow Schramm and Wein (2022) to show a recovery lower bound for low-degree polynomial algorithms. Let $\theta := \mathbf{1}\{1 \in \mathcal{S}\}$ be the indicator for the event that the first vertex belongs to the planted subgraph. We will show that if $k = o(\sqrt{n})$, then one cannot estimate θ much better than with a trivial estimator in the following sense. Let $\mathbb{R}[\mathbf{A}]_{\leq D}$ be the set of real polynomials of degree at most D whose variables are the entries of $\mathbf{A}$. The *degree-D maximum correlation* is defined as

$$\mathsf{Corr}_{\leqslant D} := \sup_{f \in \mathbb{R}[\mathbf{A}]_{\leq D}} \frac{\mathbb{E}[f(\mathbf{A}) \cdot \theta]}{\sqrt{\mathbb{E}[f(\mathbf{A})^2]}} .$$

Theorem 2.7 *Let $\mathbf{A} \sim G(n,p;\,\mathbf{H})$ be given, where $\mathbf{H} \sim G_{\mathrm{IP}}(\tilde{k},p,d)$ and $\tilde{k}$ is a binomial random variable with parameters n and $r := k/n$. Let $p \in (0,1/2)$, $d \geq 2$, and $D = o\big(\big(\frac{\log n}{\log\log n}\big)^2\big)$. If $k \leq n^{1/2-\epsilon}$ for a constant $\epsilon > 0$ then*

$$\mathsf{Corr}_{\leqslant D} = (1 + o(1))r .$$

Theorem 2.7 asserts that no degree-D estimator achieves asymptotically larger correlation than the trivial estimator $f \equiv r$, since $\frac{\mathbb{E}[r\cdot\theta]}{\sqrt{\mathbb{E}[r^2]}} = r$.

To provide evidence for a statistical-to-computational gap, it remains to prove an information-theoretic upper bound. The result below shows that a logarithmically sized triangle-dense subgraph can be partially recovered with high probability by a computationally inefficient estimator. More precisely, if $k = \Theta(n)$ then let $\hat{\mathcal{S}}$ be a uniformly random k-subset of $[n]$, while if $k = o(n)$ let

$$\hat{\mathcal{S}} \in \operatorname{argmax}_{T \subseteq [n],\, |T| = k} \sum_{i,j,l \in T} \bar{\mathbf{A}}_{ij}\bar{\mathbf{A}}_{jl}\bar{\mathbf{A}}_{li} , \tag{7}$$

which maximizes the signed triangle count over all k-subgraphs of $\mathbf{A}$.

Theorem 2.8 *Let $\mathbf{A} \sim G(n,p;\mathbf{H})$ be given, where $\mathbf{H} \sim G_{\mathrm{IP}}(k,p,d)$. For all constant $p \in (0,1/2]$ and $d \geq 2$, there exist constants $C = C(p,d) > 0$ and $\epsilon = \epsilon(d) > 0$ such that if $k \geq C\log n$, then $|\mathcal{S} \cap \hat{\mathcal{S}}| \geq \epsilon k$ with probability $1 - o(1)$.*

Recall that $r = k/n$ and $\theta = \mathbf{1}\{1 \in \mathcal{S}\}$. If we define $\hat{\theta} := \mathbf{1}\{1 \in \hat{\mathcal{S}}\}$, then by symmetry,

$$\frac{\mathbb{E}[\hat{\theta} \cdot \theta]}{\sqrt{\mathbb{E}[\hat{\theta}^2]}} = \frac{\mathbb{P}\{1 \in \mathcal{S} \cap \hat{\mathcal{S}}\}}{\sqrt{\mathbb{P}\{1 \in \hat{\mathcal{S}}\}}} \geq \frac{(1-o(1))\epsilon k/n}{\sqrt{r}} = (1-o(1))\epsilon\sqrt{r}\,.$$

This is much larger than $\mathsf{Corr}_{\leqslant D} = (1+o(1))r$ in Theorem 2.7, so there is a statistical-to-computational gap for recovery in the regime $C \log n \leq k \leq n^{1/2-\epsilon}$.

The proofs of Theorems 2.7 and 2.8 are given in Appendices D and E respectively.

3. Analysis of the algorithms

3.1. Proof of Theorem 2.5

We now prove Theorem 2.5, our main theoretical result for the spectral method. Since the spectral method is equivariant with respect to a relabeling of the vertices of the observed graph, we may assume that $\mathcal{S} = [k]$ throughout the proof without loss of generality. We also introduce the following notation: For any matrix $\mathbf{B} \in \mathbb{R}^{k\times k}$, we let $\mathbf{B}^{\#} \in \mathbb{R}^{n\times n}$ be the matrix with $\mathbf{B}$ as its top-left $k \times k$ principal minor and zeros elsewhere. In the remainder, we will frequently use that, for any matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{k\times k}$, we have $\mathbf{A}^{\#}\mathbf{B}^{\#} = (\mathbf{AB})^{\#}$ and $\mathbf{A}^{\#} \circ \mathbf{B}^{\#} = (\mathbf{A} \circ \mathbf{B})^{\#}$.

Since the proof of Theorem 2.5 involves a long list of auxiliary matrices and lemmas, we first provide a proof outline before delving into the details. At a high level, Theorem 2.5 follows from a spectral perturbation analysis using, in particular, the classical Davis–Kahan theorem (originally Davis and Kahan (1970), or see, e.g., Theorem 4.5.5 in Vershynin (2018) for a textbook treatment).

Lemma 3.1 (Davis–Kahan) *For symmetric matrices $\mathbf{A}_1, \mathbf{A}_2 \in \mathbb{R}^{n\times n}$, let $\{(\lambda_i(\mathbf{A}_1), u_i(\mathbf{A}_1)) : i \in [n]\}$ be the set of eigenpairs of $\mathbf{A}_1$ where the eigenvalues are ordered decreasingly, i.e., $\lambda_1(\mathbf{A}_1) \geq \cdots \geq \lambda_n(\mathbf{A}_1)$, and every $u_i(\mathbf{A}_1)$ is a unit vector. Define the notation for $\mathbf{A}_2$ in the same way. Then we have*

$$\min_{\zeta\in\{\pm 1\}} \|\zeta u_1(\mathbf{A}_2) - u_1(\mathbf{A}_1)\| \leq \frac{2\sqrt{2}\,\|\mathbf{A}_2 - \mathbf{A}_1\|}{\lambda_1(\mathbf{A}_1) - \lambda_2(\mathbf{A}_1)},$$

where the norm on the left-hand side is the usual Euclidean distance on $\mathbb{R}^n$ and the norm on the right-hand side is the operator norm with respect to this distance.

To apply the Davis–Kahan theorem, we will take $\mathbf{A}_2 = \mathbf{M}$. It would be nice if we could take $\mathbf{A}_1$ to be $\mathbb{E}[\mathbf{M}]$, because $\mathbb{E}[\mathbf{M}]$ has top eigenvector $\mathbf{1}_{\mathcal{S}}/\sqrt{k}$, and a large top eigengap (its top eigenvalue is k times larger than the rest, and k is growing; see the discussion after Lemma 2.4). However, it turns out that $\|\mathbf{M} - \mathbb{E}[\mathbf{M}]\|$ is at the same order as this eigengap; thus $\mathbf{A}_1 = \mathbb{E}[\mathbf{M}]$ is not a good choice in Davis–Kahan. Instead, we will define a *random* matrix $\widetilde{\mathbf{M}}$ below, whose top eigenvector is deterministically $\mathbf{1}_{\mathcal{S}}/\sqrt{k}$. Unlike $\mathbb{E}[\mathbf{M}]$, the largest and second-largest eigenvalues of $\widetilde{\mathbf{M}}$ (which are deterministic) are of the same order as one another; however, we can get a better bound on $\|\mathbf{M} - \widetilde{\mathbf{M}}\|$, so that we will ultimately choose $\mathbf{A}_1 = \widetilde{\mathbf{M}}$ in Davis–Kahan.

Towards this end, the first step is to decompose the matrix $\mathbf{M}$ into a signal term plus a few noise terms. With $\mathbf{X} := [X_1 \cdots X_k]^\top \in \mathbb{R}^{k\times d}$, we let $\mathbf{K} \in \mathbb{R}^{k\times k}$ denote the matrix $\mathbf{XX}^\top$ with its diagonal replaced by zeros, i.e.,

$$\mathbf{K} := \mathbf{XX}^\top - \mathbf{I}_k. \tag{8}$$

Note that $p\mathbf{K}^{\#} = \mathbb{E}[\bar{\mathbf{A}} \mid \mathbf{X}]$. We further define

$$\mathbf{W} := \bar{\mathbf{A}} - p\mathbf{K}^{\#}.$$

Therefore,

$$\begin{aligned}\mathbf{M} = \bar{\mathbf{A}}^2 \circ \bar{\mathbf{A}} = p^3(\mathbf{K}^2 \circ \mathbf{K})^{\#} &+ p^2(\mathbf{K}^2)^{\#} \circ \mathbf{W} + p\mathbf{W}^2 \circ \mathbf{K}^{\#} + \mathbf{W}^2 \circ \mathbf{W} \\ &+ p^2(\mathbf{K}^{\#}\mathbf{W} + \mathbf{W}\mathbf{K}^{\#}) \circ \mathbf{K}^{\#} + p(\mathbf{K}^{\#}\mathbf{W} + \mathbf{W}\mathbf{K}^{\#}) \circ \mathbf{W}.\end{aligned} \tag{9}$$

In the above decomposition, $p^3(\mathbf{K}^2 \circ \mathbf{K})^{\#}$ can be seen as the signal term, while all the other terms are noise. We will bound the spectral norms of all the noise terms in Section A.2.

To understand the matrix $\mathbf{K}^2 \circ \mathbf{K}$, we will first show that it is close to $\frac{k}{d}(\mathbf{X}\mathbf{X}^\top) \circ (\mathbf{X}\mathbf{X}^\top)$, in Lemma A.1. Next, we use the theory of spherical harmonics to write

$$\frac{1}{k}(\mathbf{X}\mathbf{X}^\top) \circ (\mathbf{X}\mathbf{X}^\top) = \mathbf{Y}\boldsymbol{\Lambda}\mathbf{Y}^\top, \tag{10}$$

where the columns of $\mathbf{Y}$ are defined from spherical harmonics and $\boldsymbol{\Lambda}$ is a diagonal matrix consisting of the corresponding eigenvalues (ordered decreasingly), both to be specified in Lemma A.2. Finally, we orthonormalize the columns of $\mathbf{Y}$ to obtain a matrix $\mathbf{Q}$, and show that $\mathbf{Y}\boldsymbol{\Lambda}\mathbf{Y}^\top$ is close to $\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\top$, in Lemma A.3. We will see that by definition the first column of $\mathbf{Q}$ is exactly $\frac{1}{\sqrt{k}}\mathbf{1}_k$ where $\mathbf{1}_k$ is the all-ones vector in $\mathbb{R}^k$.

Putting it together, we eventually choose $\mathbf{A}_1 = \widetilde{\mathbf{M}}$ in Lemma 3.1, where

$$\widetilde{\mathbf{M}} := \frac{p^3k^2}{d}(\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\top)^{\#}, \tag{11}$$

which then implies that

$$\min_{\zeta\in\{\pm 1\}} \left\| \zeta u_1(\mathbf{M}) - \frac{1}{\sqrt{k}}\mathbf{1}_{\mathcal{S}} \right\| \le \frac{2\sqrt{2}\,\|\mathbf{M} - \widetilde{\mathbf{M}}\|}{\lambda_1(\widetilde{\mathbf{M}}) - \lambda_2(\widetilde{\mathbf{M}})}. \tag{12}$$

Moreover, by (9) and (10), we obtain

$$\begin{aligned}\left\|\mathbf{M} - \widetilde{\mathbf{M}}\right\| &= \Big\| \Big(p^3\mathbf{K}^2 \circ \mathbf{K} - \frac{p^3k}{d}(\mathbf{X}\mathbf{X}^\top) \circ (\mathbf{X}\mathbf{X}^\top) + \frac{p^3k^2}{d}\mathbf{Y}\boldsymbol{\Lambda}\mathbf{Y}^\top - \frac{p^3k^2}{d}\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\top\Big)^{\#} \\ &\quad + p^2(\mathbf{K}^2)^{\#} \circ \mathbf{W} + p\mathbf{W}^2 \circ \mathbf{K}^{\#} + \mathbf{W}^2 \circ \mathbf{W} \\ &\quad + p^2(\mathbf{K}^{\#}\mathbf{W} + \mathbf{W}\mathbf{K}^{\#}) \circ \mathbf{K}^{\#} + p(\mathbf{K}^{\#}\mathbf{W} + \mathbf{W}\mathbf{K}^{\#}) \circ \mathbf{W}\Big\| \\ &\le p^3\left\|\mathbf{K}^2 \circ \mathbf{K} - \frac{k}{d}(\mathbf{X}\mathbf{X}^\top) \circ (\mathbf{X}\mathbf{X}^\top)\right\| + \frac{p^3k^2}{d}\|\mathbf{Y}\boldsymbol{\Lambda}\mathbf{Y}^\top - \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\top\| \qquad (13) \\ &\quad + p^2\|(\mathbf{K}^2)^{\#} \circ \mathbf{W}\| + p\|\mathbf{W}^2 \circ \mathbf{K}^{\#}\| + \|\mathbf{W}^2 \circ \mathbf{W}\| \\ &\quad + p^2\|(\mathbf{K}^{\#}\mathbf{W} + \mathbf{W}\mathbf{K}^{\#}) \circ \mathbf{K}^{\#}\| + p\|(\mathbf{K}^{\#}\mathbf{W} + \mathbf{W}\mathbf{K}^{\#}) \circ \mathbf{W}\|.\end{aligned}$$

In Appendix A, we will bound each of the terms in (13) to obtain the following result.

Proposition 3.2 *The matrix $\widetilde{\mathbf{M}}$ defined in* (11) *has top two eigenvalues equal to $\frac{p^3k^2}{d^2}$ and $\frac{p^3k^2}{d^2(d/2+1)}$ respectively, and top eigenvector equal to $\frac{1}{\sqrt{k}}\mathbf{1}_{\mathcal{S}}$. Moreover, for any $\epsilon > 0$, there is $C > 0$ depending only on ε such that the following holds. If $np^{3/2} \geq (\log n)^{3/2}$ and $kp^{3/4} \geq Cn^{1/2}d$, then with probability at least $1 - n^{-10}$, we have*

$$\frac{d^2}{p^3k^2}\left\|\mathbf{M} - \widetilde{\mathbf{M}}\right\| \leq \epsilon.$$

Then (4) is an immediate consequence of (12) together with Proposition 3.2. To prove (5), let us focus on the case $\left\|\hat{u} - \frac{1}{\sqrt{k}}\mathbf{1}_{\mathcal{S}}\right\| \leq \varepsilon$, and the case $\zeta = -1$ is analogous. We have

$$\left\|\hat{u} - \frac{1}{\sqrt{k}}\mathbf{1}_{\mathcal{S}}\right\|^2 = \sum_{i\in\mathcal{S}}\left(\hat{u}_i - \frac{1}{\sqrt{k}}\right)^2 + \sum_{i\in[n]\setminus\mathcal{S}}\hat{u}_i^2 \leq \varepsilon^2.$$

If $|\hat{u}_i| \leq \frac{1}{2\sqrt{k}}$ for all $i \notin \hat{\mathcal{S}}$, then

$$\varepsilon^2 \geq \sum_{i\in\mathcal{S}\setminus\hat{\mathcal{S}}}\left(\hat{u}_i - \frac{1}{\sqrt{k}}\right)^2 \geq \frac{1}{4k}|\mathcal{S}\setminus\hat{\mathcal{S}}| = \frac{1}{8k}|\mathcal{S}\triangle\hat{\mathcal{S}}|.$$

If $|\hat{u}_i| > \frac{1}{2\sqrt{k}}$ for some $i \notin \hat{\mathcal{S}}$, then we must have $|\hat{u}_i| > \frac{1}{2\sqrt{k}}$ for all $i \in \hat{\mathcal{S}}$. Therefore,

$$\varepsilon^2 \geq \sum_{i\in\hat{\mathcal{S}}\setminus\mathcal{S}}\hat{u}_i^2 \geq \frac{1}{4k}|\hat{\mathcal{S}}\setminus\mathcal{S}| = \frac{1}{8k}|\mathcal{S}\triangle\hat{\mathcal{S}}|.$$

In either case, we have $|\mathcal{S}\triangle\hat{\mathcal{S}}| \leq 8k\varepsilon^2$.

3.2. Proof of Theorem 2.6

We again assume that $\mathcal{S} = [k]$ without loss of generality, and let $\xi := \mathbf{1}_{[k]}$ for brevity. Let us start with a set of deterministic conditions that guarantee optimality.

Lemma 3.3 (Lemma 19 in Hajek et al. (2016b)) *Let $\mathbf{X}^* = \xi\xi^\top$. If there exist $n \times n$ matrices $\mathbf{D} = \mathrm{diag}(\mathbf{d})$ where $\mathbf{d}$ is an entrywise nonnegative vector, $\mathbf{S} \succeq 0$, and $\mathbf{B} \geq 0$, and $\eta, \lambda \in \mathbb{R}$ such that*

$$\begin{aligned}
&\mathbf{S} + \mathbf{M} + \mathbf{B} - \eta\mathbf{I}_n - \lambda\mathbf{J} - \mathbf{D} = 0\,,\\
&\langle\mathbf{S}, \mathbf{X}^*\rangle = 0\,,\\
&\langle\mathbf{B}, \mathbf{X}^*\rangle = 0\,,\\
&\langle\mathbf{D}, \mathbf{X}^* - \mathbf{I}_n\rangle = 0\,,\\
&\lambda_{n-1}(\mathbf{S}) > 0\,,
\end{aligned}$$

then $\mathbf{X}^*$ is the unique maximizer in (6).

We now prove Theorem 2.6. Let $\delta := n^{-10}$. It suffices to define matrices $\mathbf{D}$, $\mathbf{S}$, and $\mathbf{B}$, and real numbers $\eta, \lambda > 0$ that satisfy the conditions of Lemma 3.3 with probability at least $1-\delta$. We claim that the following assignments suffice:

$$\begin{aligned}
&\lambda := \tfrac{1}{k}\max_{k<i\le n}|(\mathbf{M}\xi)_i|\,, \\
&\eta := \min_{1\le i\le k}(\mathbf{M}\xi)_i - \lambda k\,, \\
&d_i := \mathbf{1}\{1\le i\le k\}\cdot\big((\mathbf{M}\xi)_i-\lambda k-\eta\big)\,, \quad \mathbf{D} := \operatorname{diag}(d_1,\dots,d_n)\,, \\
&b_i := \mathbf{1}\{i>k\}\cdot\big(\lambda-\tfrac{1}{k}(\mathbf{M}\xi)_i\big)\,, \qquad\qquad \mathbf{b} := (b_1,\dots,b_n)\,, \\
&\mathbf{B} := \mathbf{b}\xi^\top+\xi\mathbf{b}^\top\,, \\
&\mathbf{S} := \mathbf{D}-\mathbf{B}-\mathbf{M}+\eta\mathbf{I}_n+\lambda\mathbf{J}\,.
\end{aligned} \tag{14}$$

First notice that, by definition, we have that $d_i \ge 0$ for each i, and that $b_i \ge 0$ for each i, so that $\mathbf{B}$ is indeed entrywise nonnegative. The condition $\langle \mathbf{D}, \mathbf{X}^* - \mathbf{I}_n\rangle = 0$ is immediate since $\mathbf{d}$ is supported on its first k entries while $\mathbf{X}^*_{ii} = 1$ for all $i \in [k]$. Since $\mathbf{b}$ is orthogonal to ξ, we also have that $\langle \mathbf{B}, \mathbf{X}^*\rangle = \langle \xi, \mathbf{B}\xi\rangle = 0$, as well as $\mathbf{S}\xi = 0$; indeed, we have

$$\begin{aligned}
(\mathbf{S}\xi)_i &= d_i - (\mathbf{M}\xi)_i + \eta + \lambda k = 0, \quad i \le k, \\
(\mathbf{S}\xi)_i &= -k\lambda + (\mathbf{M}\xi)_i - (\mathbf{M}\xi)_i + k\lambda = 0, \quad i > k.
\end{aligned}$$

In particular, $\langle \mathbf{S}, \mathbf{X}^*\rangle = \langle \xi, \mathbf{S}\xi\rangle = 0$, and we conclude that, in order to show $\mathbf{S} \succeq 0$ with $\lambda_{n-1}(\mathbf{S}) > 0$, it suffices to show that $\langle v, \mathbf{S}v\rangle > 0$ uniformly over unit vectors v which are orthogonal to ξ. That is, it remains to prove that, with probability at least $1-\delta$,

$$\inf_{v\perp\xi,\|v\|=1}\langle v, \mathbf{S}v\rangle > 0.$$

For $\mathbf{S}$, we note that, if v is some unit vector which is orthogonal to ξ, we have

$$\langle v, \mathbf{B}v\rangle = \left\langle v, \mathbf{b}\xi^T v\right\rangle + \left\langle v, \xi\mathbf{b}^T v\right\rangle = 2\langle \xi, v\rangle\langle \mathbf{b}, v\rangle = 0,$$

so that

$$\langle v, \mathbf{S}v\rangle = \langle v, \mathbf{D}v\rangle - \langle v, \mathbf{B}v\rangle - \langle v, \mathbf{M}v\rangle + \eta + \lambda\langle v, 1\rangle^2 \ge \eta - \langle v, \mathbf{M}v\rangle\,,$$

where we used that $\mathbf{D}$ is positive semidefinite and that $\lambda \ge 0$. Thus it suffices to show that

$$\sup_{v\perp\xi,\|v\|=1}\langle v, \mathbf{M}v\rangle < \eta$$

with probability at least $1-\delta$. For $\widetilde{\mathbf{M}}$ defined in (11), Proposition 3.2 shows that the top eigenvector of $\widetilde{\mathbf{M}}$ is ξ, that the second-largest eigenvalue of $\widetilde{\mathbf{M}}$ is $\frac{2p^3k^2}{d^2(d+2)}$, and that $\|\mathbf{M}-\widetilde{\mathbf{M}}\| \le \frac{1}{6}\frac{p^3k^2}{d^2}$ with probability at least $1-\delta/2$. (Proposition 3.2 shows this event occurs with probability at least $1-\delta$, but the proof is easily modified to yield $1-\delta/2$.) Therefore,

$$\sup_{v\perp\xi,\|v\|=1}\langle v, \mathbf{M}v\rangle \le \sup_{v\perp\xi,\|v\|=1}\left\langle v, \widetilde{\mathbf{M}}v\right\rangle + \|\mathbf{M}-\widetilde{\mathbf{M}}\| \le \frac{2p^3k^2}{d^2(d+2)} + \frac{1}{6}\frac{p^3k^2}{d^2} \le \frac{5}{6}\frac{p^3k^2}{d^2}\,. \tag{15}$$

Applying Lemma C.1 with $\epsilon = 1/12$ yields that $\eta \ge \frac{11}{12}\frac{p^3k^2}{d^2}$ with probability at least $1-\delta/2$. Combining this with (15) completes the proof.

4. Future directions

Our work leaves more interesting problems open than it solves. Recall that we have provided evidence showing that the condition $k \geq C\sqrt{n}$ is optimal among low-degree polynomial algorithms for constant p and d, and provided a heuristic suggesting that the condition $kp^{3/4} \geq Cn^{1/2}d$ is expected for methods based on triangle counts. However, the full picture of statistical and computational thresholds for the recovery problem remains an interesting open question. Moreover, we have not studied the detection problem for planted triangle-dense subgraphs, which may have thresholds different from those for the recovery problem.

In addition, to model a subgraph with the same edge density but higher triangle density than that in the ambient Erdős–Rényi graph, the random geometric graph we use is just one natural model. As discussed in the introduction, other possibilities include a more general latent space model, an exponential family random graph model, or an Erdős–Rényi model conditional on having more triangles, to name a few. It is an interesting direction to study a potentially more general model for planted triangle-dense subgraphs, and we think our spectral method can be competitive in other settings, too, thanks to its simplicity.

Taking this one step further, one may study planted subgraphs characterized by a general homomorphism density. Is there a class of planted subgraph models where the subgraph has a higher H-density for a template graph H? How can we build a suitable graph matrix and analyze the corresponding spectral method? These questions pose substantial theoretical and algorithmic challenges.

Acknowledgments

S.v.d.P is supported by an Algorithms and Randomness Center Fellowship and an NSF Graduate Research Fellowship. C.M. is supported in part by NSF CAREER Award 2338062. B.M. is supported in part by NSF grant DMS-1760471.

References

Emmanuel Abbe, Afonso S. Bandeira, and Georgina Hall. Exact recovery in the stochastic block model. *IEEE Trans. Inform. Theory*, 62(1):471–487, 2016.

Kwangjun Ahn, Dhruv Medarametla, and Aaron Potechin. Graph matrices: Norm bounds and applications. *arXiv preprint arXiv:1604.03423*, 2016.

Noga Alon, Michael Krivelevich, and Benny Sudakov. Finding a large hidden clique in a random graph. In *Proceedings of the Eighth International Conference "Random Structures and Algorithms" (Poznan, 1997)*, volume 13, pages 457–466, 1998.

Brendan P. W. Ames. Guaranteed recovery of planted cliques and dense subgraphs by convex relaxation. *J. Optim. Theory Appl.*, 167(2):653–675, 2015.

Ery Arias-Castro and Nicolas Verzelen. Community detection in dense random networks. *Ann. Statist.*, 42(3):940–969, 2014.

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

Afonso S Bandeira, Kevin Lucca, Petar Nizić-Nikolac, and Ramon van Handel. Matrix chaos inequalities and chaos of combinatorial type. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pages 795–805, 2025.

Kiril Bangachev and Guy Bresler. Random algebraic graphs and their convergence to Erdős-Rényi. *Random Structures Algorithms*, 66(1):Paper No. e21276, 43, 2025.

Austin R Benson, David F Gleich, and Jure Leskovec. Higher-order organization of complex networks. *Science*, 353(6295):163–166, 2016.

Subhankar Bhadra and Srijan Sengupta. Detecting and localizing anomalous cliques in inhomogeneous networks using egonets. *arXiv e-prints*, pages arXiv–1807, 2018.

Aditya Bhaskara, Moses Charikar, Eden Chlamtac, Uriel Feige, and Aravindan Vijayaraghavan. Detecting high log-densities—an $O(n^{1/4})$ approximation for densest k-subgraph. In *STOC'10—Proceedings of the 2010 ACM International Symposium on Theory of Computing*, pages 201–210. ACM, New York, 2010.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: a nonasymptotic theory of independence*. Oxford University Press, 2013.

Guy Bresler and Tianze Jiang. Detection-recovery and detection-refutation gaps via reductions from planted clique. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 5850–5889. PMLR, 2023.

Guy Bresler, Chenghao Guo, and Yury Polyanskiy. Algorithmic decorrelation and planted clique in dependent random graphs: the case of extra triangles. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 2149–2158. IEEE, 2023.

Sébastien Bubeck, Jian Ding, Ronen Eldan, and Miklós Z Rácz. Testing for high-dimensional geometry in random graphs. *Random Structures & Algorithms*, 49(3):503–532, 2016.

Sourav Chatterjee and Persi Diaconis. Estimating and understanding exponential random graph models. *Ann. Statist.*, 41(5):2428–2461, 2013.

Sourav Chatterjee and S. R. S. Varadhan. The large deviation principle for the Erdős-Rényi random graph. *European J. Combin.*, 32(7):1000–1017, 2011.

Feng Dai and Yuan Xu. *Approximation theory and harmonic analysis on spheres and balls*. Springer, 2013.

Chandler Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.*, 7:1–46, 1970.

Yohann De Castro, Claire Lacour, and Thanh Mai Pham Ngoc. Adaptive estimation of nonparametric geometric graphs. *Math. Stat. Learn.*, 2(3-4):217–274, 2019.

Victor de la Peña and Evarist Giné. *Decoupling: from dependence to independence*. Springer Science & Business Media, 2012.

Victor H. de la Peña and S. J. Montgomery-Smith. Decoupling inequalities for the tail probabilities of multivariate U-statistics. *Ann. Probab.*, 23(2):806–816, 1995.

Abhishek Dhawan, Cheng Mao, and Alexander S Wein. Detection of dense subhypergraphs by low-degree polynomials. *Random Structures & Algorithms*, 66(1):e21279, 2025.

David Easley and Jon Kleinberg. *Networks, crowds, and markets: Reasoning about a highly connected world*, volume 1. Cambridge university press Cambridge, 2010.

Uriel Feige and Michael Seltser. *On the densest k-subgraph problem*. Weizmann Institute of Science. Department of Applied Mathematics and Computer Science, 1997.

Santo Fortunato. Community detection in graphs. *Phys. Rep.*, 486(3-5):75–174, 2010.

Evarist Giné, Rafał Latała, and Joel Zinn. Exponential and moment inequalities for U-statistics. In *High dimensional probability, II (Seattle, WA, 1999)*, volume 47 of *Progr. Probab.*, pages 13–38. Birkhäuser Boston, Boston, MA, 2000.

Michel X Goemans and David P Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42 (6):1115–1145, 1995.

Rishi Gupta, Tim Roughgarden, and C. Seshadhri. Decompositions of triangle-dense graphs. In *ITCS'14—Proceedings of the 2014 Conference on Innovations in Theoretical Computer Science*, pages 471–481. ACM, New York, 2014.

Bruce Hajek, Yihong Wu, and Jiaming Xu. Computational lower bounds for community detection on random graphs. In *Conference on Learning Theory*, pages 899–928. PMLR, 2015.

Bruce Hajek, Yihong Wu, and Jiaming Xu. Achieving exact cluster recovery threshold via semidefinite programming: extensions. *IEEE Trans. Inform. Theory*, 62(10):5918–5937, 2016a.

Bruce Hajek, Yihong Wu, and Jiaming Xu. Semidefinite programs for exact recovery of a hidden community. In Vitaly Feldman, Alexander Rakhlin, and Ohad Shamir, editors, *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 1051–1095, Columbia University, New York, New York, USA, 23–26 Jun 2016b. PMLR.

Peter D. Hoff, Adrian E. Raftery, and Mark S. Handcock. Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.*, 97(460):1090–1098, 2002.

Roger A. Horn. The Hadamard product. In Charles R. Johnson, editor, *Matrix Theory and Applications*, volume 40 of *Proceedings of Symposia in Applied Mathematics*, pages 87–170. American Mathematical Society, Providence, Rhode Island, 1990.

Svante Janson. Large deviations for sums of partly dependent random variables. *Random Structures & Algorithms*, 24(3):234–248, 2004.

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

Richard Kenyon, Charles Radin, Kui Ren, and Lorenzo Sadun. The phases of large networks with edge and triangle constraints. *J. Phys. A*, 50(43):435001, 22, 2017.

Aritra Konar and Nicholas D Sidiropoulos. The triangle-densest-k-subgraph problem: Hardness, Lovász extension, and application to document summarization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 4075–4082, 2022.

Siqi Liu, Sidhanth Mohanty, Tselil Schramm, and Elizabeth Yang. Testing thresholds for high-dimensional sparse random geometric graphs. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 672–677, 2022.

Suqi Liu and Miklós Z Rácz. A probabilistic view of latent space graphs and phase transitions. *Bernoulli*, 29(3):2417–2441, 2023.

Cheng Mao, Yihong Wu, and Jiaming Xu. Random geometric graphs with smooth kernels: sharp detection threshold and a spectral conjecture. *arXiv preprint arXiv:2602.14998*, 2026.

Dhruv Medarametla and Aaron Potechin. Bounds on the norms of uniform low degree graph matrices. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, volume 60 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 40, 26. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2016.

Elchanan Mossel, Joe Neeman, and Allan Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162(3):431–461, 2015.

Subhadeep Paul, Olgica Milenkovic, and Yuguo Chen. Higher-order spectral clustering under superimposed stochastic block models. *J. Mach. Learn. Res.*, 24:Paper No. [320], 58, 2023.

Goutham Rajendran and Madhur Tulsiani. Concentration of polynomial random matrices via Efron-Stein inequalities. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 3614–3653. SIAM, 2023.

Philippe Rigollet and Jan-Christian Hütter. High-dimensional statistics. *arXiv preprint arXiv:2310.19244*, 2023.

Alessandra Sala, Lili Cao, Christo Wilson, Robert Zablit, Haitao Zheng, and Ben Y Zhao. Measurement-calibrated graph models for social network experiments. In *Proceedings of the 19th international conference on World wide web*, pages 861–870, 2010.

Raman Samusevich, Maximilien Danisch, and Mauro Sozio. Local triangle-densest subgraphs. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 33–40. IEEE, 2016.

Tselil Schramm and Alexander S. Wein. Computational barriers to estimation from low-degree polynomials. *Ann. Statist.*, 50(3):1833–1858, 2022.

Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, 12 (4):389–434, 2012.

Charalampos Tsourakakis. The k-clique densest subgraph problem. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1122–1132, 2015.

Johan Ugander, Lars Backstrom, and Jon Kleinberg. Subgraph frequencies: Mapping the empirical and extremal geography of large graph collections. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1307–1318, 2013.

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge University Press, 2018.

Jiabing Wang, Rongjie Wang, Jia Wei, Qianli Ma, and Guihua Wen. Finding dense subgraphs with maximum weighted triangle density. *Information Sciences*, 539:36–48, 2020.

Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

Appendix A. Additional proofs for the spectral method

Continuing from Section 3, we bound each of the terms in (13) to prove Proposition 3.2.

A.1. Analyzing the signal term

To analyze the matrix $\mathbf{K}^2 \circ \mathbf{K}$, we first show that it is close to $(\mathbf{X}\mathbf{X}^\top) \circ (\mathbf{X}\mathbf{X}^\top)$ up to normalization.

Lemma A.1 *There is an absolute constant $C > 0$ such that the following holds for any $\delta \in (0, 1)$. If $k \geq d + \log(1/\delta)$, then with probability at least $1 - \delta$,*

$$\left\| \mathbf{K}^2 \circ \mathbf{K} - \frac{k}{d}(\mathbf{X}\mathbf{X}^\top) \circ (\mathbf{X}\mathbf{X}^\top) \right\| \leq C \frac{k^2}{d^2} \sqrt{\frac{d + \log(1/\delta)}{k}}.$$

Proof We have

$$\begin{aligned}
&\mathbf{K}^2 \circ \mathbf{K} - \frac{k}{d}(\mathbf{X}\mathbf{X}^\top) \circ (\mathbf{X}\mathbf{X}^\top) \\
&= \Big((\mathbf{X}\mathbf{X}^\top - \mathbf{I}_k)^2 - \frac{k}{d}\mathbf{X}\mathbf{X}^\top \Big) \circ (\mathbf{X}\mathbf{X}^\top) - (\mathbf{X}\mathbf{X}^\top - \mathbf{I}_k)^2 \circ \mathbf{I}_k \\
&= \Big(\mathbf{X}\mathbf{X}^\top\mathbf{X}\mathbf{X}^\top - \Big(\frac{k}{d} + 2\Big)\mathbf{X}\mathbf{X}^\top \Big) \circ (\mathbf{X}\mathbf{X}^\top) + \mathbf{I}_k \circ (\mathbf{X}\mathbf{X}^\top) - (\mathbf{X}\mathbf{X}^\top - \mathbf{I}_k)^2 \circ \mathbf{I}_k \\
&= \Big(\mathbf{X}\Big(\mathbf{X}^\top\mathbf{X} - \Big(\frac{k}{d} + 2\Big)\mathbf{I}_d \Big)\mathbf{X}^\top \Big) \circ (\mathbf{X}\mathbf{X}^\top) - (\mathbf{X}\mathbf{X}^\top)^2 \circ \mathbf{I}_k + 2\mathbf{I}_k,
\end{aligned} \tag{16}$$

where the last equality holds because $\mathbf{X}\mathbf{X}^\top$ has all ones on its diagonal.

Let us start with the first term in (16). For any matrix $L \in \mathbb{R}^{k \times k}$, by Lemma F.2,

$$\|L \circ (\mathbf{X}\mathbf{X}^\top)\| \leq \max_{i \in [k]} X_i^\top X_i \cdot \|L\| = \|L\|.$$

Therefore,

$$\left\| \Big(\mathbf{X}\Big(\mathbf{X}^\top\mathbf{X} - \Big(\frac{k}{d} + 2\Big)\mathbf{I}_d \Big)\mathbf{X}^\top \Big) \circ (\mathbf{X}\mathbf{X}^\top) \right\| \leq \|\mathbf{X}\|^2 \left\| \mathbf{X}^\top\mathbf{X} - \Big(\frac{k}{d} + 2\Big)\mathbf{I}_d \right\|.$$

By the concentration of a sample covariance matrix (see, e.g., Theorem 5.7 in Rigollet and Hütter (2023)), there is an absolute constant $C_1 > 1$ such that for any $\delta \in (0,1)$, with probability at least $1-\delta$,

$$\Big\|\frac{1}{k}\mathbf{X}^\top\mathbf{X} - \frac{1}{d}\mathbf{I}_d\Big\| \le \frac{C_1}{d}\Big(\sqrt{\frac{d+\log(1/\delta)}{k}} + \frac{d+\log(1/\delta)}{k}\Big).$$

Since $k \ge d + \log(1/\delta)$, we obtain

$$\Big\|\mathbf{X}^\top\mathbf{X} - \Big(\frac{k}{d}+2\Big)\mathbf{I}_d\Big\| \le 2 + 2C_1\frac{k}{d}\sqrt{\frac{d+\log(1/\delta)}{k}} \le 4C_1\frac{k}{d}\sqrt{\frac{d+\log(1/\delta)}{k}}$$

and

$$\|\mathbf{X}\|^2 = \|\mathbf{X}^\top\mathbf{X}\| \le \frac{k}{d} + 2C_1\frac{k}{d}\sqrt{\frac{d+\log(1/\delta)}{k}} \le (2C_1+1)\frac{k}{d}. \tag{17}$$

We conclude that

$$\Big\|\Big(\mathbf{X}\Big(\mathbf{X}^\top\mathbf{X} - \Big(\frac{k}{d}+2\Big)\mathbf{I}_d\Big)\mathbf{X}^\top\Big) \circ (\mathbf{X}\mathbf{X}^\top)\Big\| \le 4C_1(2C_1+1)\frac{k^2}{d^2}\sqrt{\frac{d+\log(1/\delta)}{k}}.$$

Moreover, the second term in (16) can be bounded as

$$\|(\mathbf{X}\mathbf{X}^\top)^2 \circ \mathbf{I}_k\| = \max_{i\in[k]} X_i^\top\mathbf{X}^\top\mathbf{X}X_i \le \|\mathbf{X}^\top\mathbf{X}\| \le (2C_1+1)\frac{k}{d}.$$

Combining the above bounds with (16) using the triangle inequality finishes the proof. ■

Next, we study the matrix $(\mathbf{X}\mathbf{X}^\top) \circ (\mathbf{X}\mathbf{X}^\top)$ and justify (10) by choosing suitable $\mathbf{Y}$ and $\mathbf{\Lambda}$. Towards this end, we use the tools of spherical harmonics (see Dai and Xu (2013) for an introduction to this topic). Let

$$D := (d-2)/2, \tag{18}$$

and consider the 0th and 2nd order Gegenbauer polynomials $C_0^D(t) = 1$ and $C_2^D(t) = 2D(D+1)t^2 - D$ respectively. Then we can write

$$t^2 = \frac{1}{2(D+1)}C_0^D(t) + \frac{1}{2D(D+1)}C_2^D(t)\,.$$

Moreover, let H_ℓ^d denote the space of real harmonic polynomials of degree ℓ on $\mathbb{R}^d$. By Corollary 1.1.4 in Dai and Xu (2013), $\dim H_0^d = 1$ and

$$m := \dim H_2^d = \binom{d+1}{2} - 1. \tag{19}$$

Let $\phi_0 \equiv 1$, and let $\phi_1, \dots, \phi_m$ be an orthonormal basis of H_2^d. The addition formula for spherical harmonics, (1.2.8) in Dai and Xu (2013), states that

$$\frac{D+2}{D}C_2^D(X_i^\top X_j) = \sum_{\ell=1}^m \phi_\ell(X_i)\phi_\ell(X_j). \tag{20}$$

Putting it together, we obtain

$$\begin{aligned}(X_i^\top X_j)^2 &= \frac{1}{2(D+1)} C_0^D(X_i^\top X_j) + \frac{1}{2D(D+1)} C_2^D(X_i^\top X_j)\\ &= \frac{1}{2(D+1)}\phi_0(X_i)\phi_0(X_j) + \frac{1}{2(D+1)(D+2)}\sum_{\ell=1}^m \phi_\ell(X_i)\phi_\ell(X_j).\end{aligned}$$

This can be rewritten as the following result.

Lemma A.2 *Let $\mathbf{Y} \in \mathbb{R}^{k\times(m+1)}$ be defined by $\mathbf{Y}_{ij} = \frac{1}{\sqrt{k}}\phi_{j-1}(X_i)$. Let the diagonal matrix $\mathbf{\Lambda} \in \mathbb{R}^{(m+1)\times(m+1)}$ be defined by $\mathbf{\Lambda}_{11} = \frac{1}{2(D+1)}$ and $\mathbf{\Lambda}_{ii} = \frac{1}{2(D+1)(D+2)}$ for $2 \le i \le m+1$. Then* (10) *holds.*

Note that the first column of $\mathbf{Y}$ is $\frac{1}{\sqrt{k}}\mathbf{1}_k$. Therefore, if $k \ge m+1$, we have the QR decomposition $\mathbf{Y} = \mathbf{QR}$ such that $\mathbf{Q} \in \mathbb{R}^{k\times(m+1)}$ has orthonormal columns with the first column also equal to $\frac{1}{\sqrt{k}}\mathbf{1}_k$, and $\mathbf{R} \in \mathbb{R}^{(m+1)\times(m+1)}$ is an upper triangular matrix. The following holds.

Lemma A.3 *Assume $k \ge m+1$ and let $\mathbf{Q}$ be as defined above. The top two eigenvalues of the matrix $\mathbf{Q\Lambda Q}^\top$ are $\mathbf{\Lambda}_{11} = \frac{1}{2(D+1)}$ and $\mathbf{\Lambda}_{22} = \frac{1}{2(D+1)(D+2)}$ respectively, and the top eigenvector is $\frac{1}{\sqrt{k}}\mathbf{1}_k$. Moreover, for any $\delta \in (0,1)$, if $k^3 \ge (2D+1)(D+2)\log\frac{2(m+1)}{\delta}$, then with probability at least $1-\delta$,*

$$\|\mathbf{Y\Lambda Y}^\top - \mathbf{Q\Lambda Q}^\top\| \le \frac{1}{2(D+1)(D+2)}\sqrt{\frac{(2D+1)(D+2)\log\frac{2(m+1)}{\delta}}{k}}.$$

Proof The first statement holds by the definition of $\mathbf{Q}$. To bound the norm, we note that

$$\|\mathbf{Y\Lambda Y}^\top - \mathbf{Q\Lambda Q}^\top\| = \|\mathbf{QR\Lambda R}^\top\mathbf{Q}^\top - \mathbf{Q\Lambda Q}^\top\| = \|\mathbf{R\Lambda R}^\top - \mathbf{\Lambda}\|.$$

In the QR decomposition $\mathbf{Y} = \mathbf{QR}$, the first columns of $\mathbf{Y}$ and $\mathbf{Q}$ are the same and $\mathbf{R}$ is upper triangular, so the first column of $\mathbf{R}$ is e_1, the first standard basis vector in $\mathbb{R}^{m+1}$. Moreover, $\mathbf{\Lambda}_{22} = \cdots = \mathbf{\Lambda}_{m+1,m+1}$ from Lemma A.2. Therefore,

$$\begin{aligned}\mathbf{R\Lambda R}^\top - \mathbf{\Lambda} &= (\mathbf{\Lambda}_{11} - \mathbf{\Lambda}_{22})e_1e_1^\top + \mathbf{\Lambda}_{22}\mathbf{RR}^\top - (\mathbf{\Lambda}_{11} - \mathbf{\Lambda}_{22})e_1e_1^\top - \mathbf{\Lambda}_{22}\mathbf{I}_{m+1}\\ &= \mathbf{\Lambda}_{22}(\mathbf{RR}^\top - \mathbf{I}_{m+1}).\end{aligned}$$

In addition, $\mathbf{Y}^\top\mathbf{Y} = \mathbf{R}^\top\mathbf{Q}^\top\mathbf{QR} = \mathbf{R}^\top\mathbf{R}$. As a result,

$$\|\mathbf{Y\Lambda Y}^\top - \mathbf{Q\Lambda Q}^\top\| = \mathbf{\Lambda}_{22}\|\mathbf{RR}^\top - \mathbf{I}_{m+1}\| = \mathbf{\Lambda}_{22}\|\mathbf{R}^\top\mathbf{R} - \mathbf{I}_{m+1}\| = \mathbf{\Lambda}_{22}\|\mathbf{Y}^\top\mathbf{Y} - \mathbf{I}_{m+1}\|.$$

Lemma 12 in De Castro et al. (2019) shows that if $\delta \in (0,1)$ and $k^3 \ge \rho(m)\log\frac{2(m+1)}{\delta}$ where $\rho(m) := \max\{1, \|\sum_{\ell=0}^m \phi_\ell^2\|_\infty - 1\}$, then with probability at least $1-\delta$,

$$\|\mathbf{Y}^\top\mathbf{Y} - \mathbf{I}_{m+1}\| \le \sqrt{\frac{\rho(m)\log\frac{2(m+1)}{\delta}}{k}}.$$

By the addition formula (20), for any $x \in \mathcal{S}^{d-1}$,

$$\sum_{\ell=0}^{m} \phi_\ell(x)^2 - 1 = \sum_{\ell=1}^{m} \phi_\ell(x)^2 = \frac{D+2}{D} C_2^D(1) = (2D+1)(D+2).$$

Combining everything finishes the proof. ■

A.2. Bounding the noise terms

We now turn to the remaining terms in (13). Each off-diagonal entry of the noise matrix $\mathbf{W}$ is obtained from centering a $\mathrm{Ber}(p)$ or $\mathrm{Ber}(p(\mathbf{K}_{ij}+1))$ variable, so it is obvious that $\mathbf{W}$ satisfies the following.

Lemma A.4 *For the noise matrix $\mathbf{W} \in \mathbb{R}^{n\times n}$ defined by $\mathbf{W}_{ii} = 0$ for $i \in [n]$, $\mathbf{W}_{ij} = \mathbf{A}_{ij} - p(\mathbf{K}_{ij}+1)$ for distinct $i, j \in [k]$, and $\mathbf{W}_{ij} = \mathbf{A}_{ij} - p$ otherwise, we have $\mathbb{E}[\mathbf{W} \mid \mathbf{X}] = 0$, $\mathrm{Var}(\mathbf{W}_{ij} \mid \mathbf{X}) \le 2p$, and $\mathbb{E}[\mathbf{W}_{ij}^4 \mid \mathbf{X}] \le 2p$ for any $i, j \in [n]$.*

We are ready to bound the noise terms in (13).

Lemma A.5 *There is an absolute constant $C > 0$ such that the following holds for any $\delta \in (0,1)$. If $k \ge d + \log(2/\delta)$ and $kp \ge C\log(2k/\delta)$, then with probability at least $1-\delta$,*

$$\|(\mathbf{K}^2)^{\#} \circ \mathbf{W}\| \le C\frac{k^{3/2}\sqrt{p}}{d}.$$

Proof Let $\mathbf{W}_{[k]}$ denote the top-left $k \times k$ principal minor of $\mathbf{W}$. By Lemma F.2,

$$\|(\mathbf{K}^2)^{\#} \circ \mathbf{W}\| = \|\mathbf{K}^2 \circ \mathbf{W}_{[k]}\| \le \max_{i\in[k]} (\mathbf{K}^2)_{ii} \cdot \|\mathbf{W}_{[k]}\|.$$

We have

$$(\mathbf{K}^2)_{ii} = \sum_{j=1}^{k} \mathbf{K}_{ij}\mathbf{K}_{ji} = X_i^\top \sum_{j\ne i} X_j X_j^\top X_i = X_i^\top \mathbf{X}^\top \mathbf{X} X_i - 1.$$

By (17), for $k \ge d + \log(2/\delta)$ and an absolute constant $C_1 > 0$,

$$\max_{i\in[k]} (\mathbf{K}^2)_{ii} \le \|\mathbf{X}^\top\mathbf{X}\| + 1 \le C_1 k/d \tag{21}$$

with probability at least $1-\delta/2$. Moreover, by Theorem F.1 and Lemma A.4, for an absolute constant $C_2 > 0$, we claim that

$$\|\mathbf{W}_{[k]}\| \le C_2\sqrt{kp + \log(k/\delta)} \tag{22}$$

with probability at least $1-\delta/2$ as long as $kp \ge C\log(2k/\delta)$. Indeed, conditionally on $\mathbf{X}$, Theorem F.1 applies with $\sigma \le \sqrt{2kp}$ and $\sigma_* \le 1$. It gives that

$$\mathbb{P}(\|\mathbf{W}_{[k]}\| \ge C_2\sqrt{kp + \log(k/\delta)}) \le k\exp(-t^2/c_\epsilon),$$

as long as $C_2\sqrt{kp+\log(k/\delta)} \geq 4(1+\epsilon)\sqrt{kp}+t$, since the right-hand side is at least $(1+\epsilon)2\sqrt{2}\sigma+t$. We satisfy this by choosing $t = C_2\sqrt{kp+\log(k/\delta)} - 4(1+\epsilon)\sqrt{kp}$. Thus the desired probability is upper-bounded by $\delta/2$ if we have $t \geq \sqrt{c_\epsilon \log(2k/\delta)}$. If we choose, say, $\epsilon = 1/2$ and $C_2 = 4(1+\epsilon)+1$, then $t \geq \sqrt{kp}$, which is at least $\sqrt{C\log(2k/\delta)}$ by assumption, as long as we choose the absolute constant to satisfy $C \geq c_{1/2}$. Our assumptions imply that $C_2\sqrt{pk+\log(k/\delta)} \leq C\sqrt{kp}$ for some absolute C, which we combine with (21) to complete the proof. ■

Lemma A.6 *There is an absolute constant $C > 0$ such that for any $\delta \in (0,1)$, if $np \geq C\log(2n/\delta)$ then it holds with probability at least $1-\delta$ that*

$$\|\mathbf{W}^2 \circ \mathbf{K}^\#\| \leq C\Big(np + \log\frac{n}{\delta}\Big).$$

Proof Let $\mathbf{W}_{1:k} \in \mathbb{R}^{n\times k}$ denote the matrix consisting of the first k columns of $\mathbf{W}$. By Lemma F.2,

$$\begin{aligned}\|\mathbf{W}^2 \circ \mathbf{K}^\#\| = \|(\mathbf{W}_{1:k}^\top\mathbf{W}_{1:k})\circ\mathbf{K}\| &\leq \|(\mathbf{W}_{1:k}^\top\mathbf{W}_{1:k})\circ(\mathbf{X}\mathbf{X}^\top)\| + \|(\mathbf{W}_{1:k}^\top\mathbf{W}_{1:k})\circ\mathbf{I}_k\|\\ &\leq 2\|\mathbf{W}_{1:k}^\top\mathbf{W}_{1:k}\| \leq 2\|\mathbf{W}\|^2.\end{aligned}$$

Moreover, by Theorem F.1 and Lemma A.4, for an absolute constant $C_1 > 0$,

$$\|\mathbf{W}\| \leq C_1\sqrt{np+\log(n/\delta)} \tag{23}$$

with probability at least $1-\delta$, as long as $np \geq C\log(2n/\delta)$. The proof of this goes as in (22); since $\mathbf{W}$ is $n\times n$ whereas $\mathbf{W}_{[k]}$ is $k\times k$, we just replace k with n everywhere in the argument. The result then follows. ■

Lemma A.7 *There is an absolute constant $C > 0$ such that the following holds for any $\delta \in (0,1)$. If $k \geq d+\log(2/\delta)$, then with probability at least $1-\delta$,*

$$\|(\mathbf{K}^\#\mathbf{W} + \mathbf{W}\mathbf{K}^\#)\circ\mathbf{K}^\#\| \leq C\frac{k}{d}\sqrt{kp+\log\frac{k}{\delta}}.$$

Proof Let $\mathbf{W}_{[k]}$ denote the top-left $k\times k$ principal minor of $\mathbf{W}$. Similar to the beginning of the proof of Lemma A.6, it holds that

$$\begin{aligned}\|(\mathbf{K}^\#\mathbf{W} + \mathbf{W}\mathbf{K}^\#)\circ\mathbf{K}^\#\| &= \|(\mathbf{K}\mathbf{W}_{[k]} + \mathbf{W}_{[k]}\mathbf{K})\circ\mathbf{K}\|\\ &\leq 2\|\mathbf{K}\mathbf{W}_{[k]} + \mathbf{W}_{[k]}\mathbf{K}\| \leq 4\|\mathbf{K}\|\|\mathbf{W}_{[k]}\|.\end{aligned}$$

By (17), for $k \geq d+\log(2/\delta)$ and an absolute constant $C_1 > 0$,

$$\|\mathbf{K}\| \leq \|\mathbf{X}^\top\mathbf{X}\| + 1 \leq C_1 k/d$$

with probability at least $1-\delta/2$. This together with (22) finishes the proof. ■

To study the term $(\mathbf{K}^\#\mathbf{W} + \mathbf{W}\mathbf{K}^\#)\circ\mathbf{W}$, we first apply a decoupling inequality.

Lemma A.8 *Let $\mathbf{W}'$ be an independent copy of $\mathbf{W}$ conditional on $\mathbf{X}$. There exists an absolute constant $C > 0$ such that for all $t > 0$,*

$$\mathbb{P}\left\{\|(\mathbf{K}^{\#}\mathbf{W} + \mathbf{W}\mathbf{K}^{\#}) \circ \mathbf{W}\| > t \mid \mathbf{X}\right\} \leq C\mathbb{P}\left\{C\|(\mathbf{K}^{\#}\mathbf{W} + \mathbf{W}\mathbf{K}^{\#}) \circ \mathbf{W}'\| > t \mid \mathbf{X}\right\}.$$

Proof Let $\Omega = [-1, 1]$ and let $\mathcal{B}$ denote the separable Banach space $\mathbb{R}^{n\times n}$ with the spectral norm. Let

$$J_n := \left\{((i, j'), (j, \ell)) \in ([n]^2)^2 : i < j', \text{ and } j < \ell, \text{ and } (i, j') \neq (j, \ell)\right\}.$$

For $((i, j'), (j, \ell)) \in J_n$, define the function $h_{ij',j\ell} : \Omega^2 \to \mathcal{B}$ by

$$h_{ij',j\ell}(x, y) := xy \cdot I_{ij'j\ell} \cdot (e_i e_{j'}^\top + e_{j'} e_i^\top),$$

where

$$I_{ij'j\ell} := \mathbf{1}\{i = j\}\,\mathbf{K}^{\#}_{j'\ell} + \mathbf{1}\{i = \ell\}\,\mathbf{K}^{\#}_{jj'} + \mathbf{1}\{j = j'\}\,\mathbf{K}^{\#}_{i\ell} + \mathbf{1}\{j' = \ell\}\,\mathbf{K}^{\#}_{ij}$$

and e_i is the ith standard basis vector. One then has

$$(\mathbf{K}^{\#}\mathbf{W} + \mathbf{W}\mathbf{K}^{\#}) \circ \mathbf{W} = \sum_{((i,j'),(j,\ell))\in J_n} h_{ij',j\ell}(\mathbf{W}_{ij'}, \mathbf{W}_{j\ell}),$$

and one can similarly check that

$$(\mathbf{K}^{\#}\mathbf{W} + \mathbf{W}\mathbf{K}^{\#}) \circ \mathbf{W}' = \sum_{((i,j'),(j,\ell))\in J_n} h_{ij',j\ell}(\mathbf{W}'_{ij'}, \mathbf{W}_{j\ell}),$$

so the result follows by applying Theorem 3.4.1 in de la Peña and Giné (2012). ■

Lemma A.9 *There is an absolute constant $C > 0$ such that if $\delta \in (0, 1)$ satisfies $np \geq \log(n/\delta)$, then with probability at least $1 - \delta$,*

$$\|(\mathbf{K}^{\#}\mathbf{W} + \mathbf{W}\mathbf{K}^{\#}) \circ \mathbf{W}\| \leq C\sqrt{\frac{np\log(n/\delta)}{d}}\left(\sqrt{kp\log(n/\delta)} + \log(n/\delta)\right).$$

Proof First, recall that $\mathbf{K}^{\#}_{ij} = \mathbf{K}_{ij} = X_i^\top X_j$ for distinct $i, j \in [k]$ and $\mathbf{K}^{\#}_{ij} = 0$ otherwise. Since $X_i^\top X_j$ is sub-Gaussian with variance $O(1/d)$, there is an absolute constant $C_1 > 0$ such that with probability at least $1 - \delta$,

$$|\mathbf{K}_{ij}| \leq C_1\sqrt{\frac{\log(n/\delta)}{d}}$$

for all $i, j \in [n]$. Let us condition on an instance of $\mathbf{X}$ such that the above bound holds.

Moreover, recall that $\mathbf{W}_{ij} \in [-1, 1]$, $\mathbb{E}[\mathbf{W} \mid \mathbf{X}] = 0$, and $\mathrm{Var}(\mathbf{W}_{ij} \mid \mathbf{X}) \leq 2p$. By Bernstein's inequality, there is an absolute constant $C_2 > 0$ such that with probability at least $1 - \delta$,

$$|(\mathbf{K}^{\#}\mathbf{W})_{ij}| = \left|\sum_{\ell=1}^{k} \mathbf{K}_{i\ell}\mathbf{W}_{\ell j}\right| \leq C_2\sqrt{\frac{\log(n/\delta)}{d}}\left(\sqrt{kp\log(n/\delta)} + \log(n/\delta)\right)$$

for all $i, j \in [n]$. We further condition on an instance of $\mathbf{W}$ such that the above bound holds.

Let $\mathbf{W}'$ be an independent copy of $\mathbf{W}$ conditional on $\mathbf{X}$. Then for an absolute constant $C_3 > 0$, we have

$$\mathbb{E}[((\mathbf{K}^{\#}\mathbf{W} + \mathbf{W}\mathbf{K}^{\#}) \circ \mathbf{W}')_{ij}^2 \mid \mathbf{X}, \mathbf{W}] \le C_3 p \frac{\log(n/\delta)}{d} \left(kp \log(n/\delta) + (\log(n/\delta))^2 \right)$$

and

$$|((\mathbf{K}^{\#}\mathbf{W} + \mathbf{W}\mathbf{K}^{\#}) \circ \mathbf{W}')_{ij}| \le C_3 \sqrt{\frac{\log(n/\delta)}{d}} \left(\sqrt{kp \log(n/\delta)} + \log(n/\delta) \right)$$

for all $i, j \in [n]$. By Theorem F.1, there are absolute constants $C_4, C_5 > 0$ such that with probability at least $1 - \delta$,

$$\begin{aligned} \|(\mathbf{K}^{\#}\mathbf{W} + \mathbf{W}\mathbf{K}^{\#}) \circ \mathbf{W}'\| &\le C_4 \Bigg(\sqrt{np \frac{\log(n/\delta)}{d} \left(kp \log(n/\delta) + (\log(n/\delta))^2\right)} \\ &\quad + \sqrt{\frac{\log(n/\delta)}{d}} \left(\sqrt{kp \log(n/\delta)} + \log(n/\delta) \right) \sqrt{\log(n/\delta)} \Bigg) \\ &\le C_5 \sqrt{\frac{\log(n/\delta)}{d}} \left(\sqrt{kp \log(n/\delta)} + \log(n/\delta) \right) \sqrt{np + \log(n/\delta)} \end{aligned}$$

Combining the above with Lemma A.8 completes the proof. ■

It remains to control $\|\mathbf{W}^2 \circ \mathbf{W}\|$. The following bound is an immediate consequence of Theorem B.1 together with the properties of $\mathbf{W}$ in Lemma A.4.

Lemma A.10 *There is an absolute constant $C > 0$ such that if $\delta \in (0, 1)$ satisfies $n \ge \log^3(n/\delta)$ and $np \ge \log^{5/3}(n/\delta)$, then with probability at least $1 - \delta$,*

$$\|\mathbf{W}^2 \circ \mathbf{W}\| \le C \left(np^{3/2} + n^{1/2} p \log(n/\delta) + \log^{3/2}(n/\delta) \right).$$

A.3. Proof of Proposition 3.2

Recall the definition of D in (18) and the definitions of $\mathbf{\Lambda}_{11}$ and $\mathbf{\Lambda}_{22}$ in Lemma A.2. The top two eigenvalues and the top eigenvector of $\widetilde{\mathbf{M}}$ are given by Lemma A.3 in view of the definition of D in (18).

To control the spectral norm, we bound the right-hand side of (13) with probability at least $1 - n^{-10}$. Towards this end, we apply Lemmas A.1, A.3, A.5, A.6, A.7, A.9, and A.10 with $\delta = n^{-10}/7$. Also recall the definitions (18) and (19). To see that these lemmas are applicable, we first check that the conditions assumed in the lemmas are all satisfied thanks to the assumption $k \ge C\sqrt{n} d / p^{3/4}$:

- $k \ge d + \log(2/\delta)$: Obvious in view of the condition $k \ge C\sqrt{n} d / p^{3/4} \ge C(d + \log n)$.
- $k \ge m + 1$: It suffices to have $k \ge d^2$ which is subsumed by $k \ge C\sqrt{n} d / p^{3/4}$ if $d \le \sqrt{n}$. Since $k \le n$, we do have $\sqrt{n} \ge Cd/p^{3/4} \ge d$.
- $k^3 \ge (2D+1)(D+2) \log \frac{2(m+1)}{\delta}$: It suffices to have $k^3 \ge Cd^2 \log n$ which is subsumed by $k \ge d^2$ proved above together with $k \ge C\sqrt{n} d / p^{3/4} \ge C \log n$.

- $kp \ge C\log(2k/\delta)$: Since $k \le n$, it suffices to show that $kp \ge C\log(14n^{11})$. As noted just after Theorem 2.5, our conditions imply $np^{3/2} \ge C$, so $kp = kp^{3/4}p^{1/4} \ge Cn^{1/2}d \cdot C^{1/6}n^{-1/6} \ge C^{7/6}n^{1/3}$ which suffices.
- $np \ge C\log(2n/\delta)$: The previous bullet point actually showed that $kp \ge C\log(2n/\delta)$, which is stronger.
- $np \ge \log(n/\delta)$: Weaker than the previous bullet point.
- $np \ge \log^{5/3}(n/\delta)$: We showed above that, under the main assumption, we have $np^{3/2} \ge C$, so that $p \ge Cn^{-2/3}$, so $np \ge Cn^{1/3} \ge \log^{5/3}(n/\delta)$.
- $n \ge \log^3(n/\delta)$: Obvious.

Now we can apply the above lemmas together with (13) to obtain

$$\begin{aligned}\left\|\mathbf{M} - \widetilde{\mathbf{M}}\right\| \le C_1\Bigg(&\frac{p^3k^2}{d^2}\sqrt{\frac{d+\log(1/\delta)}{k}} + \frac{p^3k^2}{d^2}\sqrt{\frac{\log(d/\delta)}{k}}\\ &+ \frac{p^{5/2}k^{3/2}}{d} + np^2 + p\log\frac{n}{\delta} + np^{3/2} + n^{1/2}p\log(n/\delta) + \log^{3/2}(n/\delta)\\ &+ \frac{p^2k}{d}\sqrt{kp + \log\frac{k}{\delta}} + p\sqrt{\frac{np}{d}}\left(\sqrt{kp}\log(n/\delta) + \log^{3/2}(n/\delta)\right)\Bigg)\end{aligned}$$

for an absolute constant $C_1 > 0$ with probability at least $1-\delta$ where $\delta = n^{-10}/7$.

In addition, $kp \ge \log n$ (an intermediate result of the fourth bullet point above), so the above bound simplifies to

$$\begin{aligned}&\left\|\mathbf{M} - \widetilde{\mathbf{M}}\right\|\\ &\le C_2\left(\frac{p^3k^{3/2}}{d^2}\sqrt{\log n} + \frac{p^{5/2}k^{3/2}}{d} + np^{3/2} + n^{1/2}p\log n + \log^{3/2} n + p^2\sqrt{\frac{kn}{d}}\log n\right)\end{aligned}$$

for an absolute constant $C_2 > 0$. Finally, it remains to use the assumptions $k \ge C\sqrt{n}d/p^{3/4}$ and $np^{3/2} \ge \log n$ to show that each of the following quantities is smaller than $\varepsilon/(5C_2)$ if C is sufficiently large:

- $\frac{d^2}{p^3k^2}\frac{p^3k^{3/2}}{d^2}\sqrt{\log n}$: It suffices to have $k \ge C\log n$ which is obvious.
- $\frac{d^2}{p^3k^2}\frac{p^{5/2}k^{3/2}}{d}$: It suffices to have $pk \ge Cd^2$. Since $n \ge k \ge C\sqrt{n}d/p^{3/4}$, we have $\sqrt{np} \ge Cd/p^{1/4}$, and $kp \ge C\sqrt{n}dp^{1/4} = Cn^{1/4}d(np)^{1/4} \ge Cn^{1/4}d^{3/2}/p^{1/8}$. Since we showed that $d \le \sqrt{n}$ (see the second bullet point above), it follows that $kp \ge Cd^2/p^{1/8} \ge Cd^2$.
- $\frac{d^2}{p^3k^2}np^{3/2}$: The assumption $k \ge C\sqrt{n}d/p^{3/4}$ is sufficient.
- $\frac{d^2}{p^3k^2}n^{1/2}p\log n$: Since $\frac{d}{k} \le \frac{p^{3/4}}{C\sqrt{n}}$, we have

$$\frac{d^2}{p^3k^2}n^{1/2}p\log(n) \le \frac{n^{1/2}\log(n)}{p^2}\cdot\frac{p^{3/2}}{C^2n} = \frac{\log(n)}{C^2(np)^{1/2}} = \frac{1}{C^2}\cdot\left(\frac{\log^2(n)}{np}\right)^{1/2}.$$

  Previously we showed that, under our main assumption, we have $np^{3/2} \ge C$, so that $p \ge Cn^{-2/3}$. This means that $np \ge Cn^{1/3}$, and thus $\log^2(n)/(np) = o(1)$.

- $\frac{d^2}{p^3k^2}\log^{3/2} n$: It suffices to have

$$k \geq Cd(\log n)^{3/4}/p^{3/2}$$

which is subsumed by $k \geq C\sqrt{n}d/p^{3/4}$ since $np^{3/2} \geq (\log n)^{3/2}$.

- $\frac{d^2}{p^3k^2}p^2\sqrt{\frac{kn}{d}}\log n$: It suffices to have $k \geq Cdn^{1/3}(\log n)^{2/3}/p^{2/3}$ which is subsumed by $k \geq C\sqrt{n}d/p^{3/4}$.

This completes the proof.

Appendix B. Spectral norm of a graph matrix

In this section we establish Theorem B.1, which immediately implies Lemma A.10.

Theorem B.1 *Let $\mathbf{Z} \in \mathbb{R}^{n\times n}$ be a symmetric matrix with independent entries above the diagonal and zeros on the diagonal. For all $1 \leq i < j \leq n$, assume $\|\mathbf{Z}_{ij}\|_\infty \leq 1$, $\mathbb{E}\mathbf{Z}_{ij} = 0$, $\mathbb{E}\mathbf{Z}_{ij}^2 \leq K$, and $\mathbb{E}\mathbf{Z}_{ij}^4 \leq K$ for a quantity $K \in (0, 1]$ that may depend on n. Then there is an absolute constant $C > 0$ such that if $\delta \in (0, 1)$ satisfies $n \geq \log^3(n/\delta)$ and $nK \geq \log^{5/3}(n/\delta)$, then the following holds. With probability at least $1 - \delta$, the matrix $\mathbf{L} := \mathbf{Z}^2 \circ \mathbf{Z}$ satisfies*

$$\|\mathbf{L}\| \leq C\left(nK^{3/2} + n^{1/2}K\log(n/\delta) + \log^{3/2}(n/\delta)\right).$$

A direct application of Theorem 4.1 in Rajendran and Tulsiani (2023) implies the norm of the matrix $\mathbf{L}$ from Theorem B.1 is of order $nK^{3/2}$ with an extra multiplicative logarithmic factor. The more recent work Bandeira et al. (2025) proves a general bound of the desired order on the expected spectral norm but does not directly yield a high-probability bound that we need. Our purpose in this section is to prove a high-probability spectral norm bound without any logarithmic factor in the main term.

The first step toward proving Theorem B.1 is to decouple $\mathbf{L} = \mathbf{Z}^2 \circ \mathbf{Z}$: If $\mathbf{Z}'$ is an independent copy of $\mathbf{Z}$, then we will show that it suffices to bound $\|\widetilde{\mathbf{L}}\|$ where $\widetilde{\mathbf{L}} = \mathbf{Z}^2 \circ \mathbf{Z}'$. The benefit of this decoupling—which is the main idea of the proof of Theorem B.1—is that, if we decouple and then condition on $\mathbf{Z}$, then the matrix $\widetilde{\mathbf{L}}$ has (conditionally) independent entries up to symmetry, so we can apply Theorem F.1.

Lemma B.2 *Let $\mathbf{Z}$ be as defined in Theorem B.1. Let $\mathbf{Z}'$ be an independent copy of $\mathbf{Z}$. Define the matrices $\mathbf{L} = \mathbf{Z}^2 \circ \mathbf{Z}$ and $\widetilde{\mathbf{L}} = \mathbf{Z}^2 \circ \mathbf{Z}'$. There exists an absolute constant $C > 0$ such that for all $t > 0$,*

$$\mathbb{P}\{\|\mathbf{L}\| > t\} \leq C\mathbb{P}\{C\|\widetilde{\mathbf{L}}\| > t\}.$$

Proof We will prove this result by applying the decoupling inequalities of de la Peña and Giné (2012), which means that we need to find a way to write $\mathbf{L}$ and $\widetilde{\mathbf{L}}$ in the form described there, namely as functions of some underlying independent random variables. The entries of $\mathbf{Z}$ are not independent, because of the symmetry. However, if we (abusively) write $\mathbf{Z}_{\{i,j\}}$ for the common value $\mathbf{Z}_{ij} = \mathbf{Z}_{ji}$ when $\{i, j\} \in \binom{[n]}{2}$ (i.e., $\{i, j\}$ is an unordered pair with $i \neq j$), then the random variables $\mathbf{Z}_{\{i,j\}}$ are independent as $\{i,j\}$ ranges over $\binom{[n]}{2}$. They take values in $\Omega = [-1,1]$. Let $J_n \subset \binom{[n]}{2}^3$ be the set of ordered triples of *distinct* elements of $\binom{[n]}{2}$. For each such triple $(\{i,j'\},\{j,k'\},\{k,i'\})$, define the function $h_{\{i,j'\},\{j,k'\},\{k,i'\}} : \Omega^3 \to \mathcal{B}$, where $\mathcal{B}$ is the separable Banach space of real-symmetric $n \times n$ matrices equipped with the spectral norm, by

$$h_{\{i,j'\},\{j,k'\},\{k,i'\}}(x,y,z) = \frac{1}{2}xyz \cdot \mathbf{1}\{(\{i,j'\},\{j,k'\},\{k,i'\}) \text{ form a triangle}\} \cdot (e_i e_{j'}^\top + e_{j'} e_i^\top),$$

where e_i is the ith standard basis vector, and where we say that $(\{i,j'\},\{j,k'\},\{k,i'\})$ form a triangle if all pairs of unordered pairs have nonempty intersection (i.e., if the triple is really of the form $(\{a,b\},\{b,c\},\{a,c\})$ up to reordering). Let $(\mathbf{Z}^{(a)}_{\{i,j\}})_{\{i,j\}\in\binom{[n]}{2}}$ for $a = 1,2,3$ be independent copies of $(\mathbf{Z}_{\{i,j\}})_{\{i,j\}\in\binom{[n]}{2}}$. For $a,b,c \in \{1,2,3\}$, write

$$S_{abc} := \sum_{(\{i,j'\},\{j,k'\},\{k,i'\})\in J_n} h_{\{i,j'\},\{j,k'\},\{k,i'\}}(\mathbf{Z}^{(a)}_{ij'}, \mathbf{Z}^{(b)}_{jk'}, \mathbf{Z}^{(c)}_{ki'}).$$

Then S_{111} has the same law as $\mathbf{L}$, whereas S_{122} has the same law as $\widetilde{\mathbf{L}} = \mathbf{Z}^2 \circ \mathbf{Z}'$. By the decoupling inequality in (de la Peña and Giné, 2012, Theorem 3.4.1), applied to the independent variables indexed by $\binom{[n]}{2}$, there is a universal constant $C_1 > 0$ such that

$$\mathbb{P}\{\|S_{111}\| > t\} \le C_1 \mathbb{P}\{C_1\|S_{123}\| > t\}.$$

Next, conditional on $(\mathbf{Z}^{(1)}_{\{i,j\}})_{\{i,j\}\in\binom{[n]}{2}}$, the sum S_{123} is an order-two decoupled U-statistic in the second and third coordinates. In other words, we can write it as

$$S_{123} = \sum_{(\{j,k'\},\{k,i'\})\in\widetilde{J_n}} \widetilde{h}_{\{j,k'\},\{k,i'\}}(\mathbf{Z}^{(2)}_{jk'}, \mathbf{Z}^{(3)}_{ki'}),$$

where $\widetilde{J_n} \subset \binom{[n]}{2}^2$ is the set of ordered pairs of distinct elements of $\binom{[n]}{2}$ and

$$\begin{aligned}
&\widetilde{h}_{\{j,k'\},\{k,i\}}(y,z) \\
&= \frac{yz}{2}\left(\sum_{\{i,j'\}\in\binom{[n]}{2}\setminus\{\{j,k'\},\{k,i\}\}} \mathbf{1}\{(\{i,j'\},\{j,k'\},\{k,i'\}) \text{ form a triangle}\}\mathbf{Z}^{(1)}_{ij'}(e_i e_{j'}^\top + e_{j'} e_i^\top)\right).
\end{aligned}$$

Notice that these kernels are symmetric in the sense that $\widetilde{h}_{p_1,p_2}(y,z) = \widetilde{h}_{p_2,p_1}(z,y)$. Hence the reverse decoupling inequality in (de la Peña and Giné, 2012, Theorem 3.4.1), applied conditionally on $(\mathbf{Z}^{(1)}_{\{i,j\}})_{\{i,j\}\in\binom{[n]}{2}}$, gives a universal constant $C_2 > 0$ such that

$$\mathbb{P}\{C_1\|S_{123}\| > t\} \le C_2 \mathbb{P}\{C_1 C_2 \|S_{122}\| > t\}.$$

Combining the above two inequalities gives the desired result. ■

As previously mentioned, the crux of the proof of Theorem B.1 is an application of Theorem F.1 of Bandeira and Van Handel to the partially decoupled matrix $\widetilde{\mathbf{L}}$. To apply Theorem F.1, we need good control of the variables σ and σ_*, which is given by the following two lemmas.

Lemma B.3 *Assume $\mathbf{Z} \in \mathbb{R}^{n\times n}$ satisfies the hypotheses of Theorem B.1. For all $i \in [n]$, define*

$$\sigma_i := \sqrt{\sum_{j\in[n]\setminus\{i\}} (\mathbf{Z}^2)^2_{ij}}\,,$$

where, for a matrix $\mathbf{X} \in \mathbb{R}^{n\times n}$, we write $\mathbf{X}^2_{ij} = (\mathbf{X}_{ij})^2$. Then there is an absolute constant $C > 0$ such that if $\delta \in (0,1)$ satisfies $nK \geq \log(n/\delta)$ and $n \geq \log^3(n/\delta)$, then the following holds. With probability at least $1-\delta$, we have that for all $i \in [n]$,

$$\sigma_i \leq C\left(nK + n^{1/4}K^{1/4}\log^{5/4}(n/\delta)\right).$$

Lemma B.4 *Assume $\mathbf{Z} \in \mathbb{R}^{n\times n}$ satisfies the hypotheses of Theorem B.1. Then there is an absolute constant $C > 0$ such that the following holds for any $\delta \in (0,1)$. With probability at least $1-\delta$, we have that for all $i, j \in [n]$,*

$$|(\mathbf{Z}^2)_{ij}| \leq C\left(n^{1/2}K\log^{1/2}(n/\delta) + \log(n/\delta)\right).$$

We first prove Theorem B.1 using Theorem F.1 and Lemmas B.3 and B.4.

Proof of Theorem B.1 Let $\mathbf{Z}'$ be an independent copy of $\mathbf{Z}$ and let $\widetilde{\mathbf{L}} := \mathbf{Z}^2 \circ \mathbf{Z}'$. By Lemma B.2, it suffices to prove the tail bound with $\widetilde{\mathbf{L}}$ in place of $\mathbf{L}$. In the sequel, we condition on a realization of $\mathbf{Z}$ such that the bounds in Lemmas B.3 and B.4 hold, and all the probabilities and expectations are with respect to $\mathbf{Z}'$ conditional on $\mathbf{Z}$. Define the quantities

$$\sigma := \max_{i=1,\ldots,n} \sqrt{\sum_{j=1}^{n} \mathbb{E}\Big[\big(\widetilde{\mathbf{L}}_{ij}\big)^2\Big]}\,, \qquad \sigma_* := \max_{i,j=1,\ldots,n} \|\widetilde{\mathbf{L}}_{ij}\|_\infty\,.$$

Since $\mathbb{E}(\mathbf{Z}'_{ij})^2 \leq K$, by the bound in Lemma B.3, we obtain

$$\sigma = \max_{i=1,\ldots,n} \sqrt{\sum_{j\in[n]\setminus\{i\}} (\mathbf{Z}^2)^2_{ij}\mathbb{E}(\mathbf{Z}'_{ij})^2} \leq C\sqrt{K}\left(nK + n^{1/4}K^{1/4}\log^{5/4}(n/\delta)\right).$$

Since $\|\mathbf{Z}'_{ij}\|_\infty \leq 1$, by the bound in Lemma B.4, we obtain

$$\sigma_* = \max_{i,j=1,\ldots,n} |(\mathbf{Z}^2)_{ij}| \cdot \|\mathbf{Z}'_{ij}\|_\infty \leq C\left(n^{1/2}K\log^{1/2}(n/\delta) + \log(n/\delta)\right).$$

We then claim that, for any $\delta \in (0,1)$, we have

$$\begin{aligned}
\|\widetilde{\mathbf{L}}\| &\lesssim \sigma + \sigma_*\sqrt{\log(n/\delta)} \\
&\lesssim \sqrt{K}\left(nK + n^{1/4}K^{1/4}\log^{5/4}(n/\delta)\right) + \left(n^{1/2}K\log^{1/2}(n/\delta) + \log(n/\delta)\right)\sqrt{\log(n/\delta)} \\
&\lesssim nK^{3/2} + n^{1/2}K\log(n/\delta) + \log^{3/2}(n/\delta)
\end{aligned}$$

with probability at least $1-\delta$. The first and second steps use Theorem F.1 and the estimates earlier in this proof. In the third step, we use our assumption that $nK \geq \log^{5/3}(n/\delta)$, so that $\log(n/\delta) \leq$

$(nK)^{3/5}$ and thus $\log^{5/4}(n/\delta) \leq (nK)^{3/4}$. This implies that $n^{1/4}K^{3/4}\log^{5/4}(n/\delta) \leq nK^{3/2}$. Taking a union bound completes the proof. ■

Proof of Lemma B.3 Fix $i \in [n]$. Let $\mathbf{Z}_i$ denote the ith column of $\mathbf{Z}$. Note that

$$\sigma_i^2 = \sum_{j \in [n] \setminus \{i\}} \langle \mathbf{Z}_i, \mathbf{Z}_j \rangle^2 .$$

If we define the quantities

$$S_1 := \sum_{\substack{j,r \in [n] \setminus \{i\} \\ j \neq r}} \mathbf{Z}_{ir}^2 \mathbf{Z}_{jr}^2 , \qquad S_2 := \sum_{\substack{j,r,s \in [n] \setminus \{i\} \\ \text{all distinct}}} \mathbf{Z}_{ir}\mathbf{Z}_{jr}\mathbf{Z}_{is}\mathbf{Z}_{js} ,$$

then $\sigma_i^2 = S_1 + S_2$.

Using the fact that $\|\mathbf{Z}_{jr}^2\|_\infty \leq 1$, $\mathbb{E}\mathbf{Z}_{jr}^2 \leq K$, and $\mathrm{Var}[\mathbf{Z}_{jr}^2] \leq K$, Bernstein's inequality implies that for all $r \in [n] \setminus \{i\}$, the event

$$\sum_{j \in [n] \setminus \{i,r\}} \mathbf{Z}_{jr}^2 \lesssim nK + \sqrt{nK \log\left(\frac{n}{\delta}\right)} + \log\left(\frac{n}{\delta}\right) \lesssim nK \tag{24}$$

occurs with probability at least $1 - \delta/n$, using the assumption $nK \geq \log(n/\delta)$. Hence by applying (24) n times ($n-1$ times corresponding with the indexes $r \in [n] \setminus \{i\}$ and once to the sum over r), we find that the event

$$S_1 = \sum_{r \in [n] \setminus \{i\}} \mathbf{Z}_{ir}^2 \sum_{j \in [n] \setminus \{i,r\}} \mathbf{Z}_{jr}^2 \lesssim nK \sum_{r \in [n] \setminus \{i\}} \mathbf{Z}_{ir}^2 \lesssim n^2K^2$$

occurs with probability at least $1 - \delta$, which follows from a union bound.

To bound S_2, first define $B_{rs} := \sum_{j \in [n] \setminus \{i,r,s\}} \mathbf{Z}_{jr}\mathbf{Z}_{js}$ for all $r, s \in [n] \setminus \{i\}$ such that $r \neq s$. The quantities B_{rs} satisfy

$$S_2 = \sum_{r \in [n] \setminus \{i\}} \mathbf{Z}_{ir} \sum_{s \in [n] \setminus \{i,r\}} \mathbf{Z}_{is} B_{rs} .$$

Let $\{\mathbf{Z}'_{ij}\}_{j \in [n] \setminus \{i\}}$ be an independent copy of the collection $\{\mathbf{Z}_{ij}\}_{j \in [n] \setminus \{i\}}$, and define

$$S_2' := \sum_{r \in [n] \setminus \{i\}} \mathbf{Z}_{ir} \sum_{s \in [n] \setminus \{i,r\}} \mathbf{Z}'_{is} B_{rs} . \tag{25}$$

For all $(r,s) \in ([n] \setminus \{i\})^2$ such that $r \neq s$, define $h_{r,s} : \mathbb{R}^2 \to \mathbb{R}$ by

$$h_{r,s}(x,y) := xy \cdot B_{rs} .$$

Since

$$S_2 = \sum_{\substack{(r,s) \in ([n] \setminus \{i\})^2 \\ r \neq s}} h_{r,s}(\mathbf{Z}_{ir}, \mathbf{Z}_{is}) ,$$

we can apply the decoupling inequality in Theorem 3.4.1 from de la Peña and Giné (2012) to deduce that for all $t > 0$,

$$\mathbb{P}\{|S_2| \geq t\} \leq C \cdot \mathbb{P}\left\{C \cdot |S_2'| \geq t\right\} \tag{26}$$

for an absolute constant $C > 0$. (Notice that, although the functions $h_{r,s}$ are random, the collection $(B_{rs})_{r,s\in[n]\setminus\{i\}}$ is independent of the collection $(\mathbf{Z}_{ir})_{r\in([n]\setminus\{i\})}$, because the former sees neither the ith row nor the ith column of $\mathbf{Z}$. Thus we can first condition on B_{rs}, retaining the independence of the $\mathbf{Z}_{ir}$'s necessary to apply de la Peña and Giné (2012), then use that C is universal to see that the bound also holds unconditionally.)

We will apply Bernstein's inequality three times iteratively to the terms in (25). First, since $\|\mathbf{Z}_{jr}\mathbf{Z}_{js}\|_\infty \leq 1$ and $\mathrm{Var}[\mathbf{Z}_{jr}\mathbf{Z}_{js}] \leq K^2$, Bernstein's inequality implies that for all $r, s \in [n] \setminus \{i\}$ with $r \neq s$, the event

$$|B_{rs}| \lesssim \sqrt{nK^2 \log\left(\frac{n}{\delta}\right)} + \log\left(\frac{n}{\delta}\right) \tag{27}$$

occurs with probability at least $1 - \delta/n^2$. For all $r \in [n] \setminus \{i\}$, define

$$D_r := \sum_{s\in[n]\setminus\{i,r\}} \mathbf{Z}_{is}' B_{rs}\,.$$

For fixed values of B_{rs} for $s \in [n] \setminus \{i, r\}$, Bernstein's inequality implies that the event

$$|D_r| = \left|\sum_{s\in[n]\setminus\{i,r\}} \mathbf{Z}_{is}' B_{rs}\right| \lesssim \sqrt{\sum_{s\in[n]\setminus\{i,r\}} K B_{rs}^2 \log\left(\frac{n}{\delta}\right)} + \max_s |B_{rs}| \log\left(\frac{n}{\delta}\right) \tag{28}$$

occurs with probability at least $1 - \delta/n$. We claim that conditioning on (27) for all $s \in [n] \setminus \{i, r\}$ and applying (28) will give

$$|D_r| \lesssim nK^{3/2} \log\left(\frac{n}{\delta}\right) + \log^2\left(\frac{n}{\delta}\right) \tag{29}$$

with probability at least $1-\delta/n$. Indeed, let $\tilde{\mathbf{Z}}$ be the principal submatrix of $\mathbf{Z}$ indexed by $[n]\setminus\{i, r\}$, and let $w = (\mathbf{Z}_{jr})_{j\in[n]\setminus\{i,r\}}$. Then $(B_{rs})_{s\in[n]\setminus\{i,r\}} = \tilde{\mathbf{Z}}w$ as a column vector. By Theorem F.1 and Bernstein's inequality, with probability at least $1 - \delta/n$,

$$\|\tilde{\mathbf{Z}}\| \lesssim \sqrt{nK} + \sqrt{\log(n/\delta)} \lesssim \sqrt{nK}, \qquad \|w\|_2^2 \lesssim nK + \sqrt{nK\log(n/\delta)} + \log(n/\delta) \lesssim nK,$$

where we use $nK \geq \log(n/\delta)$. Thus $\sum_{s\in[n]\setminus\{i,r\}} B_{rs}^2 \leq \|\tilde{\mathbf{Z}}\|^2\|w\|_2^2 \lesssim n^2K^2$. Combining this with (27) and (28) gives

$$|D_r| \lesssim nK^{3/2} \log^{1/2}\left(\frac{n}{\delta}\right) + n^{1/2}K \log^{3/2}\left(\frac{n}{\delta}\right) + \log^2\left(\frac{n}{\delta}\right).$$

The assumption $nK \geq \log(n/\delta)$ implies $n^{1/2}K \log^{3/2}(n/\delta) \leq nK^{3/2} \log(n/\delta)$, proving (29).

Now for fixed values of D_r for $r \in [n] \setminus \{i\}$, Bernstein's inequality implies that the event

$$|S_2'| = \left|\sum_{r\in[n]\setminus\{i\}} \mathbf{Z}_{ir} D_r\right| \lesssim \sqrt{\sum_{r\in[n]\setminus\{i\}} K D_r^2 \log\left(\frac{n}{\delta}\right)} + \max_r |D_r| \log\left(\frac{n}{\delta}\right) \tag{30}$$

occurs with probability at least $1-\delta$. (Notice that the collections $(D_r)_{r\in[n]\setminus\{i\}}$ and $(\mathbf{Z}_{ir})_{r\in[n]\setminus\{i\}}$ are independent, so we can indeed fix the values of D_r.) Hence if we condition on (29) for all $r\in[n]\setminus\{i\}$ and apply (30) together with (26), then we find that the event

$$|S_2| \lesssim n^{3/2}K^2\log^{3/2}\left(\frac{n}{\delta}\right) + n^{1/2}K^{1/2}\log^{5/2}\left(\frac{n}{\delta}\right) \tag{31}$$

occurs with probability at least $1-\delta$, using the assumption $nK \geq \log(n/\delta)$. Using a union bound over the above events, the probability of (31) is at least $1-3\delta$.

Hence combining our bounds for S_1 and S_2 together with the assumption $n \geq \log^3(n/\delta)$, we find that the event

$$\sigma_i^2 \lesssim n^2K^2 + n^{1/2}K^{1/2}\log^{5/2}\left(\frac{n}{\delta}\right)$$

occurs with probability at least $1-4\delta$, completing the proof. ■

Proof of Lemma B.4 We will use that $\|\mathbf{Z}_{ik}\mathbf{Z}_{jk}\|_\infty \leq 1$ and $\operatorname{Var}\mathbf{Z}_{ik}\mathbf{Z}_{jk} \leq K^2$, which holds for all distinct $i,j,k\in[n]$. Using Bernstein's inequality, we have that for any $\delta\in(0,1)$, with probability at least $1-\delta/n^2$,

$$|(\mathbf{Z}^2)_{ij}| = \left|\sum_{k\in[n]\setminus\{i,j\}} \mathbf{Z}_{ik}\mathbf{Z}_{jk}\right| \leq C\left(\sqrt{nK^2\log(n/\delta)} + \log(n/\delta)\right)$$

for an absolute constant $C>0$. Taking a union bound completes the proof. ■

Appendix C. Additional proofs for the semidefinite program

This section is devoted to Lemma C.1, a key step in the proof of Theorem 2.6. We will refer repeatedly to our assumptions on n, k, p, d, and δ:

$$\text{(A1)}\;\; np^{3/2} = \omega((d\log n)^2)\,, \qquad \text{(A2)}\;\; kp^{3/4} \geq Cn^{1/2}d\,,$$

where $C>0$ is a sufficiently large constant.

Lemma C.1 *For all fixed $\epsilon\in(0,1)$, if η is defined as in* (14) *and (A1) and (A2) hold, then for all large enough n depending on ϵ, we have $\eta > (1-\epsilon)\cdot p^3k^2/d^2$ with probability at least $1-n^{-10}/2$.*

Proof Let $\delta := n^{-10}/2$. Recall that $\eta := \min_{1\leq i\leq k}(\mathbf{M}\xi)_i - \max_{k<i\leq n}|(\mathbf{M}\xi)_i|$. The proof resembles the spectral method in Section 3 and analyzes the two cases $1\leq i\leq k$ and $k<i\leq n$. In both cases we have $(\mathbf{M}\xi)_i = \sum_{j=1}^k \mathbf{M}_{ij}$.

Case 1: $i\leq k$. By (9), it holds that

$$\sum_{j=1}^k \mathbf{M}_{ij} \geq \underbrace{p^3\sum_{j=1}^k(\mathbf{K}^2\circ\mathbf{K})_{ij}}_{T_1} - \underbrace{p^2\left|\sum_{j=1}^k(\mathbf{K}^2)_{ij}\mathbf{W}_{ij}\right|}_{T_2} - \underbrace{p\left|\sum_{j=1}^k(\mathbf{W}^2)_{ij}\mathbf{K}_{ij}\right|}_{T_3} - \underbrace{\left|\sum_{j=1}^k(\mathbf{W}^2)_{ij}\mathbf{W}_{ij}\right|}_{T_4}$$
$$- \underbrace{p^2\left|\sum_{j=1}^k(\mathbf{K}^{\#}\mathbf{W}+\mathbf{W}\mathbf{K}^{\#})_{ij}\mathbf{K}_{ij}\right|}_{T_5} - \underbrace{p\left|\sum_{j=1}^k(\mathbf{K}^{\#}\mathbf{W}+\mathbf{W}\mathbf{K}^{\#})_{ij}\mathbf{W}_{ij}\right|}_{T_6}.$$

Case 2: $i > k$. By (9), it holds that

$$\Big|\sum_{j=1}^{k}\mathbf{M}_{ij}\Big| \leq \underbrace{\Big|\sum_{j=1}^{k}(\mathbf{W}^2)_{ij}\mathbf{W}_{ij}\Big|}_{T_7} + \underbrace{p\Big|\sum_{j=1}^{k}(\mathbf{W}\mathbf{K}^{\#})_{ij}\mathbf{W}_{ij}\Big|}_{T_8}$$

since all the other terms vanish for $i > k$.

For all $1 \leq i \leq k$ and $k < j \leq n$, define the event

$$E_{ij} := \left\{(\mathbf{M}\xi)_i - |(\mathbf{M}\xi)_j| \leq (1-\epsilon)\frac{p^3k^2}{d^2}\right\} .$$

Using the inclusion

$$E_{ij} \subseteq \left\{T_1 \leq \left(1-\frac{\epsilon}{2}\right)\frac{p^3k^2}{d^2}\right\} \cup \bigcup_{l=2}^{8}\left\{T_l \geq \frac{\epsilon}{14}\frac{p^3k^2}{d^2}\right\} ,$$

we calculate that

$$\begin{aligned}\mathbb{P}\left\{\eta \leq (1-\epsilon)\frac{p^3k^2}{d^2}\right\} &\leq \sum_{1\leq i\leq k}\sum_{k<j\leq n}\mathbb{P}\left\{E_{ij}\right\} \\ &\leq n^2\left(\mathbb{P}\left\{T_1 \leq \left(1-\frac{\epsilon}{2}\right)\frac{p^3k^2}{d^2}\right\} + \sum_{i=2}^{8}\mathbb{P}\left\{T_i \geq \frac{\epsilon}{14}\frac{p^3k^2}{d^2}\right\}\right) .\end{aligned}$$

We will show in the second half of the proof that assumptions (A1) and (A2) imply the right-hand sides of Lemmas C.3 to C.5, C.7 and C.8 are all $o(p^3k^2/d^2)$ after multiplying by the corresponding factor of p. For the moment, we use this fact to complete the proof of the lemma. We apply Lemma C.2 with $\lambda = 1-\epsilon/2$ and Lemmas C.3 to C.5, C.7 and C.8 with $\delta = n^{-12}/8$. Notice that T_7 is the same as T_4; additionally, T_8 appears as the object S_2 in the proof of Lemma C.8, so the bound from Lemma C.8 also bounds T_8. We then obtain

$$\mathbb{P}\left\{\eta \leq (1-\epsilon)\frac{p^3k^2}{d^2}\right\} \leq n^2\left(e^{-ck/d} + \frac{7}{n^{12}}\right) \leq e^{-ck/2d} + \frac{7}{8n^{10}} \leq \frac{1}{n^{10}} ,$$

where in the last step we used that $k/d \gg \log n$, which follows easily from (A2) (this also implies that $k \geq d + \log(1/\delta)$, which we needed to apply these lemmas).

It remains to verify the right-hand sides of Lemmas C.3 to C.5, C.7 and C.8 are $o(p^3k^2/d^2)$ after multiplying by the corresponding factor of p. We will repeatedly use the following easy consequences of (A1) and (A2). Let $L := \log(1/\delta)$, $L_k := \log(k/\delta)$, and $L_n := \log(n/\delta)$. Since $\delta = n^{-10}/2$, we have $L_n \lesssim \log n$. From (A2), we have $pk \geq Cn^{2/3}d^{4/3}k^{-1/3}$, hence

$$\frac{d\log^{4/3}n}{pk} \lesssim \frac{d\log^{4/3}n}{n^{2/3}d^{4/3}k^{-1/3}} = \frac{k^{1/3}\log^{4/3}n}{n^{2/3}d^{1/3}} \leq \frac{\log^{4/3}n}{n^{1/3}} = o(1) . \tag{32}$$

We verify the terms one at a time:

(LC.3) Dividing by p^3k^2/d^2, the two constituent terms become $\sqrt{dL/pk}$ and dL/pk, respectively, which are $o(1)$ by $L \lesssim \log n$ and (32).

(LC.4) For the first term, we calculate

$$\frac{p^2\sqrt{nk/d}\,L_n}{p^3k^2/d^2} = \frac{d^{3/2}\sqrt{n}\,L_n}{pk^{3/2}} \lesssim \log n \cdot \left(\frac{d}{nk}\right)^{1/6} = o(1)\,,$$

where the inequality used $L_n \lesssim \log n$ and (A2) in the form $p \gtrsim (\sqrt{n}d/k)^{4/3}$. For the second term,

$$\frac{p\sqrt{kp/d}\,L_n^2}{p^3k^2/d^2} = \left(\frac{dL_n^{4/3}}{pk}\right)^{3/2} = o(1)\,,$$

where we used $L_n \lesssim \log n$ and (32).

(LC.5) For the first term, we calculate

$$\frac{\sqrt{nk}p^{3/2}\sqrt{L}}{p^3k^2/d^2} = \frac{d^2\sqrt{nL}}{p^{3/2}k^{3/2}} = o(1)\,,$$

where the $o(1)$ follows since $\frac{d^2\sqrt{nL}}{p^{3/2}k^{3/2}} = o(1)$ is equivalent to $k \gg (nL)^{1/3}d^{4/3}/p$, which in turn holds for the following reason: (A2) asserts $k \gtrsim \sqrt{n}d/p^{3/4}$, and taking the ratio

$$R := \frac{(nL)^{1/3}d^{4/3}/p}{\sqrt{n}d/p^{3/4}} = \frac{(dL)^{1/3}}{n^{1/6}p^{1/4}}\,,$$

we notice that $R = o(1)$ if and only if $np^{3/2} \gg (dL)^2$, which is precisely (A1). For the second term,

$$\frac{p\sqrt{k}L_n^{3/2}}{p^3k^2/d^2} = \frac{d^2L_n^{3/2}}{p^2k^{3/2}} \lesssim \frac{k^{7/6}(\log n)^{3/2}}{n^{4/3}d^{2/3}} \le \frac{\log^{3/2} n}{n^{1/6}d^{2/3}} = o(1)\,,$$

where the first inequality uses $L_n \lesssim \log n$ and (A2) in the form $p^2 \gtrsim (\sqrt{n}d/k)^{8/3}$, and the second inequality uses $k \le n$. For the third term, the condition $\frac{L_n^2}{p^3k^2/d^2} = o(1)$ is equivalent to $k \gg (dL_n)/p^{3/2}$, which holds for the following reason: again comparing with the condition $k \gtrsim \sqrt{n}d/p^{3/4}$ from (A2), the ratio

$$\frac{dL_n/p^{3/2}}{\sqrt{n}d/p^{3/4}} = \frac{L_n}{\sqrt{n}p^{3/4}}$$

is $o(1)$ if and only if $np^{3/2} \gg L_n^2$, which holds by (A1).

(LC.7) Identical to (LC.3).

(LC.8) For the first term, we calculate

$$\frac{p^2kL_k/\sqrt{d}}{p^3k^2/d^2} = \frac{d^{3/2}L_k}{pk} \lesssim \frac{d^{1/6}k^{1/3}\log n}{n^{2/3}} \le \frac{d^{1/6}\log n}{n^{1/3}} = o(1)\,,$$

where we used $L_k \lesssim \log n$ and (A2) in the form $p \gtrsim (\sqrt{n}d/k)^{4/3}$. Finally, (A2) implies $d \lesssim k/\sqrt{n} \le \sqrt{n}$, hence $\frac{d^{1/6}\log n}{n^{1/3}} \lesssim \frac{\log n}{n^{1/4}} = o(1)$. For the second term,

$$\frac{p\sqrt{pk/d}\,L_k^2}{p^3k^2/d^2} = \frac{d^{3/2}L_k^2}{p^{3/2}k^{3/2}} \lesssim \frac{\sqrt{k}(\log n)^2}{n\sqrt{d}} \le \frac{(\log n)^2}{\sqrt{n}} = o(1)\,,$$

where we used $L_k \lesssim \log n$ and then (A2) in the form $p^{3/2} \gtrsim (\sqrt{n}d/k)^2 = nd^2/k^2$.

Having verified the right-hand sides of Lemmas C.3 to C.5, C.7 and C.8 are of the correct order, the proof is complete. ■

Lemma C.2 *Let $S := \sum_{j=1}^k (\mathbf{K}^2 \circ \mathbf{K})_{1j}$. For every fixed $\lambda \in (0,1)$ there exists a constant $c > 0$ such that*

$$\mathbb{P}\left\{S \le \lambda \frac{k^2}{d^2}\right\} \le e^{-ck/d}\,.$$

Proof Throughout the proof we condition on X_1. We can rewrite

$$S = \sum_{\substack{2\le j,l\le k\\ j\neq l}} \langle X_1, X_j\rangle\langle X_1, X_l\rangle\langle X_j, X_l\rangle\,. \tag{33}$$

Define the centered kernel $h : (\mathbb{S}^{d-1})^2 \to \mathbb{R}$ by

$$h(x,y) := \langle X_1, x\rangle\langle X_1, y\rangle\langle x, y\rangle - \frac{1}{d}\langle X_1, x\rangle^2 - \frac{1}{d}\langle X_1, y\rangle^2 + \frac{1}{d^2}\,. \tag{34}$$

For $Y \sim \mathrm{Unif}(\mathbb{S}^{d-1})$ and any fixed $x \in \mathbb{S}^{d-1}$, we have

$$\mathbb{E}_Y[\langle X_1, x\rangle\langle X_1, Y\rangle\langle x, Y\rangle \mid X_1] = \langle X_1, x\rangle \cdot X_1^\top \mathbb{E}[YY^\top]x = \frac{1}{d}\langle X_1, x\rangle^2\,,$$
$$\mathbb{E}_Y[\langle X_1, Y\rangle^2 \mid X_1] = \frac{1}{d}\langle X_1, X_1\rangle^2 = \frac{1}{d}\,,$$

which implies

$$\mathbb{E}_Y[h(x,Y) \mid X_1] = 0\,. \tag{35}$$

By symmetry the same holds with x and y swapped.

Writing $W_j := \langle X_1, X_j\rangle^2 - 1/d$, we directly expand (33) and (34) to obtain

$$S = \mu + \frac{2(k-2)}{d}\sum_{j=2}^k W_j + U\,, \quad \mu := \frac{(k-1)(k-2)}{d^2}\,, \quad U := \sum_{\substack{2\le j,l\le k\\ j\neq l}} h(X_j, X_l). \tag{36}$$

In particular, $\mathbb{E}[S \mid X_1] = \mu$. If we set $t := (1-\lambda)\mu$ then we find from (36) that

$$\{S \le \mu - t\} \subseteq \left\{U \le -\frac{t}{2}\right\} \cup \left\{\frac{2(k-2)}{d}\sum_{j=2}^k W_j \le -\frac{t}{2}\right\}. \tag{37}$$

We will bound the two terms separately.

The random variables $W_2, \ldots, W_k$ are independent given X_1, satisfy $\|W_j\|_{\psi_1} \lesssim 1/d$, and are mean-zero. Hence by the subexponential Bernstein inequality (e.g. Theorem 2.8.1 in Vershynin (2018)) conditional on X_1, there exists $c > 0$ such that

$$\mathbb{P}\left\{ \sum_{j=2}^{k} W_j \le -s \right\} \le \exp\left(-c \min\left\{ \frac{s^2 d^2}{k}, \, sd \right\} \right)$$

for all $s > 0$. Taking $s := \frac{td}{4(k-2)} = \frac{(1-\lambda)(k-1)}{4d}$, we obtain

$$\mathbb{P}\left\{ \frac{2(k-2)}{d} \sum_{j=2}^{k} W_j \le -\frac{t}{2} \right\} \le e^{-c(1-\lambda)^2 k} . \tag{38}$$

We now move on to the U-statistic term. Let $\mathbf{X}' := \{X'_j\}_{j\in[k]}$ and $\mathbf{X}'' := \{X''_j\}_{j\in[k]}$ be independent copies of $\{X_j\}_{j\in[k]}$, and define the decoupled U-statistic

$$U' := \sum_{\substack{2 \le j,l \le k \\ j \ne l}} h(X'_j, X''_l) .$$

By the decoupling inequality for order-2 U-statistics (de la Peña and Montgomery-Smith, 1995, Theorem 1), there is an absolute constant $C' > 0$ such that for all $x > 0$,

$$\mathbb{P}\left\{ |U| \ge x \right\} \le C' \cdot \mathbb{P}\left\{ C' \cdot |U'| \ge x \right\} . \tag{39}$$

We now estimate tail probabilities for U' using an exponential inequality for U-statistics due to Giné et al., which we apply conditionally on X_1. Equation (35) shows that, conditional on X_1, the kernel h is canonical in the sense needed to apply (Giné et al., 2000, Theorem 3.3). This result is written in terms of various norms and expectations of h, which we now estimate using the simpler forms of A, B, C, and D appearing before (Giné et al., 2000, Corollary 3.4).

Bound for A. It is easy to see that $A = \|h\|_\infty \le 1 + 2/d + 1/d^2 \le 4$.

Bounds for B and C. We will use that for any fixed $v, w \in \mathbb{S}^{d-1}$,

$$\mathbb{E}[\langle v, Y \rangle^2 \langle w, Y \rangle^2] = \frac{1 + 2\langle v, w \rangle^2}{d(d+2)} .$$

For any fixed $x \in \mathbb{S}^{d-1}$, using $(a+b+c+d)^2 \le 4(a^2+b^2+c^2+d^2)$, we find that there exists a universal constant $C > 0$ such that

$$\begin{aligned} \mathbb{E}_Y[h(x,Y)^2 \,|\, X_1] &\le 4\,\mathbb{E}_Y[\langle X_1, x \rangle^2 \langle X_1, Y \rangle^2 \langle x, Y \rangle^2] \\ &\qquad + \frac{4}{d^2} \langle X_1, x \rangle^4 + \frac{4}{d^2} \mathbb{E}_Y[\langle X_1, Y \rangle^4] + \frac{4}{d^4} \\ &\le 4 \langle X_1, x \rangle^2 \, \frac{1 + 2\langle X_1, x \rangle^2}{d(d+2)} + O\left(\frac{1}{d^2} \right) \le \frac{C}{d^2}. \end{aligned} \tag{40}$$

The same estimate holds with x and Y interchanged, hence

$$B^2 = (k-1)\big(\|\mathbb{E}_Y h(\cdot, Y)^2\|_\infty + \|\mathbb{E}_X h(X, \cdot)^2\|_\infty \big) \lesssim \frac{k}{d^2} ,$$

where the implicit constant is universal. Similarly, taking expectation of (40) over $x = X$ and using $\mathbb{E}[\langle X_1, X\rangle^2] = 1/d$ gives $\mathbb{E}[h(X,Y)^2] \lesssim 1/d^3$, which implies

$$C^2 = (k-1)^2\,\mathbb{E}[h(X,Y)^2] \lesssim \frac{k^2}{d^3}\,,$$

where the implicit constant is again universal.

Bound for D. From (Giné et al., 2000, p. 20), we have $D = (k-1)\|h\|_{L_2\to L_2}$ and $\|h\|_{L_2\to L_2} \le \|h\|_{L_2} = \sqrt{\mathbb{E}[h(X,Y)^2]}$ by Cauchy–Schwarz. Thus since $\mathbb{E}[h(X,Y)^2] \lesssim 1/d^3$,

$$D \le (k-1)\|h\|_{L_2} \lesssim \frac{k}{d^{3/2}}\,,$$

where the implicit constant is universal.

Using the above bounds on A, B, C, and D, we apply (Giné et al., 2000, Theorem 3.3) to U', conditional on X_1, to obtain for all $x > 0$,

$$\mathbb{P}\left\{|U'| \ge x \,|\, X_1\right\} \le L\exp\left(-\frac{1}{L}\min\left\{\frac{x^2}{C^2}, \frac{x}{D}, \frac{x^{2/3}}{B^{2/3}}, \frac{x^{1/2}}{A^{1/2}}\right\}\right),$$

where $L > 0$ is an absolute constant. Applying our upper bounds for A, B, C, and D (which do not depend on X_1), and then taking expectations over X_1 and using (39) gives

$$\mathbb{P}\left\{|U| \ge x\right\} \le C\exp\left(-c\min\left\{\frac{x^2d^3}{k^2}, \frac{xd^{3/2}}{k}, \frac{x^{2/3}d^{2/3}}{k^{1/3}}, x^{1/2}\right\}\right).$$

Setting $x := t/2 = (1-\lambda)\mu/2$ and using $\mu \sim k^2/d^2$ yields

$$\mathbb{P}\left\{U \le -\frac{t}{2}\right\} \le \mathbb{P}\left\{|U| \ge \frac{t}{2}\right\} \le C\exp\left(-c\frac{k}{d}\right)\,, \tag{41}$$

where we used that λ is fixed and the last term in the minimum dominates. Combining (37), (38), and (41) completes the proof. ■

Lemma C.3 *There is an absolute constant $C > 0$ such that the following holds. Assume $k \ge d + \log(1/\delta)$. For all $1 \le i \le k$ and $\delta \in (0,1)$,*

$$\left|\sum_{j=1}^{k}(\mathbf{K}^2)_{ij}\mathbf{W}_{ij}\right| \le C\left(\sqrt{\frac{pk^3}{d^3}\log\left(\frac{1}{\delta}\right)} + \frac{k}{d}\log\left(\frac{1}{\delta}\right)\right)$$

with probability at least $1-\delta$.

Proof Fix $1 \le i \le k$. Let $S := \sum_{j=1}^{k}(\mathbf{K}^2)_{ij}\mathbf{W}_{ij}$ and $B_j := (\mathbf{K}^2)_{ij}$ for all $j \in [k]\setminus\{i\}$. Throughout the proof, we condition on X_i. We have $\mathrm{Var}[\mathbf{W}_{ij} \,|\, \mathbf{K}] \le 2p$, so conditioned on $\mathbf{K}$, Bernstein's inequality implies

$$|S| \lesssim \sqrt{p\left(\sum_{j\in[k]\setminus\{i\}} B_j^2\right)\log\left(\frac{1}{\delta}\right)} + \left(\max_{j\in[k]\setminus\{i\}}|B_j|\right)\log\left(\frac{1}{\delta}\right)$$

with probability at least $1-\delta/3$. In the remainder, we will show that the events

$$\sum_{j\in[k]\setminus\{i\}} B_j^2 \lesssim \frac{k^3}{d^3}+\frac{k^{5/2}}{d^3}\sqrt{\log\left(\frac{1}{\delta}\right)}\,, \qquad \max_{j\in[k]\setminus\{i\}} |B_j| \lesssim \frac{k}{d}\,, \tag{42}$$

occur with probability $1-2\delta/3$. It is easy to check that the statement of the lemma follows from the above bound on $|S|$, (42), and the assumption $k \geq d+\log(1/\delta)$.

If we define the quantities

$$Y := \sum_{l=1}^{k} \mathbf{K}_{il} X_l\,, \qquad Z := (\mathbf{K}_{ij})_{j\in[k]\setminus\{i\}}\,, \qquad \mathbf{G} := \sum_{l\in[k]\setminus\{i\}} X_l X_l^\top\,,$$

then $B_j = \langle X_j, Y\rangle - \mathbf{K}_{ij}$ and

$$\begin{aligned}\sum_{j\in[k]\setminus\{i\}} B_j^2 &= Y^\top \mathbf{G} Y - 2\|Y\|^2 + \|Z\|^2 \\ &\leq \|\mathbf{G}\|\|Y\|^2 + \|Z\|^2 \leq \|\mathbf{G}\|^2\|Z\|^2 + \|Z\|^2\,.\end{aligned} \tag{43}$$

To see the last inequality in (43), let $\mathbf{H} \in \mathbb{R}^{(k-1)\times d}$ have rows $(X_l^\top)_{l\in[k]\setminus\{i\}}$, and notice

$$\|Y\|^2 = \sum_{l,m=1}^{k} \mathbf{K}_{il}\mathbf{K}_{im}\langle X_l, X_m\rangle = Z^\top(\mathbf{H}\mathbf{H}^\top)Z \leq \|\mathbf{H}\mathbf{H}^\top\|\|Z\|^2 = \|\mathbf{G}\|\|Z\|^2\,,$$

since $\|\mathbf{H}\mathbf{H}^\top\| = \|\mathbf{H}^\top\mathbf{H}\| = \|\mathbf{G}\|$. Using that $\mathbb{E}\mathbf{G} = \frac{k-1}{d} I_d$, concentration of the sample covariance matrix (e.g., Theorem 5.7 in Rigollet and Hütter (2023)) implies the event

$$\|\mathbf{G}\| \lesssim \frac{k}{d} + \frac{k}{d}\left(\sqrt{\frac{d+\log(1/\delta)}{k}} + \frac{d+\log(1/\delta)}{k}\right) \asymp \frac{k}{d}$$

occurs with probability at least $1-\delta/6$, where we used $k \geq d+\log(1/\delta)$. Since $\mathbb{E}[\mathbf{K}_{ij}^2] = 1/d$ and $\|\mathbf{K}_{ij}^2 - 1/d\|_{\psi_1} \lesssim 1/d$, the subexponential Bernstein inequality implies

$$\|Z\|^2 \lesssim \frac{k}{d} + \frac{1}{d}\sqrt{k\log\left(\frac{1}{\delta}\right)} + \frac{1}{d}\log\left(\frac{1}{\delta}\right) \tag{44}$$

with probability at least $1-\delta/6$. Continuing from (43) and using our bounds on $\|\mathbf{G}\|$ and $\|Z\|^2$, and dropping redundant terms (using the assumption $k \geq d+\log(1/\delta)$), we find that the event

$$\sum_{j\in[k]\setminus\{i\}} B_j^2 \lesssim \left(\frac{k}{d}\right)^2\left(\frac{k}{d} + \frac{1}{d}\sqrt{k\log\left(\frac{1}{\delta}\right)}\right) = \frac{k^3}{d^3} + \frac{k^{5/2}}{d^3}\sqrt{\log\left(\frac{1}{\delta}\right)}$$

occurs with probability at least $1-\delta/3$, which yields the first bound in (42).

We now prove the bound on $\max_j |B_j|$. We have $B_j = \sum_{l=1}^k \mathbf{K}_{il}\mathbf{K}_{jl}$. Conditional on X_i and X_j, the summands $\mathbf{K}_{il}\mathbf{K}_{jl}$ have mean $\mathbf{K}_{ij}/d$ and ψ_1-norm $\lesssim 1/d$. Hence by the subexponential Bernstein inequality and a union bound, the event

$$\max_{j\in[k]\setminus\{i\}} |B_j| \lesssim \frac{k}{d} + \frac{1}{d}\left(\sqrt{k\log\left(\frac{k}{\delta}\right)} + \log\left(\frac{k}{\delta}\right)\right) \lesssim \frac{k}{d}$$

occurs with probability at least $1-\delta/3$, where we again used $k \geq d + \log(1/\delta)$. ■

Lemma C.4 *There is an absolute constant $C > 0$ such that if (A2) holds, then we have the following. Assume $k \geq d + \log(1/\delta)$. For all $1 \leq i \leq k$ and $\delta \in (0,1)$,*

$$\left|\sum_{j=1}^k (\mathbf{W}^2)_{ij}\mathbf{K}_{ij}\right| \leq C\left(p\sqrt{\frac{nk}{d}}\log\left(\frac{n}{\delta}\right) + \sqrt{\frac{kp}{d}}\log^2\left(\frac{n}{\delta}\right)\right)$$

with probability at least $1-\delta$.

Proof Fix $1 \leq i \leq k$. Let $\mathbf{W}'$ be a conditionally independent copy of $\mathbf{W}$ given $\mathbf{K}$. Define the sums

$$S := \sum_{j=1}^k\sum_{l=1}^n \mathbf{W}_{il}\mathbf{W}_{jl}\mathbf{K}_{ij}\,, \qquad S' := \sum_{j=1}^k\sum_{l=1}^n \mathbf{W}_{il}\mathbf{W}'_{jl}\mathbf{K}_{ij}\,. \tag{45}$$

For all $l \in [n]\setminus\{i\}$ let $B_l := \sum_{j=1}^k \mathbf{W}'_{jl}\mathbf{K}_{ij}$, so that $S' = \sum_{l=1}^n \mathbf{W}_{il}B_l$. We claim that there exists a universal c such that, for all $t > 0$, we have the decoupling inequality

$$\mathbb{P}(|S| \geq t) \leq c\mathbb{P}(c|S'| \geq t). \tag{46}$$

This will follow from Theorem 3.4.1 in de la Peña and Giné (2012), applied conditionally on $\mathbf{K}$, if we can write S and S' as sums of functions of independent random variables. As in the proof of Lemma B.2, when $\{i,j\} \in \binom{[n]}{2}$, we write $\mathbf{W}_{\{i,j\}}$ for the common value $\mathbf{W}_{ij} = \mathbf{W}_{ji}$. These are conditionally independent given $\mathbf{K}$, and take values in $\Omega = [-1,1]$. Let $J_n \subset \binom{[n]}{2}^2$ be the set of ordered pairs of *distinct* elements. Say that $(\{a,b\},\{c,d\}) \in J_2$ is "good" if $\{a,b\}\cap\{c,d\} \neq \emptyset$ and $i \in \{a,b\}\setminus\{c,d\}$. Such pairs must actually be of the form $(\{a,b\},\{c,d\}) = (\{i,\ell\},\{j,\ell\})$ for some $j,\ell \in [n]\setminus\{i\}$. Write J_n^{good} for the set of all such pairs, and $\eta : J_n^{\text{good}} \to [n]\setminus\{i\}$ for the function which "extracts j" in the sense of returning the element which is neither i nor shared. This is all well-defined, and it allows us to define $h_{\{a,b\},\{c,d\}} : \Omega^2 \to \mathbb{R}$ by

$$h_{\{a,b\},\{c,d\}}(x,y) = xy\mathbf{1}\{(\{a,b\},\{c,d\}) \text{ is good}\}(\mathbf{K}^{\#})_{i,\eta(\{a,b\},\{c,d\})}.$$

With this definition, one can check that

$$S = \sum_{j,\ell=1}^n \mathbf{W}_{i\ell}\mathbf{W}_{j\ell}(\mathbf{K}^{\#})_{ij} = \sum_{(\{a,b\},\{c,d\})\in J_n} h_{\{a,b\},\{c,d\}}(\mathbf{W}_{ab},\mathbf{W}_{cd})$$

$$S' = \sum_{j,\ell=1}^n \mathbf{W}_{i\ell}\mathbf{W}_{j\ell}(\mathbf{K}^{\#})_{ij} = \sum_{(\{a,b\},\{c,d\})\in J_n} h_{\{a,b\},\{c,d\}}(\mathbf{W}_{ab},\mathbf{W}'_{cd}),$$

so that Theorem 3.4.1 of de la Peña and Giné (2012) indeed implies (46).

Conditional on $\mathbf{K}$ and $\mathbf{W}'$, the random variables $\{B_l \mathbf{W}_{il}\}_{l \in [n] \setminus \{i\}}$ are independent, centered, bounded in absolute value by $|B_l|$, and have variance at most $2pB_l^2$. Hence Bernstein's inequality implies

$$|S'| \lesssim \sqrt{p\left(\sum_{l \in [n] \setminus \{i\}} B_l^2\right) \log\left(\frac{1}{\delta}\right)} + \left(\max_{l \in [n] \setminus \{i\}} |B_l|\right) \log\left(\frac{1}{\delta}\right) \tag{47}$$

with conditional probability at least $1 - \delta/(3c)$.

We first bound the variance term. For each $l \in [n] \setminus \{i\}$, let us define $Y_l := \sum_{j=1}^k \mathbf{W}'_{jl} X_j$ so that $B_l = X_i^\top Y_l$. We then calculate that

$$\sum_{l \in [n] \setminus \{i\}} B_l^2 \leq X_i^\top \mathbf{X}^\top \mathbf{W}'_{1:k,:} (\mathbf{W}'_{1:k,:})^\top \mathbf{X} X_i \leq \|\mathbf{X}\|^2 \|\mathbf{W}'\|^2 .$$

By concentration of the sample covariance matrix, with probability at least $1 - \delta/(6c)$,

$$\|\mathbf{X}\|^2 = \|\mathbf{X}^\top \mathbf{X}\| \lesssim \frac{k}{d} + \frac{1}{d} \log\left(\frac{1}{\delta}\right) \lesssim \frac{k}{d}$$

where we used $k \geq d + \log(1/\delta)$.

Conditional on $\mathbf{K}$, the entries of $\mathbf{W}'$ are independent, centered, bounded by 1, and have variance at most $2p$. Applying Lemma C.6 to $\mathbf{W}'$ with $\delta/(6c)$ in place of δ, we obtain

$$\|\mathbf{W}'\| \lesssim \sqrt{np} + \sqrt{\log\left(\frac{n}{\delta}\right)}$$

with conditional probability at least $1 - \delta/(6c)$ given $\mathbf{K}$. Combining these bounds, we obtain

$$\sum_{l \in [n] \setminus \{i\}} B_l^2 \lesssim \|\mathbf{X}\|^2 \|\mathbf{W}'\|^2 \lesssim \frac{k}{d} \cdot \left(np + \log\left(\frac{n}{\delta}\right)\right) \tag{48}$$

with probability at least $1 - \delta/(3c)$.

To control the second term in (47), if we condition on $\mathbf{K}$, Bernstein's inequality and a union bound yield

$$\max_{l \in [n] \setminus \{i\}} |B_l| \lesssim \sqrt{p\left(\sum_{j \in [k] \setminus \{i\}} \mathbf{K}_{ij}^2\right) \log\left(\frac{n}{\delta}\right)} + \log\left(\frac{n}{\delta}\right)$$

with probability at least $1 - \delta/(6c)$. The bound from (44) applies directly to $\sum_{j \in [k] \setminus \{i\}} \mathbf{K}_{ij}^2$ with probability at least $1 - \delta/(6c)$, so we obtain

$$\max_{l \in [n] \setminus \{i\}} |B_l| \lesssim \sqrt{\frac{pk}{d}} \log\left(\frac{n}{\delta}\right) + \log\left(\frac{n}{\delta}\right)$$

with probability at least $1 - \delta/(3c)$, where we used $k \geq d + \log(1/\delta)$. By combining the bounds for the variance and max terms, using the decoupling inequality, and $k \geq d + \log(1/\delta)$, we obtain

$$p\sqrt{\frac{nk}{d}} \log\left(\frac{n}{\delta}\right) + \sqrt{\frac{pk}{d}} \log^{3/2}\left(\frac{1}{\delta}\right) + \sqrt{\frac{pk}{d}} \log^2\left(\frac{n}{\delta}\right) + \log^2\left(\frac{n}{\delta}\right) . \tag{49}$$

The second term in (49) is clearly at most the third term. The fourth term is at most the third term since $\sqrt{pk/d} \geq 1$, which in turn follows from (A2). Indeed, (A2) and $k \leq n$ imply $p \gtrsim n^{-2/3}$, hence we have $pk/d \gtrsim \sqrt{n}p^{1/4} \gtrsim n^{1/3} \geq 1$. This completes the proof of the lemma. ■

Lemma C.5 *There is an absolute constant $C > 0$ such that we have the following. Let $1 \leq i \leq k$ and $\delta \in (0, 1)$. If (A1), (A2), and $\delta \geq n^{-C}$ hold, then*

$$\left|\sum_{j=1}^{k}(\mathbf{W}^2)_{ij}\mathbf{W}_{ij}\right| \leq C\left(\sqrt{nk}p^{3/2}\sqrt{\log\left(\frac{1}{\delta}\right)} + p\sqrt{k}\log^{3/2}\left(\frac{n}{\delta}\right) + \log^2\left(\frac{n}{\delta}\right)\right)$$

with probability at least $1 - \delta$.

Proof Fix $1 \leq i \leq k$ and let

$$S := \sum_{j=1}^{k}\sum_{l=1}^{n}\mathbf{W}_{il}\mathbf{W}_{jl}\mathbf{W}_{ij}\,, \qquad S' := \sum_{j=1}^{k}\sum_{l=1}^{n}\mathbf{W}_{il}\mathbf{W}'_{lj}\mathbf{W}''_{ij}\,,$$

where $\mathbf{W}', \mathbf{W}'' \in \mathbb{R}^{n\times k}$ are conditionally independent copies of $\mathbf{W}$ given $\mathbf{K}$. We claim that, for a universal constant $c > 0$ and all $t > 0$,

$$\mathbb{P}\{|S| \geq t\} \leq c\,\mathbb{P}\{c|S'| \geq t\}.$$

Conditional on $\mathbf{K}$, the random variables $\{\mathbf{W}_{\{a,b\}}\}_{\{a,b\}\in\binom{[n]}{2}}$ are independent. Let $J_n \subset \binom{[n]}{2}^3$ be the set of ordered triples of distinct unordered pairs. For $(\{a,b\},\{c,d\},\{e,f\}) \in J_n$, define

$$\begin{aligned} h_{\{a,b\},\{c,d\},\{e,f\}}(x,y,z) = xyz\cdot \mathbf{1}\{&\{a,b\},\{c,d\},\{e,f\}\text{ form a triangle}\} \\ &\cdot \mathbf{1}\{i \in \{a,b\}\cap\{e,f\}\}\cdot\mathbf{1}\{\{e,f\} = \{i,j\}\text{ for some } j\in[k]\}. \end{aligned}$$

The nonzero terms are exactly those with $(\{a,b\},\{c,d\},\{e,f\}) = (\{i,l\},\{l,j\},\{i,j\})$ for some $j \in [k]$ and $l \in [n]\setminus\{i,j\}$. Therefore, using the convention that the diagonal of $\mathbf{W}$ vanishes,

$$\sum_{(\{a,b\},\{c,d\},\{e,f\})\in J_n} h_{\{a,b\},\{c,d\},\{e,f\}}(\mathbf{W}_{ab},\mathbf{W}_{cd},\mathbf{W}_{ef}) = \sum_{j=1}^{k}\sum_{l=1}^{n}\mathbf{W}_{il}\mathbf{W}_{lj}\mathbf{W}_{ij} = S,$$

and similarly

$$\sum_{(\{a,b\},\{c,d\},\{e,f\})\in J_n} h_{\{a,b\},\{c,d\},\{e,f\}}(\mathbf{W}_{ab},\mathbf{W}'_{cd},\mathbf{W}''_{ef}) = \sum_{j=1}^{k}\sum_{l=1}^{n}\mathbf{W}_{il}\mathbf{W}'_{lj}\mathbf{W}''_{ij} = S'.$$

Theorem 3.4.1 in de la Peña and Giné (2012), applied conditionally on $\mathbf{K}$ to the independent variables indexed by $\binom{[n]}{2}$, gives the desired conditional decoupling inequality. Taking expectations over $\mathbf{K}$ gives the claimed unconditional version.

Let $B_l := \sum_{j=1}^k \mathbf{W}'_{lj}\mathbf{W}''_{ij}$ for $l \in [n]$, so that $S' = \sum_{l\in[n]} B_l \mathbf{W}_{il}$. Conditional on $\mathbf{K}$, $\mathbf{W}'$, and $\mathbf{W}''$, the variables $\{B_l \mathbf{W}_{il}\}_{l\in[n]\setminus\{i\}}$ are independent, centered, bounded in absolute value by $|B_l|$, and have variance at most $2pB_l^2$. From Bernstein's inequality,

$$|S'| \lesssim \sqrt{p\left(\sum_{l\in[n]\setminus\{i\}} B_l^2\right)\log\left(\frac{1}{\delta}\right)} + \left(\max_{l\in[n]\setminus\{i\}} |B_l|\right)\log\left(\frac{1}{\delta}\right) \tag{50}$$

with probability at least $1 - \delta/(3c)$.

Let us write $B := (B_j)_{j\in[n]\setminus\{i\}}$ and $V := \|B\|^2$. Define the random vector $Y := (\mathbf{W}''_{ij})_{j\in[k]}$ with $Y_i := 0$. We then have

$$V = \|\mathbf{W}'Y\|^2 - B_i^2 \leq \|\mathbf{W}'\|^2\|Y\|^2 .$$

From Lemma C.6 below, we have the bound $\|\mathbf{W}'\| \lesssim \sqrt{np}$ with conditional probability at least $1 - \delta/(6c)$ given $\mathbf{K}$, where the $\sqrt{\log(n/\delta)}$ term is $O(\sqrt{np})$ due to (A1). Conditional on $\mathbf{K}$, Bernstein's inequality implies

$$\|Y\|^2 \lesssim kp + \sqrt{kp\log\left(\frac{1}{\delta}\right)} + \log\left(\frac{1}{\delta}\right) \lesssim kp + \log\left(\frac{1}{\delta}\right)$$

with probability at least $1-\delta/(6c)$. The fourth bullet point in the proof of Proposition 3.2 shows that $\log(n) = o(kp)$ when (A2) holds; thus, since $\delta \geq n^{-C}$, we have that $\|Y\|^2 \lesssim kp$ with probability at least $1 - \delta/(6c)$. It then follows that

$$V \lesssim nkp^2$$

with probability at least $1 - \delta/(3c)$.

We now bound the second term in (50). Conditional on $\mathbf{K}$, the summands $\mathbf{W}'_{lj}\mathbf{W}''_{ij}$ in B_l are centered, bounded in absolute value by 1, and have variance at most $4p^2$, hence Bernstein's inequality and a union bound imply

$$\max_{l\in[n]\setminus\{i\}} |B_l| \lesssim \sqrt{kp^2\log\left(\frac{n}{\delta}\right)} + \log\left(\frac{n}{\delta}\right)$$

with probability at least $1 - \delta/(3c)$. Combining the above bounds and taking a union bound, we have

$$|S'| \lesssim \sqrt{nk}p^{3/2}\sqrt{\log\left(\frac{1}{\delta}\right)} + \sqrt{k}p\log^{3/2}\left(\frac{n}{\delta}\right) + \log^2\left(\frac{n}{\delta}\right)$$

with probability at least $1 - \delta/c$. The lemma now follows by applying the decoupling inequality cited above to S and S'. ■

Lemma C.6 *There is an absolute constant $C > 0$ such that the following holds. For all $\delta \in (0, 1)$,*

$$\|\mathbf{W}\| \leq C\left(\sqrt{np} + \sqrt{\log\left(\frac{n}{\delta}\right)}\right)$$

with conditional probability at least $1 - \delta$ given $\mathbf{K}$.

Proof Condition on $\mathbf{K}$ throughout the proof. If we define the quantities

$$\sigma := \max_{i=1,\dots,n} \sqrt{\sum_{j=1}^{n} \mathbb{E}[\mathbf{W}_{ij}^2 \,|\, \mathbf{K}]}\,, \qquad \sigma_* := \max_{i,j=1,\dots,n} \|\mathbf{W}_{ij}\|_\infty\,,$$

then $\sigma_* \leq 1$ and $\sigma^2 \leq 2np$, since $\mathbb{E}[\mathbf{W}_{ij}^2 \,|\, \mathbf{K}] \leq 2p$. The hypotheses of Theorem F.1 are satisfied for $\mathbf{W}$ since it is symmetric with independent centered entries above the diagonal. Hence applying Theorem F.1 and taking $t \asymp \sqrt{\log(n/\delta)}$ implies the event

$$\|\mathbf{W}\| \lesssim \sigma + t \lesssim \sqrt{np} + \sqrt{\log\left(\frac{n}{\delta}\right)}$$

occurs with probability at least $1 - \delta$. ■

Lemma C.7 *There is an absolute constant $C > 0$ such that the following holds. Assume $k \geq d + \log(1/\delta)$. For all $1 \leq i \leq k$ and $\delta \in (0, 1)$,*

$$\left|\sum_{j=1}^{k} (\mathbf{K}^{\#}\mathbf{W} + \mathbf{W}\mathbf{K}^{\#})_{ij}\mathbf{K}_{ij}\right| \leq C\left(\sqrt{\frac{pk^3}{d^3}\log\left(\frac{1}{\delta}\right)} + \frac{k}{d}\log\left(\frac{1}{\delta}\right)\right)$$

with probability at least $1 - \delta$.

Proof Fix $1 \leq i \leq k$. Define the sums

$$S := \sum_{j=1}^{k} (\mathbf{K}^{\#}\mathbf{W} + \mathbf{W}\mathbf{K}^{\#})_{ij}\mathbf{K}_{ij}\,, \quad S_1 := \sum_{1\leq j<l\leq k} \mathbf{K}_{il}\mathbf{W}_{jl}\mathbf{K}_{ij}\,, \quad S_2 := \sum_{j,l=1}^{k} \mathbf{W}_{il}\mathbf{K}_{jl}\mathbf{K}_{ij}\,,$$

so that $S = 2S_1 + S_2$. Notice $|S_2|$ is precisely the quantity analyzed in Lemma C.3 (applied with $\delta/2$), so it remains to bound $|S_1|$. Conditional on $\mathbf{K}$, the summands in S_1 are independent, mean-zero, and variance at most $2p\mathbf{K}_{il}^2\mathbf{K}_{ij}^2$. Hence by applying Bernstein's inequality given $\mathbf{K}$, we have that with probability at least $1 - \delta/2$,

$$\begin{aligned}
|S_1| &\lesssim \sqrt{p\Big(\sum_{1\leq j<l\leq k} \mathbf{K}_{ij}^2\mathbf{K}_{il}^2\Big)\log\left(\frac{1}{\delta}\right)} + \log\left(\frac{1}{\delta}\right) \\
&\lesssim \left(\frac{k}{d} + \frac{1}{d}\sqrt{k\log\left(\frac{1}{\delta}\right)} + \frac{1}{d}\log\left(\frac{1}{\delta}\right)\right)\sqrt{p\log\left(\frac{1}{\delta}\right)} + \log\left(\frac{1}{\delta}\right) \\
&\lesssim \frac{k\sqrt{p}}{d}\sqrt{\log\left(\frac{1}{\delta}\right)} + \log\left(\frac{1}{\delta}\right)\,,
\end{aligned}$$

where in the second inequality we used $\sum_{j<l}\mathbf{K}_{ij}^2\mathbf{K}_{il}^2 \leq (\sum_j \mathbf{K}_{ij}^2)^2$ and applied (44) to the right-hand side, and in the third inequality we used $k \geq d + \log(1/\delta)$. Combining the bounds for $|S_1|$ and $|S_2|$, and dropping redundant terms, implies the statement of the lemma. ■

Lemma C.8 *There is an absolute constant $C > 0$ such that the following holds. Assume (A2) and $k \geq d + \log(1/\delta)$. For all $1 \leq i \leq k$ and $\delta \in (0,1)$,*

$$\left| \sum_{j=1}^{k} (\mathbf{K}^{\#}\mathbf{W} + \mathbf{W}\mathbf{K}^{\#})_{ij} \mathbf{W}_{ij} \right| \leq C \left(\frac{pk}{\sqrt{d}} \log\left(\frac{k}{\delta}\right) + \sqrt{\frac{pk}{d}} \log^2\left(\frac{k}{\delta}\right) \right)$$

with probability at least $1 - \delta$.

Proof Fix $1 \leq i \leq k$. Define the sums

$$S := \sum_{j=1}^{k} (\mathbf{K}^{\#}\mathbf{W} + \mathbf{W}\mathbf{K}^{\#})_{ij} \mathbf{W}_{ij}\,, \quad S_1 := \sum_{j,l=1}^{k} \mathbf{K}_{il} \mathbf{W}_{jl} \mathbf{W}_{ij}\,, \quad S_2 := \sum_{j,l=1}^{k} \mathbf{W}_{il} \mathbf{K}_{jl} \mathbf{W}_{ij}\,,$$

so that $S = S_1 + S_2$. First note that S_1 is precisely S as defined in (45), except with the sum over $l \in [k]$ instead of $j \in [n]$. Hence it follows from Lemma C.4 (applied with $\delta/2$ and $n = k$) that

$$|S_1| \lesssim \frac{pk}{\sqrt{d}} \log\left(\frac{k}{\delta}\right) + \sqrt{\frac{pk}{d}} \log^2\left(\frac{k}{\delta}\right)$$

with probability at least $1 - \delta/2$.

We will now bound $|S_2|$. Let $c > 0$ be the constant from the decoupling inequality for order-2 kernels (Theorem 3.4.1 in de la Peña and Giné (2012)). Let $\mathbf{W}'$ be an independent copy of $\mathbf{W}$ (that is, $\mathbf{W}$ and $\mathbf{W}'$ are conditionally i.i.d. given $\mathbf{K}$). Let $B_l := \sum_{j=1}^{k} \mathbf{W}'_{ij} \mathbf{K}_{jl}$ for all $l \in [k] \setminus \{i\}$, and define $S_2' := \sum_{l=1}^{k} B_l \mathbf{W}_{il}$. Conditional on $\mathbf{K}$ and $\mathbf{W}'$, Bernstein's inequality implies

$$|S_2'| \lesssim \sqrt{p \Bigg(\sum_{l \in [k] \setminus \{i\}} B_l^2 \Bigg) \log\left(\frac{1}{\delta}\right)} + \Big(\max_{l \in [k] \setminus \{i\}} |B_l| \Big) \log\left(\frac{1}{\delta}\right)$$

with probability at least $1 - \delta/(6c)$.

Let $Y := \sum_{j=1}^{k} \mathbf{W}'_{ij} X_j$, and let $\mathbf{G} \in \mathbb{R}^{d \times d}$ and $\mathbf{H} \in \mathbb{R}^{(k-1) \times d}$ be defined as in Lemma C.3. Also let $Z := (\mathbf{W}'_{ij})_{j \in [k] \setminus \{i\}}$. Notice that $B_l = \langle Y, X_l \rangle - \mathbf{W}'_{il}$, which implies

$$\begin{aligned}\sum_{l \in [k] \setminus \{i\}} B_l^2 &\leq 2 Y^\top \left(\sum_{l \in [k] \setminus \{i\}} X_l X_l^\top \right) Y + 2 \sum_{l \in [k] \setminus \{i\}} (\mathbf{W}'_{il})^2 \\ &= 2 Y^\top \mathbf{G} Y + 2\|Z\|^2 \leq 2\|\mathbf{G}\| \|Y\|^2 + 2\|Z\|^2\,.\end{aligned}$$

From Lemma C.3 we have $\|\mathbf{G}\| \lesssim \frac{k}{d}$ with probability at least $1 - \delta/(18c)$. Conditional on $\mathbf{X}$, the random vectors $V_j := \mathbf{W}'_{ij} X_j$ for $j \in [k] \setminus \{i\}$ are independent, mean-zero, and satisfy $\|V_j\| \leq 1$ and $\mathbb{E}[\|V_j\|^2 \,|\, \mathbf{X}] \leq 2p$. Thus from the vector Bernstein inequality (e.g. (Tropp, 2012, Theorem 1.6)),

$$\|Y\| \lesssim \sqrt{kp \log\left(\frac{d}{\delta}\right)} + \log\left(\frac{d}{\delta}\right)$$

with probability at least $1 - \delta/(18c)$. The scalar Bernstein inequality implies

$$\|Z\|^2 \lesssim kp + \sqrt{kp \log\left(\frac{1}{\delta}\right)} + \log\left(\frac{1}{\delta}\right)$$

with probability at least $1 - \delta/(18c)$. Combining the above three bounds, the event

$$\sum_{l \in [k] \setminus \{i\}} B_l^2 \lesssim \frac{k^2 p}{d} \log\left(\frac{d}{\delta}\right) + \frac{k}{d} \log^2\left(\frac{d}{\delta}\right)$$

occurs with probability at least $1 - \delta/(6c)$.

For the maximum term, conditional on $\mathbf{K}$, Bernstein's inequality gives

$$|B_l| \lesssim \sqrt{p \left(\sum_{j \in [k] \setminus \{i\}} \mathbf{K}_{jl}^2\right) \log\left(\frac{k}{\delta}\right)} + \log\left(\frac{k}{\delta}\right)$$

for all $l \in [k] \setminus \{i\}$ with conditional probability at least $1 - \delta/(12c)$. Further, applying (44) with failure probability $\delta/(12ck)$ and taking a union bound over $l \in [k] \setminus \{i\}$ gives

$$\max_{l \in [k] \setminus \{i\}} \sum_{j \in [k] \setminus \{i\}} \mathbf{K}_{jl}^2 \lesssim \frac{k}{d} + \frac{1}{d}\sqrt{k \log\left(\frac{k}{\delta}\right)} + \frac{1}{d} \log\left(\frac{k}{\delta}\right) \lesssim \frac{k}{d}$$

with probability at least $1 - \delta/(12c)$, where the last inequality uses $k \geq d + \log(1/\delta)$. Combining the last two displays and absorbing constants, we obtain

$$\max_{l \in [k] \setminus \{i\}} |B_l| \lesssim \sqrt{\frac{pk}{d} \log\left(\frac{k}{\delta}\right)} + \log\left(\frac{k}{\delta}\right) \lesssim \sqrt{\frac{pk}{d}} \log\left(\frac{k}{\delta}\right) + \log\left(\frac{k}{\delta}\right)$$

with probability at least $1 - \delta/(6c)$.

Combining the above bounds and using $k \geq d + \log(1/\delta)$ implies

$$|S_2'| \lesssim \frac{kp}{\sqrt{d}} \log\left(\frac{d}{\delta}\right) + \sqrt{\frac{pk}{d}} \log^2\left(\frac{k}{\delta}\right) + \log^2\left(\frac{k}{\delta}\right),$$

with probability at least $1 - \delta/(2c)$. By the decoupling inequality, the same bound holds for $|S_2|$ with probability at least $1 - \delta/2$. It remains to absorb the third term into the second on the right-hand side using the same argument as at the end of the proof of Lemma C.4 together with (A2). ■

Appendix D. Computational lower bound

We prove Theorem 2.7 in this section following Schramm and Wein (2022). Before stating the proof, we need several definitions. Let $\mathbf{B} \sim G(n, p)$ coupled to $\mathbf{A}$ such that for all distinct pairs $1 \leq i, j \leq n$ not both in $\mathcal{S}$, we have $\mathbf{A}_{ij} = \mathbf{B}_{ij}$, and the collections $\{\mathbf{A}\}_{i,j \in [n]}$ and $\{\mathbf{B}\}_{i,j \in \mathcal{S}}$

are independent. For a subset $\alpha \subseteq \binom{[n]}{2}$ (which we identify with the binary vector with entries $\mathbf{1}\{e \in \alpha\}$, $e \in \binom{[n]}{2}$) and an $n \times n$ matrix $\mathbf{Y}$, define

$$\mathbf{Y}^{\alpha} := \prod_{ij \in \alpha} \mathbf{Y}_{ij} \, .$$

We will also view α as a graph whose vertex set, denoted $V(\alpha)$, is induced by its edges. With these definitions, the following calculation is the correct modification of the top of p. 17 in the technical appendix of Schramm and Wein (2022):

$$\begin{aligned}\mathbb{E}[\bar{\mathbf{A}}^{\alpha} \,|\, \mathbf{B}] &= \mathbb{E}_{\mathcal{S}}\Big[\bar{\mathbf{B}}^{\alpha \setminus \binom{\mathcal{S}}{2}} \cdot \mathbb{E}\Big[\bar{\mathbf{A}}^{\alpha \cap \binom{\mathcal{S}}{2}} \,\Big|\, \mathcal{S}\Big] \,\Big|\, \mathbf{B}\Big] \\ &= \sum_{\beta \subseteq \alpha} \mathbb{E}\Big[\mathbf{1}\{\alpha \setminus \tbinom{\mathcal{S}}{2} = \beta\} \cdot \bar{\mathbf{A}}^{\alpha \cap \binom{\mathcal{S}}{2}}\Big] \cdot \bar{\mathbf{B}}^{\beta} = \sum_{\beta \subseteq \alpha} M_{\alpha\beta} \bar{\mathbf{B}}^{\beta} \, ,\end{aligned}$$

where

$$M_{\alpha\beta} := \mathbb{E}\left[\mathbf{1}\big\{\alpha \setminus \tbinom{\mathcal{S}}{2} = \beta\big\} \cdot \bar{\mathbf{A}}^{\alpha \cap \binom{\mathcal{S}}{2}}\right] \, .$$

The following lemma now follows from the proof of Theorem 2.2 of Schramm and Wein (2022) (see also p. 17 of the technical appendix).

Lemma D.1 (Schramm and Wein (2022)) *Let $\mathbf{A}$ and $\mathbf{B}$ be as defined above. Suppose $M_{\alpha\alpha} \neq 0$ for all $\alpha \subseteq \binom{[n]}{2}$ with $|\alpha| \leq D$. Then*

$$\mathsf{Corr}^2_{\leqslant D} \leq \sum_{\alpha \subseteq \binom{[n]}{2},\, |\alpha| \leq D} w_{\alpha}^2 \, ,$$

where w_{α} are defined recursively as

$$w_{\emptyset} = \mathbb{E}[\theta] \, , \quad w_{\alpha} := \frac{1}{M_{\alpha\alpha}} \left(\mathbb{E}[\bar{\mathbf{A}}^{\alpha} \cdot \theta] - \sum_{\beta \subsetneq \alpha} M_{\alpha\beta} w_{\beta} \right) \, , \quad \alpha \subseteq \tbinom{[n]}{2} \, , \quad |\alpha| \geq 1 \, ,$$

where $\beta \subsetneq \alpha$ means β is a proper subgraph of α.

Proof of Theorem 2.7 The proof of $\mathsf{Corr}^2_{\leqslant D} \leq (1 + o(1)) r^2$ is nearly the same as that of Lemma H.2(*i*) in the technical appendix of Schramm and Wein (2022), so we only describe the parts that are different. We only need to confirm that Lemma H.4 of Schramm and Wein (2022) still holds in our setting, for which it suffices to prove the same bounds on $M_{\alpha\alpha}$, $|M_{\alpha\beta}|$, and $|\mathbb{E}[\bar{\mathbf{A}}^{\alpha} \cdot \theta]|$. Note that Lemma H.3 in their article holds true unchanged in our setting.

For all $\alpha \subseteq \binom{[n]}{2}$ we have

$$\begin{aligned}M_{\alpha\alpha} &= \mathbb{E}\left[\mathbf{1}\big\{\alpha \setminus \tbinom{\mathcal{S}}{2} = \alpha\big\} \cdot \bar{\mathbf{A}}^{\alpha \cap \binom{\mathcal{S}}{2}}\right] \\ &= \mathbb{P}\big\{\alpha \setminus \tbinom{\mathcal{S}}{2} = \alpha\big\} \geq \mathbb{P}\{V(\alpha) \cap \mathcal{S} = \emptyset\} = (1 - r)^{|V(\alpha)|} \, .\end{aligned}$$

For fixed $\beta \subseteq \alpha$, since $\alpha \setminus \binom{\mathcal{S}}{2} = \beta$ implies $V(\alpha \setminus \beta) \subseteq \mathcal{S}$, we calculate

$$\begin{aligned}
|M_{\alpha\beta}| &= \left|\mathbb{E}_{\mathcal{S}}\left[\mathbf{1}\{\alpha \setminus \tbinom{\mathcal{S}}{2} = \beta\} \cdot \mathbb{E}\left[\bar{\mathbf{A}}^{\alpha \cap \binom{\mathcal{S}}{2}} \,\middle|\, \mathcal{S}\right]\right]\right| \\
&= \mathbb{P}\big\{\alpha \setminus \tbinom{\mathcal{S}}{2} = \beta\big\} \cdot \left|\mathbb{E}\left[\bar{\mathbf{A}}^{\alpha} \,|\, V(\alpha) \subseteq \mathcal{S}\right]\right| \\
&= \mathbb{P}\big\{\alpha \setminus \tbinom{\mathcal{S}}{2} = \beta\big\} \cdot \left|\mathbb{E}\left[\prod_{ij \in \alpha} \frac{\mathbf{A}_{ij} - p}{\sqrt{p(1-p)}}\right]\right| \\
&= \mathbb{P}\big\{\alpha \setminus \tbinom{\mathcal{S}}{2} = \beta\big\} \cdot \left(\frac{p}{1-p}\right)^{|\alpha|/2} \left|\mathbb{E}\left[\prod_{ij \in \alpha} \langle X_i, X_j \rangle\right]\right| \\
&\le \mathbb{P}\big\{\alpha \setminus \tbinom{\mathcal{S}}{2} = \beta\big\} \le r^{|V(\alpha \setminus \beta)|}.
\end{aligned}$$

Finally we bound $|\mathbb{E}[\bar{\mathbf{A}}^{\alpha} \cdot \theta]|$. Let $X_1, \ldots, X_{|V(\alpha)|}$ be i.i.d. uniformly distributed on the sphere $\mathbb{S}^{d-1}$. Since $\mathbb{E}\left[\bar{\mathbf{A}}^{\alpha} \,|\, \mathcal{S}\right] \cdot \theta \neq 0$ only if $V(\alpha) \cup \{1\} \subseteq \mathcal{S}$, we have

$$\begin{aligned}
\mathbb{E}[\bar{\mathbf{A}}^{\alpha} \cdot \theta] &= \mathbb{E}_{\mathcal{S}}\left[\mathbb{E}\left[\bar{\mathbf{A}}^{\alpha} \,|\, \mathcal{S}\right] \cdot \theta\right] \\
&= \mathbb{P}\{V(\alpha) \cup \{1\} \subseteq \mathcal{S}\} \cdot \mathbb{E}\left[\bar{\mathbf{A}}^{\alpha} \,|\, V(\alpha) \subseteq \mathcal{S}\right] \le r^{|V(\alpha) \cup \{1\}|} \le r^{|V(\alpha)|},
\end{aligned}$$

where the first inequality holds via the same calculation as for $|M_{\alpha\beta}|$. The above bounds on $M_{\alpha\alpha}$, $|M_{\alpha\beta}|$, and $|\mathbb{E}[\bar{\mathbf{A}}^{\alpha} \cdot \theta]|$ match those in Schramm and Wein (2022), so indeed Lemma H.4 holds in our setting, completing the proof. ■

Appendix E. Information-theoretic upper bound

We prove Theorem 2.8 in this section. For clarity, let $\mathcal{S}^*$ (instead of $\mathcal{S}$) denote the vertex set of the planted subgraph in $\mathbf{A}$ throughout the proof. If $k \ge cn$ for a constant $c > 0$ and we take $\epsilon \in (0, c^2)$ then $|\mathcal{S}^* \cap \hat{\mathcal{S}}| \ge \epsilon k$ with high probability (recall $\hat{\mathcal{S}}$ is defined as a random k-subset of $[n]$ when $k = \Theta(n)$). Thus in the remainder we may assume $k = o(n)$, where $\hat{\mathcal{S}}$ is defined by (7). For all subsets $A, B, C \subseteq [n]$ define

$$\tau(A, B, C) := \sum_{i \in A, j \in B, l \in C} \bar{\mathbf{A}}_{ij} \bar{\mathbf{A}}_{jl} \bar{\mathbf{A}}_{li}$$

and write $\tau(A) \equiv \tau(A, A, A)$. Moreover, define

$$\begin{aligned}
&T_1(A) := \tau(A, \mathcal{S}^* \setminus A, \mathcal{S}^* \setminus A), \quad && T_2(A) := \tau(A, A, \mathcal{S}^* \setminus A), \quad && T_3(A) := \tau(A), \\
&T_4(A, B) := \tau(B, \mathcal{S}^* \setminus A, \mathcal{S}^* \setminus A), \quad && T_5(A, B) := \tau(B, B, \mathcal{S}^* \setminus A), \quad && T_6(B) := \tau(B).
\end{aligned}$$

At times we will write $T_1, \ldots, T_6$ without their respective arguments to ease notation. Define

$$\Delta(A, B) := 3T_1(A) + 3T_2(A) + T_3(A) - 3T_4(A, B) - 3T_5(A, B) - T_6(B).$$

By definition of Δ, if $\mathcal{S} \in \binom{[n]}{k}$, $A = \mathcal{S}^* \setminus \mathcal{S}$, and $B = \mathcal{S} \setminus \mathcal{S}^*$, then $\mathcal{S}^* \setminus A = \mathcal{S}^* \cap \mathcal{S}$ and

$$\Delta(A, B) = \tau(\mathcal{S}^*) - \tau(\mathcal{S}).$$

Let $\epsilon > 0$ be a sufficiently small constant. The optimality of $\hat{\mathcal{S}}$ implies $\Delta(\mathcal{S}^* \setminus \hat{\mathcal{S}}, \hat{\mathcal{S}} \setminus \mathcal{S}^*)$ is nonpositive, hence

$$\begin{aligned}
\mathbb{P}\{|\mathcal{S}^* \cap \hat{\mathcal{S}}| \le \epsilon k\} &\le \mathbb{P}\left\{\tau(\mathcal{S}) \ge \tau(\mathcal{S}^*) \text{ for some } \mathcal{S} \in \binom{[n]}{k} \text{ with } |\mathcal{S}^* \cap \mathcal{S}| \le \epsilon k\right\} \\
&= \mathbb{P}\left\{\Delta(A,B) \le 0 \text{ for some } A \in \binom{\mathcal{S}^*}{\ge (1-\epsilon)k},\ B \in \binom{[n]\setminus \mathcal{S}^*}{|A|}\right\} \\
&\le \sum_{A \in \binom{\mathcal{S}^*}{\ge (1-\epsilon)k}} \mathbb{P}\left\{\Delta(A,B) \le 0 \text{ for some } B \in \binom{[n]\setminus\mathcal{S}^*}{|A|}\right\},
\end{aligned}$$

and let us write P_A to denote the summand on the right-hand side. Let $s := p^3k^3/2d^2$, and for all A define the event

$$E_A := \{T_1(A) \le -s/12 \text{ or } T_2(A) \le -s/12 \text{ or } T_3(A) \le s\}.$$

We then have

$$\begin{aligned}
P_A &\le \mathbb{P}\{E_A\} + \mathbb{P}\left\{\Delta(A,B) \le 0 \text{ for some } B \in \binom{[n]\setminus\mathcal{S}^*}{|A|} \text{ and } E_A^c \text{ holds}\right\} \\
&\le \sum_{i=1}^{2} \mathbb{P}\{T_i \le -s/12\} + \mathbb{P}\{T_3 \le s\} \\
&\quad + \mathbb{P}\left\{3T_4 + 3T_5 + T_6 \ge s/2 \text{ for some } B \in \binom{[n]\setminus\mathcal{S}^*}{|A|}\right\} \\
&\le \sum_{i=1}^{2} \mathbb{P}\{T_i \le -s/12\} + \mathbb{P}\{T_3 \le s\} + \sum_{i=4}^{6} \mathbb{P}\left\{T_i \ge s/14 \text{ for some } B \in \binom{[n]\setminus\mathcal{S}^*}{|A|}\right\}.
\end{aligned}$$

Summarizing, and applying another union bound, we write

$$\begin{aligned}
\mathbb{P}\{|\mathcal{S}^* \cap \hat{\mathcal{S}}| \le \epsilon k\} \le &\sum_{A \in \binom{\mathcal{S}^*}{\ge(1-\epsilon)k}} \left(\mathbb{P}\{T_1 \le -s/12\} + \mathbb{P}\{T_2 \le -s/12\} + \mathbb{P}\{T_3 \le s\}\right) \\
&+ \sum_{i=4}^{6} \sum_{A \in \binom{\mathcal{S}^*}{\ge(1-\epsilon)k}} \sum_{B \in \binom{[n]\setminus\mathcal{S}^*}{|A|}} \mathbb{P}\{T_i \ge s/14\}.
\end{aligned} \tag{51}$$

If $H(x) = -x\log_2 x - (1-x)\log_2(1-x)$ is the binary entropy then we have

$$\binom{k}{\ge (1-\epsilon)k} \le 2^{H(\epsilon)k} \qquad \text{and} \qquad \binom{k}{\ge(1-\epsilon)k}\binom{n-k}{k} \le 2^{H(\epsilon)k} e^{k\log(en/k)}. \tag{52}$$

To complete the proof, it suffices to show that for all $A \in \binom{\mathcal{S}^*}{\ge(1-\epsilon)k}$ and $B \in \binom{[n]\setminus\mathcal{S}^*}{|A|}$, each of the summands in (51) is $o(\cdot)$ of the reciprocal of the corresponding term in (52).

Tail bounds for T_1, T_2, and T_3. We will prove a tail bound for T_1 in two steps: first we control the conditional expectation $M := \mathbb{E}[T_1 \mid \mathbf{X}]$; then we control T_1 given M. The function

$$f(x_i\,;\, i \in \mathcal{S}^*) := p^3 \sum_{i\in A} \sum_{\substack{j,l \in \mathcal{S}^*\setminus A \\ j \ne l}} \langle x_i, x_j\rangle\langle x_j, x_l\rangle\langle x_l, x_i\rangle$$

satisfies $M = f(X_i\,;\,i \in \mathcal{S}^*)$, and f satisfies the bounded differences property with parameters $c_i = 2p^3|\mathcal{S}^* \setminus A|^2$ for $i \in A$, and $c_i := 2p^3|A| \cdot |\mathcal{S}^* \setminus A|$ for $i \in \mathcal{S}^* \setminus A$. If we let $\lambda := \mathbb{E}M + s/24$ then by the bounded differences inequality (Theorem 6.2 in Boucheron et al. (2013)),

$$\mathbb{P}\left\{M \le \mathbb{E}M - \lambda\right\} \le \exp\left(-\frac{k^5}{Cd^4|A| \cdot |\mathcal{S}^* \setminus A|^3}\right) \le \exp\left(-\frac{k}{C\epsilon^3 d^4}\right) \tag{53}$$

for an absolute constant $C > 0$ (that may vary between lines), where we used that $|A| \le k$ and $|\mathcal{S}^* \setminus A| \le \epsilon k$. Let E be the event on the left-hand side of (53). Using Theorem 2.1 from Janson (2004) and redefining $\lambda := M + s/24$,

$$\mathbb{P}\left\{T_1 \le M - \lambda \text{ and } E^c \text{ holds}\right\} \le \exp\left(-\frac{s^2}{C|A|^2|\mathcal{S}^* \setminus A|^2}\right) \le \exp\left(-\frac{p^6k^2}{C\epsilon^2 d^4}\right). \tag{54}$$

To apply Theorem 2.1 from Janson (2004), we simply used that the maximum degree of the dependency graph of the variables $\{\bar{\mathbf{A}}_{ij}\bar{\mathbf{A}}_{jl}\bar{\mathbf{A}}_{li}\}_{i\in A,\, j,l \in \mathcal{S}^*\setminus A}$ (conditioned on $\mathbf{X}$) is at most $3|A| - 2$ (an edge joins two variables in the dependency graph if and only if they are not independent). Combining (53) and (54), we have

$$\begin{aligned}\mathbb{P}\{T_1 \le -s/12\} &\le \mathbb{P}\{E\} + \mathbb{P}\left\{T_1 \le M - \lambda \text{ and } E^c \text{ holds}\right\} \\ &\le \exp\left(-\frac{k}{C\epsilon^3 d^4}\right) + \exp\left(-\frac{p^6k^2}{C\epsilon^2 d^4}\right).\end{aligned}$$

Using precisely the same argument as for T_1, we calculate that

$$\mathbb{P}\{T_2 \le -s/12\} \le \exp\left(-\frac{k}{C\epsilon d^4}\right) + \exp\left(-\frac{p^6k^2}{C\epsilon d^4}\right).$$

For the term T_3, the argument only needs to be modified slightly. Note that

$$\mathbb{E}M = p^3 \sum_{i,j,l \in A} 1/d^2 \ge 0.99p^3k^3/d^2 = 1.98s.$$

In this case, the function

$$f(x_i\,;\,i \in \mathcal{S}^*) := p^3 \sum_{\substack{i,j,l \in A \\ \text{distinct}}} \langle x_i, x_j\rangle\langle x_j, x_l\rangle\langle x_l, x_i\rangle$$

satisfies the bounded differences property with parameters $c_i = 2p^3|A|^2$, so the variance term in (Boucheron et al., 2013, Theorem 6.2) is $p^6|A|^5$. Therefore, choosing $\lambda = 0.4s$ in the above argument yields

$$\mathbb{P}\{T_3 \le s\} \le \exp\left(-\frac{k}{Cd^4}\right) + \exp\left(-\frac{p^6k^2}{Cd^4}\right).$$

If $\epsilon > 0$ is taken to sufficiently small depending on C and d, then above tail bounds are all $o(2^{-H(\epsilon)k})$.

Tail bounds for T_4, T_5, and T_6. For T_4 and T_5, we may assume $\mathcal{S}^* \setminus A \neq \emptyset$ (otherwise these terms are 0 trivially). We first bound T_4. Notice that $\mathbb{E}[T_4 \,|\, \mathbf{X}] = 0$ with probability 1, since for

$i \in B$ and $j \in \mathcal{S}^* \setminus A$, the random variable $\mathbf{A}_{ij}$ is centered and independent of all else. To apply Theorem 2.1 from Janson (2004), note that the maximum degree of the dependency graph of the random variables $\{\bar{\mathbf{A}}_{ij}\bar{\mathbf{A}}_{jl}\bar{\mathbf{A}}_{li}\}_{i\in B, j,l\in\mathcal{S}^*\setminus A}$ (conditional on $\mathbf{X}$) is at most $3k$. We thus have

$$\begin{aligned}\mathbb{P}\{T_4 \geq s/6 \,|\, \mathbf{X}\} &\leq \exp\left(-2 \cdot \frac{(s/6)^2}{3k \cdot |B| \cdot |\mathcal{S}^* \setminus A|^2}\right) \\ &\leq \exp\left(-\frac{p^6 k^2}{216\epsilon^2 d^4}\right) = o(2^{-H(\epsilon)k} e^{-k\log(en/k)}),\end{aligned}$$

where we used that $|B| \leq k$ and $|\mathcal{S}^* \setminus A| \leq \epsilon k$, and in the last step we used that $k \geq C\log n$ for a large enough constant $C = C(p,d) > 0$. Taking an expectation with respect to $\mathbf{X}$ completes the proof.

Analogous tail bounds for T_5 and T_6 are proved in exactly the same manner as T_4. For T_5, the maximum degree of the dependency graph of the variables $\{\bar{\mathbf{A}}_{ij}\bar{\mathbf{A}}_{jl}\bar{\mathbf{A}}_{li}\}_{i,j\in B, l\in\mathcal{S}^*\setminus A}$ (conditional on $\mathbf{X}$) is again at most $3k$. For T_6, we need not even condition on $\mathbf{X}$ since T_6 is simply the signed triangle count of an Erdős–Rényi random graph.

Appendix F. Probability and linear algebra tools

The following theorem is useful for multiple proofs in this work.

Theorem F.1 (Corollary 3.12 and Remark 3.13 in Bandeira and van Handel (2016)) *Let X be an $n \times n$ symmetric matrix whose entries X_{ij} are centered random variables, independent up to symmetry. Define the quantities*

$$\sigma := \max_{i=1,\dots,n} \sqrt{\sum_{j=1}^{n} \mathbb{E}X_{ij}^2}\,, \qquad \sigma_* := \max_{i,j=1,\dots,n} \|X_{ij}\|_\infty\,,$$

where we write $X_{ij}^2 = (X_{ij})^2$. Then for all $0 < \epsilon \leq \frac{1}{2}$ there exists a universal constant c_ϵ such that for every $t \geq 0$,

$$\mathbb{P}\{\|X\| \geq (1+\epsilon)2\sqrt{2}\sigma + t\} \leq n\exp\left(-\frac{t^2}{c_\epsilon \sigma_*^2}\right).$$

The following is a useful result about the spectral norm of a Hadamard product of matrices.

Lemma F.2 ((3.7.9) in Horn (1990)) *For matrices $B \in \mathbb{R}^{n\times n}$, $X, Y \in \mathbb{R}^{m\times n}$, and $A = X^\top Y$, we have*

$$\|A \circ B\| \leq \sqrt{\|\mathbf{I}_n \circ (X^\top X)\| \cdot \|\mathbf{I}_n \circ (Y^\top Y)\|} \cdot \|B\|.$$

In particular, if $X = Y$, then

$$\|A \circ B\| \leq \max_{i\in[n]} A_{ii} \cdot \|B\|.$$