A non-backtracking method for long matrix and tensor completion

Ludovic Stephan
École Polytechnique Fédérale de Lausanne (EPFL), IdePHICS Lab

Yizhe Zhu
Department of Mathematics, University of California Irvine

Editors: Shipra Agrawal and Aaron Roth

Abstract

We consider the problem of low-rank rectangular matrix completion in the regime where the matrix $M$ of size $n \times m$ is “long”, i.e., the aspect ratio $m/n$ diverges to infinity. Such matrices are of particular interest in the study of tensor completion, where they arise from the unfolding of a low-rank tensor. In the case where the sampling probability is $\frac{d}{\sqrt{mn}}$, we propose a new spectral algorithm for recovering the singular values and left singular vectors of the original matrix $M$ based on a variant of the standard non-backtracking operator of a suitably defined bipartite weighted random graph, which we call a non-backtracking wedge operator. When $d$ is above a Kesten-Stigum-type sampling threshold, our algorithm recovers a correlated version of the singular value decomposition of $M$ with quantifiable error bounds. This is the first result in the regime of bounded $d$ for weak recovery and the first for weak consistency when $d \to \infty$ arbitrarily slowly without any polylog factors. As an application, for low-CP-rank orthogonal $k$-tensor completion, we efficiently achieve weak recovery with sample size $O\left(\frac{n^k}{2}\right)$ and weak consistency with sample size $\omega\left(\frac{n^k}{2}\right)$. A similar result is obtained for low-multilinear-rank tensor completion with $O\left(\frac{n^k}{2}\right)$ many samples.

Keywords: Tensor completion, long matrix completion, non-backtracking operator

1. Introduction

Matrix completion is the problem of reconstructing a matrix from a (usually random) subset of entries, leveraging prior structural knowledge such as rank and incoherence. When the two dimensions $n, m$ of the matrix $M$ are comparable, this problem has been widely studied in the past decades, with a fairly general pattern emerging. Roughly speaking, if the ground truth matrix is low-rank and eigenvectors are sufficiently incoherent (or delocalized), sampling $\Omega(n \log n)$ many entries uniformly at random with a certain efficient optimization procedure, one can exactly recover the ground truth matrix (Keshavan et al., 2009, 2010; Candes and Tao, 2010; Candes and Plan, 2010; Recht, 2011; Candes and Recht, 2012; Jain et al., 2013). This is not the case when the sample size is only $O(n)$, where full completion is provably impossible under uniform sampling. However, an approximation of the ground truth matrix is still possible (Heiman et al., 2014; Gamarnik et al., 2017; Brito et al., 2022) with efficient algorithms, and more refined estimates of the eigenvectors were studied in (Bordenave et al., 2022b).

In recent years, attention has shifted to “long” matrix models where the row and column sizes of the matrix are not proportional. This is the case, for example, in the spiked long matrix and tensor models (Montanari and Wu, 2022; Ding et al., 2020; Ben Arous et al., 2023; Montanari and Richard, 2014). This situation also occurs in bipartite graph clustering and community detection, where the goal is to recover the community structure in the smaller vertex set (Florescu and Perkins, 2016;
Ndaoud et al., 2021; Cai et al., 2021; Braun and Tyagi, 2023; Braun, 2023) for a bipartite stochastic block model, and in subspace recovery problems (Cai et al., 2021; Zhou et al., 2021).

Our work is closely related to tensor completion, which is the higher-order analog of matrix completion. Indeed, tensor problems are highly related to their long matrix counterparts through the unfolding operation for an order-\(k\) tensor of size \(n^k\):

\[
\text{unfold}_{a,b} : \mathbb{R}^{n^k} \rightarrow \mathbb{R}^{n_a \times n_b},
\]

where \(a + b = k\). For example, an order-3 tensor of size \(n\) yields an unfolded matrix of size \(n \times n^2\); in general, uneven unfoldings (i.e., with \(a \neq b\)) of a tensor give rise to long rectangular matrices. In those tensor problems, the ground truth is assumed to have a low CP-rank, which implies that the unfolded matrix is low-rank. In low-rank \(k\)-tensor completion, the best known polynomial-time algorithms require \(\tilde{O}(n^{k/2})\) revealed entries (Jain and Oh, 2014; Montanari and Sun, 2018; Xia and Yuan, 2019; Cai et al., 2022), while information theoretically \(\tilde{O}(n)\) suffices (Ghadermarzy et al., 2019; Harris and Zhu, 2021; Harris et al., 2023). In (Barak and Moitra, 2022), the authors showed lower bounds for a family of algorithms based on the sum-of-squares hierarchy for order-3 noisy tensor completion, where \(\Omega(n^{3/2})\) many samples are needed, by connecting it to the literature on refuting random 3-SAT. They also conjectured such a sample size is needed for any polynomial-time algorithm. This is aligned with other tensor estimation problems where a diverging statistical-to-computational gap is expected (Montanari and Richard, 2014; Wein et al., 2019; Jagannath et al., 2020; Arous et al., 2020; Dudeja and Hsu, 2024; Auddy and Yuan, 2022).

However, so far no known work has reached the exact threshold \(\Omega(n^{3/2})\), only approaching it up to polylog factors, in a similar fashion to the matrix completion phase diagram described above. For exact recovery results on tensor completion, all the methods in the literature are either convex optimization or spectral initialization with iterative optimization steps. Our method achieves weak consistency without any log factors, which could improve the sample size for spectral initialization in the work (Jain and Oh, 2014; Cai et al., 2022; Xia and Yuan, 2019; Xia et al., 2021).

The applications we mentioned above are related to the same question:

*Can we improve the sample complexity for estimating a one-sided structure in a long matrix?*

Specifically, we explore the use of the *non-backtracking operator* in this context. The non-backtracking operator is an important object in the study of spectral graph theory (Bass, 1992; Terras, 2010). Recently, it has been a powerful tool for understanding the spectrum of a very sparse random matrix (Bordenave et al., 2018; Stephan and Massoulié, 2022; Bordenave, 2020; Bordenave and Collins, 2019; Brito et al., 2022; Benaych-Georges et al., 2020; Alt et al., 2021b,a; Dumitriu and Zhu, 2022; Stephan and Zhu, 2022) and for designing efficient algorithms in community detection and matrix completion (Krzakala et al., 2013b; Bordenave et al., 2018; Stephan and Massoulié, 2022; Bordenave et al., 2022b; Stephan and Zhu, 2022; Zhu and Zhu, 2023). In the bipartite graph setting, the non-backtracking spectrum was analyzed in (Brito et al., 2022; Dumitriu and Zhu, 2022, 2021) when the two sides have a bounded aspect ratio. For the spectrum of the adjacency matrix of a random bipartite graph with a diverging aspect ratio, only a few results were obtained in (Zhu, 2023; Dumitriu and Zhu, 2023) for random bipartite biregular graphs, and in (Guruswami et al., 2022) for the adjacency matrix of bipartite random graphs with random sign flips.

In this work, we aim to address this gap by defining a non-backtracking operator suitable for sparse matrix estimation with a diverging aspect ratio, which we call *a non-backtracking wedge matrix*. Our work sheds light on the fundamental limits and practical algorithms for sparse long matrix estimation with a focus on one-sided structure recovery. To the best of our knowledge, this is
Non-backtracking tensor completion

The first result for a long $n \times m$ low-rank matrix completion with sampling probability $p = \frac{d}{\sqrt{mn}}$ in the regime of bounded $d$ for weak recovery and the first result for weak consistency when $d \to \infty$ arbitrarily slowly without any polylog factors. These two types of recovery guarantees are well studied in the literature of community detection (Abbe, 2018). In fact, a direct adaptation of the bipartite non-backtracking operator we define here leads to a simpler spectral algorithm for the detection of the bipartite stochastic block model studied in (Florescu and Perkins, 2016). Below, we summarize our results informally. See Theorems 2 and 3 for the precise statements.

**Theorem 1 (long matrix completion, informal)** With $O(\sqrt{mn})$ many samples, if the left singular vectors of a long $n \times m$ matrix ($m \gg n$) are sufficiently delocalized, and the singular values are above a specific Kesten-Stigum threshold, then the non-backtracking wedge operator can recover the ground truth singular values and find correlated estimates of the left singular vectors.

The Kesten-Stigum threshold in Theorem 1 is analogous to a similar threshold for community detection in the stochastic block model, which characterizes a threshold of the signal-to-noise ratio for the existence of polynomial time algorithms to achieve weak recovery; see (Abbe, 2018) for more details. We discuss the Kesten-Stigum bound in matrix completion in Section 2.4.

We then apply our main results to the low-rank tensor completion problem under an orthogonality assumption on the $r$ components in a tensor decomposition. This is the first efficient algorithm to achieve weak recovery of a low-rank tensor with $\Omega(n^{k/2})$ samples without any extra polylog($n$) factors. In particular, an explicit $n^{k/2}$ threshold is obtained for rank-1 tensor completion when the entries are $\pm 1$. The weak recovery results can serve as a good spectral initialization step for other optimization procedures (Montanari and Sun, 2018; Jain and Oh, 2014; Cai et al., 2022). Unlike the common approach of unfolding a tensor to a matrix as square as possible (i.e., applying unfold $\lfloor k/2 \rfloor, \lceil k/2 \rceil$ to $T$) in (Mu et al., 2014; Montanari and Sun, 2018), we unfold $T$ to an $n \times n^{k-1}$ matrix in the most unbalanced way, and this allows us to directly estimate individual components of the tensor decomposition with sample complexity $O(n^{k/2})$. Note that simply applying the results in (Bordenave et al., 2022b) to unfold $\lfloor k/2 \rfloor, \lceil k/2 \rceil$ to $T$ will have sample complexity $O(n^{\lceil k/2 \rceil})$, and one cannot directly estimate any individual component in $T$. Therefore, our algorithm has a better computational complexity for tensor completion for all integer values $k$. The long matrix completion result also has a direct application in low-multilinear-rank tensor completion. Detailed statements and discussion on tensor completion are given in Sections 2.5 and 2.6.

Although there is a strong connection between the detection problem for matrix completion and community detection in the stochastic block model (Angelini et al., 2015; Stephan and Massoulié, 2022; Bordenave et al., 2018, 2022b), the statistical-to-computational gap in the hypergraph stochastic block model behaves very differently from tensor completion (Pal and Zhu, 2021; Stephan and Zhu, 2022; Gu and Polyanskiy, 2023). In hypergraph community detection, efficient spectral methods can achieve weak recovery with the sample complexity down to the information-theoretical threshold up to a constant factor in all cases. However, in tensor completion, there is a conjectured diverging gap between the information-theoretical sample complexity and sample complexity for efficient algorithms (Barak and Moitra, 2022).

The non-backtracking wedge operator we defined in Section 2.1 is tailored for the situation where the underlying sparse bipartite graph has a diverging aspect ratio. It is similar in spirit to the non-backtracking operator for hypergraphs (Angelini et al., 2015; Liu et al., 2022; Stephan and Zhu, 2022), and the nomadic walk operator in (Mohanty et al., 2020; Guruswami et al., 2022) for (signed) random bipartite biregular graphs. The main difference compared to (Guruswami et al.,
is that we work with a non-centered random matrix model without degree homogeneity, and we only count one-sided non-backtracking walks. In addition, a detailed eigenvector analysis is provided while (Guruswami et al., 2022; Mohanty et al., 2020) are only about extreme eigenvalues. In terms of tensor estimation with path-counting arguments, another relevant work is the estimation of a spiked tensor model under heavy-tail noise (Ding et al., 2020), where the authors use $n \times n$ symmetric matrices to count the number of self-avoiding walks of length $c \log n$. However, those matrices are usually more expensive to compute and require a new parameter (the length of the paths considered) that needs tuning in practice.

2. Setting and main results

2.1. Detailed setting for long matrix completion

We consider a rectangular matrix $M$ of size $n \times m$, with aspect ratio $\alpha = m/n$. We write the singular value decomposition of $M$ as

$$M = \sum_{i=1}^{n} \nu_i \phi_i \psi_i^T,$$

where the $\nu_i$ are the singular values of $M$, and the $(\phi_i)_{i \in [r]}$ (resp. the $(\psi_i)_{i \in [r]}$) are an orthonormal family of its left (resp. right) singular vectors. We order the singular values of $M$ in decreasing order, that is $\nu_1 \geq \cdots \geq \nu_n$. Our observed matrix is of the form

$$A = \frac{\sqrt{mn}}{d} (X \circ M),$$

where $X$ is an i.i.d Bernoulli matrix:

$$X_{ij} \overset{i.i.d}{\sim} \text{Ber} \left( \frac{d}{\sqrt{mn}} \right).$$

We will use the shortcut $p = d/\sqrt{mn}$ as the sampling probability. The scaling of $A$ is chosen so that the expected value of $A$ is exactly $M$. Our goal is to weakly recover the eigendecomposition:

$$MM^T = \sum_{i=1}^{n} \nu_i^2 \phi_i \phi_i^T.$$

The non-backtracking wedge matrix We can view the matrix $A$ as the weighted biadjacency matrix of a bipartite graph $G = (V_1 \cup V_2, E)$ with $|V_1| = n$, $|V_2| = m$. We call $V_1$ a left vertex set and $V_2$ a right vertex set. A path of length 2 in $G$ denoted by $(e_1, e_2, e_3)$ is called a wedge if $e_1, e_3 \in V_1, e_2 \in V_2$, and $e_1 \neq e_3$. Define the set of oriented wedges $E$ as

$$E = \{(e_1, e_2, e_3) \in V_1 \times V_2 \times V_1 : \{e_1, e_2\} \in E, \{e_2, e_3\} \in E, e_1 \neq e_3\}.$$  

The non-backtracking wedge matrix $B$ associated to $G$ is defined on $E \times E$ by

$$B_{ef} = 1_{e \to f} A_{f_1 f_2} A_{f_3 f_2}, \quad \text{where} \quad 1_{e \to f} = 1_{e_3 = f_1} 1_{e_2 \neq f_2}.$$
Figure 1: Example of a non-backtracking path between two wedges \((e_1, e_2, e_3)\) and \((f_1, f_2, f_3)\).

Equivalently, \(1_{e \rightarrow f}\) is one if and only if \((e_1, e_2, e_3, f_2, f_3)\) is a non-backtracking walk of length 4 in \(G\). See Figure 1 for an example. Note that the non-backtracking wedge matrix \(B\) is not symmetric: its index set is the oriented wedge set defined in (2), and \(B_{ef} \neq B_{fe}\) in (3).

For comparison, we recall the definition of the non-backtracking operator for weighted graphs as follows. Let \(A\) be a weighted adjacency matrix of a graph \(G = (V, E)\). Define the oriented edge set \(E\) for \(G\) as \(E = \{(i, j) : \{i, j\} \in E\}\). The non-backtracking operator \(B\) of associated with \(A\) is a non-Hermitian operator of size \(|E| \times |E|\). For any \((u, v), (x, y) \in E\), \(B\) is defined as

\[
B_{(u,v),(x,y)} = \begin{cases} A_{xy} & v = x, u \neq y, \\ 0 & \text{otherwise.} \end{cases}
\]  

In the case of an \(n \times m\) bipartite weighted graph defined in (1), the non-backtracking wedge operator only counts non-backtracking walks starting from a vertex in \(V_1\) while the non-backtracking operator defined in (4) counts non-backtracking walks from both \(V_1\) and \(V_2\).

**Intuition for the non-backtracking wedge matrix** It has been shown that the non-backtracking operator, as a non-hermitian matrix, outperforms the standard (weighted) adjacency matrix when the sample size is very small in many graph clustering and matrix completion problems (Bordenave et al., 2018, 2022b; Stephan and Massoulié, 2022; Stephan and Zhu, 2022). The key difference is that the spectrum of the non-backtracking operator is more informative and stable when the underlying graph is very sparse with a constant average degree. It is also a general phenomenon that the spectrum of non-hermitian random matrices is less sensitive to a row or column with a large \(\ell_2\)-norm (Coste, 2023; Bordenave et al., 2022a).

In terms of the sample size, the non-backtracking matrix in (4) was used in (Bordenave et al., 2022b; Stephan and Massoulié, 2022) to achieve the optimal scale \(O(n)\) for square matrix completion. However, for a low-rank tensor in a matrix form of size \(n \times n^2\), the work (Bordenave et al., 2022b) only achieves weak recovery with sample size \(O(n^2)\), while with the non-backtracking wedge operator, we only need \(O(n^{1.5})\). The standard way in (Montanari and Sun, 2018; Cai et al., 2022) to obtain \(\tilde{O}(n^{1.5})\) sample size is to apply a spectral method on the hollowed Gram matrix \(Z = AA^T - \text{diag}(AA^T)\). For \(i \neq j\), we have

\[
Z_{ij} = \sum_{\text{wedge } i \rightarrow k \rightarrow j} A_{ik}A_{kj}.
\]

Hence, the spectrum of \(Z\) depends on wedge walk counts. Our new non-backtracking operator only counts the wedge walks that are non-backtracking, which avoids the localization effect from high-degree vertices (Benaych-Georges et al., 2019) in the weighted graph corresponding to \(Z\). This new operator introduces a tool to “regularize” the spectrum of a heavy-tailed long random matrix without any tuning parameter, which might be useful for other applications.
2.2. Quantifying the convergence

The alignment between the spectral decomposition of $B$ and the SVD of $A$ will be governed by a wealth of different quantities. For ease of understanding, we divide those parameters into two categories, depending on their role in the convergence.

**Threshold parameters** These parameters describe which parts of the SVD of $M$ will be reflected in the spectral decomposition of $B$.

(i) **Variance parameter:**
\[ \rho = \|Q\|, \quad \text{where} \quad Q = \sqrt{mn}(M \circ M). \] (5)

(ii) **Amplitude parameter:**
\[ L = \sqrt{mn} \max_{x \in [n], y \in [m]} |M_{xy}|. \]

(iii) **Detection thresholds**
\[ \theta = \max(\theta_1, \theta_2), \]
where the two parameters $\theta_1$ and $\theta_2$ are defined as
\[ \theta_1 = \sqrt{\frac{\rho}{d}} \quad \text{and} \quad \theta_2 = \frac{L}{d}. \]

(iv) **Detection rank:**
\[ r_0 = \max_i \{i \in [n] : \nu_i > \theta\}. \]
Equivalently, it is the number of singular values of $M$ above the detection threshold $\theta$.

(v) **Inverse signal-to-noise ratios (inverse SNRs):** for $i \in [r_0]$,
\[ \tau_i = \frac{\theta}{\nu_i} \in (0, 1). \]
The largest of those inverse SNRs, $\tau_{r_0}$, will be denoted by $\tau$.

**Complexity parameters** Those parameters quantify how intrinsically “easy” the matrix $M$ is to recover. In particular, contrary to the previous ones, they are all invariant under any rescaling of $M$.

(i) **Rescaled amplitude parameter:**
\[ K = \frac{L^2}{\rho}. \] (6)

(ii) **Aspect log-ratio:**
\[ \eta = \frac{\log(m)}{\log(n)} - 1. \] (7)
Equivalently, it is the solution of the equation $m = n^{1+\eta}$. When $M$ is obtained from the $(a, b)$ unfolding of a tensor, we get $\eta = \frac{b-a}{a}$.

(iii) **Rank:**
\[ r = \text{rank}(M) = \max \{i \in [n] : \nu_i \neq 0\}. \]

(iv) **Delocalization parameter:**
\[ \kappa = \sqrt{n} \max_i \|\phi_i\|_\infty. \]
Note that $\kappa \geq 1$ by definition. We require no such conditions on the right singular vectors $\psi_i$. 

Simplifying assumptions  We also make a few simplifying assumptions throughout the paper; although our results are still valid when those assumptions aren’t satisfied, the proofs become much messier. Recall the aspect ratio is defined by $\alpha = \frac{m}{n} \geq 1$. We first assume that

$$\liminf_n d > 1 \quad \text{and} \quad d \leq n. \quad (8)$$

We also assume that

$$\eta \geq 2 \log(d) + 2 \log \log_d(n) \left( \log(n) \right),$$

or equivalently

$$\alpha \geq \frac{d^2 (\log d n)^2}{\log(n)}. \quad (9)$$

We note that if $d > n$, our results become trivially true (see the definition of $\ell$ in Theorem 2, and the bounds thereafter).

2.3. Main results for long matrix completion

Our first result shows that the eigenvalues of $B$ exhibit a structure similar to those observed in (Stephan and Massoulié, 2022; Bordenave et al., 2022b): namely, most eigenvalues of $B$ lie in a circle of known radius, with a few outlier eigenvalues correlated with the singular values of $M$.

**Theorem 2 (Singular value estimation)** Let

$$\ell = \lceil \varepsilon \log_d(n) \rceil, \quad \text{with} \quad \varepsilon = \frac{(\eta/2) \wedge 1}{25}.$$  

There exist constants $C_0, n_0 \geq 1$ that depend polynomially on $r, K, \kappa, d, \log n$ such that, if $n \geq n_0$ and $C_0 \tau^2 \ell \leq 1$, the following holds on an event with probability at least $1 - cd^{-4 n^{-1/4}}$:

(i) There exists an ordering $\lambda_1, \ldots, \lambda_{r_0}$ of the top $r_0$ eigenvalues of $B$ such that for any $i \in [r_0]$, $|\lambda_i - \nu_i^2| \leq C_0 \nu_i^2 \tau^2 \ell \delta_i, \ell$.

(ii) All other eigenvalues of $B$ have modulus at most $C_0^{1/\ell} \theta^2$.

Our second result concerns quantifying the overlaps between the top eigenvectors of $B$ and the singular vectors of $M$. For $i \in [r_0]$, we define the relative eigengap

$$\delta_{i, \ell} = \min_{\nu_j \neq \nu_i} \left| 1 - \left( \frac{\nu_j}{\nu_i} \right)^{2\ell} \right|. \quad (10)$$

Let $\xi_i^R, \xi_i^L$ a right (resp. left) unit eigenvector of $B$ associated with $\lambda_i$, and define

$$\zeta_i^R(x) := \sum_{e \in E, e_1 = x} A_{e_1 e_2} A_{e_3 e_2} \xi_i^R(e) \quad \text{and} \quad \zeta_i^L(x) := \sum_{e \in E, e_1 = x} \xi_i^L(e). \quad (11)$$

**Theorem 3 (Left singular vector estimation)** Under the same setting and conditions as Theorem 2, with probability at least $1 - cd^{-4 n^{-1/4}}$, for $i \in [r_0]$, there exists a unit left singular vector $\phi_i^l$ of $M$ associated with $\nu_i$ such that

$$\left| \langle \zeta_i^R, \phi_i^l \rangle - \frac{1}{\sqrt{\gamma_i}} \right| \leq \frac{C_0 \tau_i^2 \ell}{\delta_{i, \ell}} \quad \text{and} \quad \left| \langle \zeta_i^L, \phi_i^l \rangle - \frac{1}{\sqrt{\gamma_i}} \right| \leq \frac{C_0 \tau_i^2 \ell}{\delta_{i, \ell}}, \quad (12)$$

where $\gamma_i$ is defined as

$$\gamma_i = \begin{pmatrix} 1, \left( I - \frac{QQ^*}{\nu_i^4 d^2} \right)^{-1} (\phi_i \circ \phi_i) \end{pmatrix}. \quad (13)$$
2.4. Discussion

Comparison to (Bordenave et al., 2022b) In the seminal work of (Bordenave et al., 2022b), the authors provide a spectral algorithm to weakly recover a low-rank rectangular matrix of size $m \times n$ with $O(m + n)$ many samples, where a Kesten-Stigum bound for estimation singular values and both left and right eigenvectors are obtained. In our low-rank long matrix setting ($m \gg n$), their sample complexity scales as $O(m)$. In contrast, our non-backtracking wedge operator is able to estimate singular values and left singular vectors of a long matrix with sample size $O(\sqrt{mn})$. This improvement in sample complexity to recover one-sided structure is useful in tensor completion; see Section 2.5 for more details.

Overlap convergence and weak consistency In full generality, the overlaps $\gamma_i$ in Theorem 3 have a complicated expression. However, this can be simplified under a homogeneity assumption, namely

$$D_x := \sum_{y \in V_1, z \in V_2} Q_{xz} Q_{yz}$$

is independent of $x$. In this case, the Perron-Frobenius theorem implies that $D_x = \rho^2$, and hence

$$\langle \zeta_i^R, \phi_i \rangle = \sqrt{1 - \tau_i^4} + o(1), \quad \langle \zeta_i^L, \phi_i \rangle = \sqrt{1 - \tau_i^4} + o(1).$$

This implies a weak recovery of the singular vectors $\phi_i$. In general, as $d \to \infty$, we have $(\gamma_i - 1) \sim C/d^2$, and hence (12) implies the eigenvectors of $B$ are asymptotically aligned with the left singular vectors of $M$. Under the additional assumption that all singular values of $M$ are distinct, the top $r$ eigenvectors of $B$ are asymptotically aligned with all left singular vectors of $M$, and eigenvalues are consistent estimators for $\nu_i, i \in [r]$: this is a form of weak consistency for matrix completion.

Optimality of the bounds and thresholds We have made no effort towards getting closed-form exponents in the polynomials $C_0$ and $n_0$, which would probably not be sharp. An explicit bound on $C_0, n_0$ in the proportional case for matrix completion is worked out in (Bordenave et al., 2022b, Theorem 5.4). Similarly, although we do expect (when $d, \eta$ are fixed) that the rate of convergence is indeed of the form $n^{-c}$, we didn’t try to optimize the value of such $c$ as much as possible.

Our work shows a threshold behavior for long matrix completion: one can weakly recover the components in $MM^\top$ whose signal-to-noise ratio is above the threshold. On the other hand, the optimality of $\theta$ is still an open and interesting question since if this threshold can be lowered, more information about $M$ can be recovered. The first threshold $\theta_1$ is inherent to the problem and can be found in many similar problems (e.g., the Kesten-Stigum threshold in community detection (Krzakala et al., 2013a; Bordenave et al., 2018; Mossel et al., 2015, 2018)): it is related to an intrinsic property on the existence of a pseudo-eigenvector in a Galton-Watson tree with Poisson offspring distribution of mean $d$, see Section E for details. But whether $\theta_2$ is needed or is an artifact of proof is still unclear. We note that in a similar problem regarding the configuration model (Coste, 2021), such a seemingly spurious threshold turns out to be the sharp one (albeit only for one eigenvalue). Note that as $d$ grows, $\theta_2$ decays faster than $\theta_1$, therefore $\theta = \theta_1$ for sufficiently large $d$.

Recovering right singular vectors We draw attention to the fact that our result is almost completely independent of the right singular vectors $\psi_i$ of $M$, both on the level of assumptions (except for their influence on $K$) and results. We believe this is not an artifact of proof: indeed, most of the right vertices (a proportion $1 - \alpha^{-1}$ of them) are isolated, and it is impossible to recover any
signal from those. This is a completely different behavior compared with (Bordenave et al., 2022b), where the authors show that when \( \alpha \) is bounded, both left and right singular vectors of \( M \) can be recovered. However, it matches recent results from (Montanari and Wu, 2022) in the framework of long matrix reconstruction showing different reconstruction thresholds for the left and right singular vectors. Our work also shows that one can effectively estimate the rank of a long matrix of size \( n \times m \) with \( O(\sqrt{mn}) \) random samples when the rank \( r = O(1) \). When \( m, n \) are proportional, a similar result was obtained in (Saade et al., 2015) based on the Bethe-Hessian matrix.

**Parameter scaling** All of the results in Theorems 2 and 3 are non-asymptotic, and with more care in the proof, we would be able to give explicit expressions for \( \mathcal{C} \). Parameter scaling

\[
K, \kappa, r = n^{o(1)}, \quad \text{and} \quad \log(d)^2 = o(\log(n)).
\]

### 2.5. Application in low-CP-rank tensor completion

Throughout this part, our goal will be to recover a tensor \( T \) of order \( k \geq 3 \), which has a CP decomposition (see, e.g., (Kolda and Bader, 2009)) with \( r \) components:

\[
T = \sum_{i=1}^{r} \nu_i w_i^{(1)} \otimes w_i^{(2)} \otimes \cdots \otimes w_i^{(k)},
\]

where \( \|w_i^{(j)}\|_2 = 1 \) for all \( i \in [r], j \in [k] \). The smallest integer \( r \) for the existence of the decomposition (14) is called the CP-rank of \( T \), and we will call it the rank of \( T \) for simplicity in this section.

It is known that for a general low-rank tensor \( T \), finding its CP-decomposition is NP-hard (Hillar and Lim, 2013). Even with the complete information of the ground truth \( T \), identifying each component in \( \{w_i^{(j)}, i \in [r], j \in [k]\} \) is not feasible. Since we consider asymmetric tensor completion, it will be useful to consider a general version of unfolding:

**Definition 4** Let \( S, T \) be a partition of \([k]\) in two subsets. The \((S, T)\)-unfolding of \( T \); denoted as unfold\(_{S,T}(T)\), is a matrix \( M \in \mathbb{R}^{n_S \times n_T} \) defined as follows: for any multi-index \((i_s)_{s \in [k]}\),

\[
M_{(i_s)_{s \in S},(i_s)_{s \in T}} = T_{(i_s)_{s \in [k]}}.
\]

In particular, when \( T \) is as in (14), we have

\[
\text{unfold}_{S,T}(T) = \sum_{i=1}^{r} \nu_i \left( \bigotimes_{j \in S} w_j \right) \left( \bigotimes_{j \in T} w_j \right)^*.
\]

For simplicity, we define the mode-\( i \) unfolding of \( T \) as \( \text{unfold}_i(T) := \text{unfold}_{\{i\},[k]\setminus\{i\}}(T) \).

However, we make an important remark: Equation (15) is not necessarily the SVD of the tensor unfolding. Indeed, the CP decomposition does not always have orthogonal components, and hence, the vectors in (15) are not necessarily orthogonal. Generally, for a given tensor \( T \), the spectra of the unfolded matrices are not the same for different unfoldings, and there is a hierarchy of unfolded matrix norms (Wang et al., 2017). In the following, we will denote by \( \tilde{T} \) an observed tensor from \( T \) with sampling probability \( p = \frac{d}{n^{k/2}} \). For an unfolding mode \( i \), we let \( M^{(i)} = \text{unfold}_i(\tilde{T}) \), and \( B^{(i)} \) be its associated non-backtracking wedge matrix.
2.5.1. The Rank-1 Case

When \( r = 1 \), we write

\[
T = w^{(1)} \otimes \cdots \otimes w^{(k)},
\]

where we choose \( \nu_1 = 1 \) in (14) for convenience. Then we have

\[
\text{unfold}_i(T) := \text{unfold}_{\{i, [k]\} \setminus \{i\}}(T) = w^{(i)}(w^{(1)} \otimes \cdots \otimes w^{(i-1)} \otimes w^{(i+1)} \otimes \cdots \otimes w^{(k)})^\top.
\]

We now write down the corresponding model parameters for \( M^{(i)} \). It can be checked that they are the same for all \( i \), independent of the choice for unfolding, except for \( \kappa \):

\[
\begin{align*}
\theta_1 &= \sqrt{\frac{n k/2 \prod_{j=1}^k \| w^{(j)} \|_4^2}{d}}, & \theta_2 &= \frac{n k/2 \prod_{j=1}^k \| w^{(j)} \|_\infty}{d}, \\
K &= n k/2 \prod_{j=1}^k \| w^{(j)} \|_2^2, & \kappa^{(i)} &= \sqrt{n} \| w^{(i)} \|_\infty \leq \sqrt{n} \max_i \| w^{(i)} \|_\infty =: \kappa_{\max}.
\end{align*}
\]

Note that we always have \( \theta_1^2 \leq \theta_2 \). Applying Theorems 2 and 3 to \( \text{unfold}_i(T) \), we obtain the following corollary:

**Corollary 5 (Rank-1 tensor completion)** Let \( k \geq 3 \), and define \( \ell \) as

\[
\ell = \lfloor c_1 \log_d(n) \rfloor, \quad c_1 = \frac{(k/2 - 1) \wedge 1}{25}.
\]

For any constant \( \varepsilon > 0 \), assume that

\[
d > (1 + \varepsilon) n k/2 \prod_{j=1}^k \| w^{(j)} \|_\infty.
\]

Then there exists \( C_0, n_0 \geq 1 \) that depend polynomially on \( K, \kappa_{\max}, d, \log n \) such that for any \( i \in [k], \quad n \geq n_0 \) and \( C_0 \theta^{2 \ell} \leq 1 \), the following holds on an event with probability at least \( 1 - cd^4 n^{-1/4} \):

(i) The top eigenvalue \( \lambda^{(i)} \) of \( B^{(i)} \) satisfies

\[
|\lambda^{(i)} - 1| \leq C_0 \theta^{2 \ell},
\]

where \( \theta = \max(\theta_1, \theta_2) \) with \( \theta_1, \theta_2 \) in (17).

(ii) All other eigenvalues of \( B^{(i)} \) have modulus at most \( C_0^{1/\ell} \theta^2 \).

(iii) There exist associated left and right eigenvectors \( \xi^{(i),L}, \xi^{(i),R} \) of \( B^{(i)} \) such that

\[
\left| \langle \xi^{(i),L}, w^{(i)} \rangle - \sqrt{1 - \theta^2} \right| \leq C_0 \theta^{2 \ell} \quad \text{and} \quad \left| \langle \xi^{(i),R}, w^{(i)} \rangle - \sqrt{1 - \theta^2} \right| \leq C_0 \theta^{2 \ell},
\]

where \( \xi^{(i),L}, \xi^{(i),R} \) are defined in (11).

Hence, successively applying tensor unfolding \( k \) times to the observed tensor is enough to estimate all of the components of \( T \). When all components \( w_1, \ldots, w_k \) are vectors in \( \frac{1}{\sqrt{n}} \{ \pm 1 \}^n \), with \( (1 + \varepsilon)n k/2 \) samples, we have

\[
\langle \xi^{(i),L}, w^{(i)} \rangle = \frac{\sqrt{(1 + \varepsilon)^2 - 1}}{1 + \varepsilon} + O(n^{-c}).
\]
2.5.2. Low-rank orthogonal components estimation

Our long matrix completion results aim to estimate individual left singular vectors above a Kesten-Stigum-type threshold. However, in the tensor case, without orthogonal assumption on the vectors \([w_i^{(1)}, \ldots, w_i^{(k)}]\), such a reduction to the estimation of left singular vectors in long matrix completion is impossible. Therefore, we consider the same setting as in (Jain and Oh, 2014; Potechin and Steurer, 2017), and assume that \(T\) can be written as

\[
T = \sum_{j=1}^{r} \nu_j w_j^{(1)} \otimes w_j^{(2)} \otimes \cdots \otimes w_j^{(k)},
\]

(20)

where for each \(i \in [k]\), \([w_1^{(i)}, \ldots, w_r^{(i)}]\) forms an orthonormal basis, and \(\nu_1 \geq \nu_2 \geq \cdots \geq \nu_r > 0\). In contrast with the rank-one case, we don’t have access to an orthogonal CP-decomposition of \(T \circ T\), and hence most parameters used in Theorem 2 depend on the unfolding mode \(i\); we shall denote them by \(\rho^{(i)}, Q^{(i)}, \) and so on. We can still provide the following estimates of \(\rho^{(i)}\) and \(L^{(i)}\):

\[
n^{k/2}\|T\|_2 \leq \rho^{(i)} \leq n^{k/2}\|T\|_F, \quad \text{and} \quad L^{(i)} = L := n^{k/2}\|T\|_{\infty}.
\]

**Corollary 6 (Low-rank orthogonal tensor completion)** Let \(T\) be a \(k\)-tensor with an orthogonal decomposition (20) and \(k \geq 3\), and define \(\ell\) as

\[
\ell = \left[ c_1 \log_d(n) \right], \quad c_1 = \frac{(k/2 - 1) \wedge 1}{25}.
\]

For any \(i \in [k]\), there exists \(C_0^{(i)}, n_0^{(i)}\) that depend polynomially on \(K^{(i)}, \kappa^{(i)}, r^{(i)}, d^{(i)}, \log(n)\), such that, if \(n \geq n_0\) and \(C_0^{(i)} \left( \tau^{(i)} \right)^{2\ell} \leq 1\), the following holds with probability at least \(1 - cd^4n^{-1/4}\):

(i) There exists an ordering \(\lambda_1^{(i)}, \ldots, \lambda_{r_0}^{(i)}\) of the top \(r_0\) eigenvalues of \(B^{(i)}\) such that for any \(j \in [r_0^{(i)}]\),

\[
|\lambda_j - \nu_j^2| \leq C_0^{(i)} \nu_j^2 \left( \tau^{(i)} \right)^{2\ell}.
\]

(ii) All other eigenvalues of \(B^{(i)}\) have modulus at most \(C_0^{(i)} \left( \tau^{(i)} \right)^{1/\ell} \left( \rho^{(i)} \right)^2\).

(iii) For \(j \in [r_0^{(i)}]\), there exist associated left and right eigenvectors \(\xi_j^L, \xi_j^R\) of \(B^{(i)}\) such that

\[
\left| \langle \xi_j^{(i)},R, w_j^{(i)} \rangle - \frac{1}{\sqrt{\gamma_j^{(i)}}} \right| \leq \frac{C_0^{(i)} \left( \tau_j^{(i)} \right)^{2\ell}}{\delta_j,\ell}, \quad \left| \langle \xi_j^{(i)},L, w_j^{(i)} \rangle - \frac{1}{\sqrt{\gamma_j^{(i)}}} \right| \leq \frac{C_0^{(i)} \left( \tau_j^{(i)} \right)^{2\ell}}{\delta_j,\ell},
\]

where \(\gamma_j\) is defined as

\[
\gamma_j^{(i)} = \left( \frac{1}{\sqrt{\nu_j^2}}, \left( I - Q^{(i)}Q^{(i)*} \right)^{-1}(w_j^{(i)} \circ w_j^{(i)}) \right),
\]

\(\delta_j,\ell\) is defined in (10), and \(\xi_j^{(i),R}, \xi_j^{(i),L}\) are defined in (11).
We obtain a weak recovery of an orthogonal tensor decomposition with sample size $O(n^{k/2})$, which removes the $\polylog(n)$ factors in (Jain and Oh, 2014; Potechin and Steurer, 2017). Same as in the discussion for long matrix completion in Section 2.4, under the additional assumption that $\{\nu_i, i \in [r]\}$ are separated, with sample size $\omega(n^{k/2})$, weak consistency for $T$ can be achieved:

**Theorem 7** Let $T$ have an orthogonal decomposition as in (20), such that $\delta_{i,\ell} \geq c$ for some constant $c > 0$. For $i \in [k], j \in [r]$, define the estimator

$$\hat{T} = \sum_{j=1}^r \hat{v}_j \hat{w}_j^{(1)} \otimes \cdots \otimes \hat{w}_j^{(k)} \text{ for } \hat{w}_j^{(i)} = \zeta_j^{(i)}, \text{ and } \hat{v}_j = \sqrt{\lambda_j^{(1)}}.$$

Then, with probability at least $1 - cd^4 n^{-1/4}$, for large enough $d$,

$$\frac{\|T - \hat{T}\|_F}{\|T\|_F} \leq \frac{\sqrt{K_{\tau} K \tau}}{d} + C_1 \left(\frac{\tau}{d}\right)^{\ell-1}, \text{ where } \tau = \nu_i^{-2} \max_{i} \rho_{i}^{(i)}.$$

Theorem 7 is proved in Appendix I. To the best of our knowledge, this is the first result that shows weak recovery (in the sense $\|T - \hat{T}\|_F = o(1)$) in the regime $d \to \infty$, as opposed to $d \gtrsim \log(n)^c$.

### 2.5.3 Discussion on Tensor Completion

**Comparison to nearly square unfolding** Corollaries 5 and 6 provide an efficient algorithm that outputs an estimate for each component $w_{i}^{(0)}$. This directly implies that one can have a correlated estimation of $T$ with sample size $O(n^{k/2})$. This improves the sample complexity $O(n^{[k/2]})$ when directly applying the matrix completion algorithm in (Bordenave et al., 2022b) to unfold $[k/2],[k/2]$ $(T)$ when $k$ is odd. This mode-$i$ unfolding approach is more computationally efficient than the nearly square unfolding (Montanari and Sun, 2018) for component-wise estimation. For example, in the rank-1 case, applying the matrix completion algorithm to unfold $[k/2],[k/2]$ $(T)$, one can find a correlated estimation of $w_{(1)} \otimes \cdots \otimes w_{([k/2])}$. However, a further recursive unfolding is needed to recover individual components. The advantage of single-mode unfolding was also discussed in (Ben Arous et al., 2023) for the spiked tensor model.

**Optimal sample complexity for rank-one completion** We note that although our method does reach the $O(n^{k/2})$ scaling for rank-one tensor completion, this is not the optimal sample complexity. Indeed, for a boolean rank-one tensor, completion is equivalent to $k$-XORSAT, and hence (Creignou et al., 2003) implies the following theorem:

**Theorem 8** Assume that $T = x \otimes k$ for a given vector $x \in \mathbb{R}^n$ such that $\|x\|_2 = 1$. Then there exists a constant $c_k < 1$ such that if $p \geq c_k n^{-(k-1)}$, there exists an algorithm that outputs an estimator $\hat{x}$ for $x$ satisfying

$$\frac{\langle \hat{x}, x \rangle}{\|\hat{x}\|_2} \geq \frac{1}{\sqrt{n} \|x\|_{\infty}}.$$

Hence, for the specific case of rank-one tensors with bounded entries, a near-optimal algorithm can reach a sample complexity of $O(n)$. However, the mapping to a XORSAT problem disappears for tensors of rank $r > 1$, as well as with the addition of any type of noise (see, e.g., (Barak and Moitra, 2022)). Therefore, the conjectured $n^{k/2}$ threshold in (Barak and Moitra, 2022) for general low-rank tensor completion does not apply to this specific noiseless example. The proof of Theorem 8 is given in Appendix J.
Subspace recovery for general low-rank tensors Without the orthogonal decomposition assumption of the tensor $T$, one cannot directly associate the singular vectors of the unfolded matrix to the components of $T$. In this general case, the best one can hope for is to recover the subspace spanned by the spikes from $T$. Although our result is of independent interest in long matrix completion, it can also be used directly in non-orthogonal tensor completion. First, when $d \to \infty$, the results of (Montanari and Sun, 2018) imply that a good reconstruction of the left singular space of $M$ translates directly to weak consistency for tensor reconstruction. On the other hand, several algorithms such as the ones in (Montanari and Sun, 2018; Liu and Moitra, 2020; Cai et al., 2022) make use of a first spectral estimate of the CP factors of the tensor $T$ before a refined optimization step; a popular choice for such an estimate is usually a truncated SVD of $A$. We conjecture that even for fixed $d$ when the SVD of $A$ contains no information about the one of $M$, the embedded eigenvectors of $B$ can be used to perform an informed initialization of those algorithms.

2.6. Application in low-multilinear-rank tensor completion Besides CP-decomposition, another popular choice for tensor decomposition is the Tucker decomposition (Tucker, 1966), which is also called higher-order singular value decomposition (HOSVD) (De Lathauwer et al., 2000). For an order $k$-tensor $T$, we write its HOSVD as

$$T = \sum_{q_1=1}^{r_1} \sum_{q_2=1}^{r_2} \cdots \sum_{q_k=1}^{r_k} s_{q_1,\ldots,q_k} \phi_{q_1}^{(1)} \otimes \cdots \otimes \phi_{q_k}^{(k)},$$

where $(r_1,\ldots,r_k)$ is called the multilinear rank of $T$, $U_j = [\phi_1^{(j)},\ldots,\phi_r^{(j)}] \in \mathbb{R}^{n \times r_j}$ is a matrix with orthonormal columns, and $S \in \mathbb{R}^{r_1 \times \cdots \times r_k}$ is the core tensor. One can show that the columns of $U_j$ are the left singular vectors of the mode-$j$ unfolding of $T$ denoted by unfold$_j(T) \in \mathbb{R}^{n \times n^{k-1}}$. See (Kolda and Bader, 2009) for more details. Tensor completion with low-multilinear-rank was studied in (Kressner et al., 2014; Yuan and Zhang, 2017; Montanari and Sun, 2018). Recovering an HOSVD under additive noise was considered in (Lebeau et al., 2024).

We can apply the result for long matrix completion in Section 2.3 and to unfold$_j(T)$ to recover the mode-$j$ components of the HOSVD. We denote the SVD of unfold$_j(T)$ as

$$\text{unfold}_j(T) = \sum_{i=1}^{r_j} \nu_i^{(j)} \phi_i^{(j)} \psi_i^{(j)}.$$  

Let $B^{(j)}$ be the non-backtracking wedge operator for the observed matrix of unfold$_j(T)$ after sampling each entry with probability $p = \frac{d}{n^2 r_j}$. The following corollary follows directly from Theorem 3.

**Corollary 9 (Low-multilinear-rank tensor completion)** Let $T$ be a tensor defined in (21). With the same notations as in Section 2.2 and Theorem 3 applied to unfold$_j(T)$ and $B^{(j)}$, there exist constants $C_0, n_0 \geq 1$ that depend polynomially on $r, K, \kappa, d, \log n$ such that, if $n \geq n_0$ and $C_0 \tau^{2\ell} \leq 1$, the following holds on an event with probability at least $1 - cd^4 n^{-1/4}$: there exists a unit left singular vector $\phi_i^{(j)}$ of unfold$_j(T)$ associated with $\nu_i^{(j)}$ such that

$$\left| \langle \xi_i^R, \phi_i^{(j)} \rangle - \frac{1}{\sqrt{\gamma_i}} \right| \leq \frac{C_0 \tau^{2\ell}}{\delta_i,\ell} \quad \text{and} \quad \left| \langle \xi_i^L, \phi_i^{(j)} \rangle - \frac{1}{\sqrt{\gamma_i}} \right| \leq \frac{C_0 \tau^{2\ell}}{\delta_i,\ell}.$$
By applying Corollary 9 to different unfoldings, we can achieve weak recovery for the orthonormal components $U^{(j)}$, $1 \leq j \leq k$, with $O(n^{k/2})$ many samples. Our method, however, does not provide estimations for the core tensor $S$, the same as the work on low-multilinear-rank tensor approximation (Lebeau et al., 2024).

Organization of the appendix

Numerical simulations are provided in Appendix A. Appendix B is devoted to stating the main technical ingredient of Theorems 2 and 3. It consists of a detailed spectral structure of the matrix $B$ (or, more precisely, of its power $B^\ell$), with precise expressions given for the pseudo-eigenvectors of $B$ and their relationships with each other. We also provide a sketch of the proof of how this technical result implies our main theorems. The rest of the sections are then devoted to proving Theorem 10. In short:

(i) In Appendix C, we collect a few simple inequalities that will be useful in the proof.

(ii) In Appendix D, we perform a local study of the neighborhoods of the weighted bipartite random graph $G$; we show a local convergence (in a stronger sense than the one of Benjamini and Schramm (Benjamini and Schramm, 2001)) of $G$ to an appropriately defined Galton-Watson tree $T$. The structure of this tree is tailored to our bipartite setting and is similar to the one of (Florescu and Perkins, 2016). In short, the tree $T$ encodes the “local” properties of $G$, in the sense that most neighborhoods of $G$ have the same distribution as the tree $T$.

(iii) In Appendix E, we study this tree in-depth to extract information about specific processes on $T$. Those processes give rise to the pseudo-eigenvectors of $B$. The main ingredient of this Appendix is a reduction to the setting of (Stephan and Massoulié, 2022), which approximates the bipartite graph $G$ with a randomly weighted (but non-bipartite) version.

(iv) Appendices F and G are devoted to bounding the bulk eigenvalues of $G$. The method used – a so-called tangle-free decomposition and a variant of the trace method (Füredi and Komlós, 1981) – is similar to the one of (Bordenave et al., 2018, 2022b; Stephan and Massoulié, 2022; Stephan and Zhu, 2022). One of the major technical novelty is in the trace method proof of Theorem 2(ii). The $B$ matrix counts non-backtracking wedges, but the appearance of different wedges in a random graph $G$ is dependent. It is a big challenge to estimate the high-power trace of a random matrix with dependent entries compared to the standard moment method proof techniques in the random matrix theory literature, and such a technical challenge does not appear in (Bordenave et al., 2018, 2022b; Stephan and Massoulié, 2022). Additional proofs based on the trace method are included in Appendix H.

Finally, we include the proof of Theorem 7 in Appendix I, and the proof of Theorem 8 in Appendix J.

Acknowledgments

Y.Z. is partially supported by NSF-Simons Research Collaborations on the Mathematical and Scientific Foundations of Deep Learning and an AMS-Simons Travel Grant.
References


Appendix A. Numerical experiments

To illustrate our results, we considered a rectangular matrix \( M \) of size \( n \times m \), of the form

\[
M = \frac{1}{\sqrt{mn}} \left( \nu_1 \mathbf{1}_n \mathbf{1}_m^\top + \nu_2 \mathbf{u} \mathbf{v}^\top \right)
\]  

(22)

where \( \mathbf{u} \) and \( \mathbf{v} \) are vectors with entries in \( \{-1, 1\} \) orthogonal to \( \mathbf{1}_n \) and \( \mathbf{1}_m \), respectively. In this setting, the thresholds \( \theta_1 \) and \( \theta_2 \) are \( \theta_1 = \sqrt{\nu_1^2 + \nu_2^2} \) and \( \theta_2 = \frac{\nu_1 + \nu_2}{d} \). For all our experiments, we fixed \( d = 3, \nu_1 = 1, \nu_2 = 0.9 \), and \( m = n^2 \) to mimic the unfolding of a tensor of order 3.

The left part of Figure 2 shows the spectrum of the non-backtracking wedge matrix \( B \) constructed as in (3), with \( n = 4000 \). The structure shown in Theorem 2 is easily apparent: all but two of the eigenvalues are confined in a circle of radius close to \( \theta^2 \), with two real outlier eigenvalues around \( \nu_1^2 \) and \( \nu_2^2 \), respectively. In the right part, we plot the reduced eigenvector \( \zeta_L^\top \) associated with \( \lambda_2 \) for \( n = 30000 \) against the theoretical vector \( \mathbf{u} \). We can check that \( \langle \text{sign}(\zeta_L^\top), \mathbf{u} \rangle_n \approx 0.845 \), which corresponds to recovering 92.3\% of the entries of \( \mathbf{u} \). On the other hand, Figure 3 shows the same
Figure 2: **Left:** Spectrum of the non-backtracking wedge matrix of the matrix $M$ in (22). The bulk eigenvalues are in orange, while the outliers $\lambda_1$ and $\lambda_2$ are in green. The circle of radius $\theta^2$ and the theoretical locations at $\nu_2^1$ and $\nu_2^2$ are plotted in solid black and dashed light purple, respectively. **Right:** Reduced second eigenvector $\zeta_L^2$ of $B$ (see (11) for the embedding), in green. The theoretical singular vector $u$ is plotted in black. Both vectors are rescaled to have norm $\sqrt{n}$.

Figure 3: Spectrum density (left) and second eigenvector (right) of the matrix $Z = AA^T - \text{diag}(AA^T)$. The second eigenvector of $Z$ is normalized and plotted against the vector $u$ as in Figure 2.

plots for the matrix $Z = AA^T - \text{diag}(AA^T)$, widely used in tensor completion works (Montanari and Sun, 2018; Cai et al., 2022). The spectrum of $Z$ is less well-behaved than the one of $B$, and in particular, there is no visible outlier. This is confirmed by plotting the second eigenvector $\xi_2$ of $Z$ against $u$: the former is very localized, with no visible correlation with $u$. This time, the sign correlation between $\xi_2$ and $u$ is $\frac{\langle \text{sign}(\xi_2), u \rangle}{n} \approx 0.01 \approx \frac{1.65}{\sqrt{n}}$, which is indistinguishable from the variance of a random guess.
Appendix B. Spectral structure of the matrix $B$

In this section, we state the main intermediary result used in our paper, which characterizes almost fully the structure of the top eigenvectors of $B$. Theorems 2 and 3 then stem from a perturbation analysis already done in (Stephan and Massoulié, 2022; Bordenave et al., 2022b; Stephan and Zhu, 2022), which we will sketch here.

B.1. Preliminaries

We begin with a few important definitions and properties. Define the start and terminal matrices $S \in \mathbb{R}^{V_1 \times E}$ and $T \in \mathbb{R}^{E \times V_1}$ by

$$S_{xe} = 1_{e_1 = x} \quad \text{and} \quad T_{ex} = 1_{e_3 = x}.$$ 

This implies that for any vector $\phi \in \mathbb{R}^{V_1}$, $[S^* \phi](e) = \phi_{e_1}$ and $[T \phi](e) = \phi(e_3)$. We shall also need the so-called edge inversion operator $J$ and a diagonal weight operator $\Delta$, defined as

$$J_{e,f} = 1\{f = e^{-1}\} \quad \text{and} \quad \Delta_{e,e} = A_{e_1 e_2} A_{e_3 e_2},$$

where $e^{-1} = (e_3, e_2, e_1)$. As a shortcut, we will denote

$$S_\Delta = S\Delta, \quad T_\Delta = \Delta T, \quad J_\Delta = \Delta J = J\Delta$$

the weighted versions of $S, T, J$ respectively; and for any vector $\chi \in \mathbb{R}^E$, we let

$$\tilde{\chi} = J_\Delta \chi.$$ 

Those matrices are related by the following identities:

$$S = T^* J, \quad T = S^* J,$$

$$S_\Delta T = ST_\Delta = AA^*,$$  

$$B = TS_\Delta - J_\Delta.$$ 

In particular, those identities imply the following property, named the parity-time symmetry in (Bordenave et al., 2018):

$$J_\Delta B = B^* J_\Delta.\quad (27)$$

Since we aim to recover the left singular vectors of $M$, it will be useful to consider the “folded” equivalents of $M$ and $Q$, namely signal and variance matrices:

$$P = MM^* = \sum_{i=1}^r \nu_i^2 \phi_i \phi_i^* \quad \text{and} \quad \Phi = QQ^*$$

Finally, we define the oriented equivalents

$$\chi_i = T \phi_i$$

of the left singular vectors of $M$. 

22
B.2. Main technical results

We are now in a position to state our main technical result. Let

$$\ell = \lfloor \varepsilon \log_d(n) \rfloor \quad \text{with} \quad \varepsilon = \frac{(\eta/2) \wedge 1}{25}. $$

The candidate right (resp. left) eigenvectors of $B^\ell$ are the following: for $i \in [r_0]$,

$$u_i = \frac{B^\ell \chi_i}{\nu_i^{2\ell+2}} \quad \text{and} \quad \hat{u}_i = \left( B^* B^\ell \right) \frac{\chi_i}{\nu_i^{2\ell+2}}$$

with associated (pseudo)-eigenvalue $\nu_i^{2\ell}$. We shall collect those eigenvalues and eigenvectors in the matrices

$$\Sigma^\ell = \text{diag}(\nu_i^{2\ell}) \in \mathbb{R}^{r_0 \times r_0}, U, \hat{U} \in \mathbb{R}^{E \times r_0}.$$

Finally, we define for all $t \geq 0$ the matrix

$$\Gamma(t)_{ij} = \sum_{s=0}^t \left( 1, \Phi_s(\phi_i \circ \phi_j) \right) (\nu_i \nu_j d)^{2s}. $$

Then our results are as follows:

**Theorem 10** There exists an absolute constants $c, C_0 > 0$ such that for $n \geq C_0 K^{12}$, with probability at least $1 - cd^4 n^{-1/4}$, the following inequalities hold:

$$\left\| U^* U - d^2 \Gamma(t) \right\| \leq C_1 n^{-\varepsilon/2} \tau^{4\ell}, \quad \left\| U^* \hat{U} - \left( \Gamma(t+1) - I_{r_0} \right) \right\| \leq C_1 n^{-\varepsilon/2} \tau^{4\ell},$$

$$\left\| U^* \hat{U} - I_{r_0} \right\| \leq C_1 n^{-\varepsilon/2} \tau^{4\ell}, \quad \left\| \hat{U}^* B^\ell \hat{U} - \Sigma^\ell \right\| \leq C_1 n^{-\varepsilon/2} \tau^{4\ell} \theta^{2\ell},$$

$$\left\| B^\ell P_{\text{im}(\hat{U})^\perp} \right\| \leq C_2 \theta^{2\ell}, \quad \left\| P_{\text{im}(U)^\perp} B^\ell \right\| \leq C_2 \theta^{2\ell},$$

with $C_1 = c r \kappa d^6 \log(n)^{7/2}$ and $C_2 = c r \kappa d^6 K^{20} \log^{14}(n)$, where $P_{\text{im}(U)^\perp}$ is the projection operator onto the orthogonal complement of the vector space spanned by column vectors of $U$.

The statements (32) and (33) are proved in Section E.1, and (34) is proved in Section F.

B.3. Sketch of proof for Theorems 2 and 3

Theorems 2 and 3 stem from Theorem 10 via non-Hermitian perturbation arguments. Details can be found in (Stephan and Massoulié, 2022; Bordenave et al., 2022b). Following the steps in (Stephan and Massoulié, 2022, Section 5.2), the bounds in Theorem 10 imply that $B$ is nearly diagonalized by $U, \hat{U}$, in the sense that for a constant $C_3$ depending polynomially on $r, K, \kappa, d, \log n$,

$$\left\| B^\ell - U^* \Sigma^\ell \hat{U} \right\| \leq C_3 \theta^{2\ell}$$

and $U, \hat{U}$ are nearly orthogonal. In turn, an application of the Bauer-Fike Theorem (Bauer and Fike, 1960) yields that for $i \in [r_0]$,

$$\lambda_i^\ell = \nu_i^{2\ell} + C_4 \theta^{2\ell} = \nu_i^{2\ell} \left( 1 + C_4 \left( \frac{\theta}{\nu_i} \right)^{2\ell} \right),$$

$$C_3 = cr \kappa d^6 \left( \log(n)^{7/2} \right).$$
and taking the $1/\ell$-th power on both sides, we get

$$\lambda_i = \nu_i^2 \left( 1 + C_5 \left( \frac{\theta}{\nu_i} \right)^{2\ell} \right),$$

where $C_4, C_5$ depend polynomially on $r, K, \kappa, d, \log n$. All the other eigenvalues of $B$ satisfy

$$|\lambda| \leq C_5^{1/\ell} \theta^2.$$ 

This corresponds to Theorem 2.

Now, let $\sigma = C_4 \theta^2$. For $i \in [r_0]$, we define $M_i = \{ j \in [r_0] : \nu_j = \nu_i \}$, and the eigengap

$$\delta_i = \min_{j \neq M_i} |\nu_j - \nu_i|.$$

We will assume in the following that $\delta_i \geq 2\sigma$; otherwise, the bounds we show are trivial. Recall the definition of $u_i$ from (30), and the definition of $\chi_i$ from (29). By the proof of Theorem 8 in (Stephan and Massoulié, 2022), if $\xi_i$ is an eigenvector of $B$ associated to $\lambda_i$, there exists a vector $\tilde{u}_i$ in $\text{vect}(u_j : j \in I)$ such that

$$\|\xi_i - \tilde{u}_i\| \leq \frac{3\sigma}{\delta_i} \leq \frac{6\sigma}{\delta_i}.$$

From the definition of $u_i$, $i \in [r_0]$ in (30), by linearity, we can write

$$\tilde{u}_i = \frac{B^T \phi_i}{\nu_i^{2\ell}},$$

where $\phi_i \in \text{vect}(\phi_j : j \in I)$ is a unit vector. For simplicity, we assume $\phi_i = \phi_i$ (and hence $\tilde{u}_i = u_i$); the proof holds, mutatis mutandis, in the general case.

Note that using the definition of $S_\Delta$ from (23), $S_\Delta S_\Delta^*$ is a diagonal matrix such that

$$\langle S_\Delta S_\Delta^* xx \rangle = \sum_{e \in E : x = e_1} (A_{e_1 e_2} A_{e_3 e_2})^2 \leq \left( \frac{L}{d} \right)^4 \sum_{e_2 \in [m], e_3 \neq x} X_{xe_2} X_{e_1 e_2}.$$ 

Here $\sum_{e_2, e_3} X_{xe_2} X_{e_1 e_2}$ counts the number of wedges in $G$ starting from $x$, which is bounded by $cd^2 \log n$ for all $x \in [n]$ for an absolute constant $c > 1$ with probability at least $1 - n^{-1}$ from Proposition 16. Therefore $\|S_\Delta\| \leq cK \theta^2 \sqrt{\log n}$ and

$$\left\| S_\Delta \xi_i - S_\Delta \frac{u_i}{\|u_i\|} \right\| \leq \frac{c \sigma K^2 \theta^2 \sqrt{\log n}}{\delta_i}.$$

On the other hand, using Lemma 22 and (52) in Lemma 23, and recalling matrix relations an definitions in (23), (24), (29), we have

$$\langle S_\Delta u_i, \phi_i \rangle = \langle u_i, \chi_i \rangle = \nu_i^2 (1 + O(C_0 n^{-\varepsilon/2} \ell_i^{2\ell}))$$

$$\|S_\Delta u_i\|^2 = \nu_i^4 \left( \Gamma_i^{(\ell + 1)} + O(C_0 n^{-\varepsilon/2} u_i^{2\ell}) \right).$$
for a constant $C_0$ depending polynomially on $r, K, \kappa, d, \log n$. From (50) in Lemma 23,

$$\|u_i\|^2 = d^2 \Gamma_{ii}^{(t)} + O(C_0\ell^{-2}r_i^{4\ell}).$$

Since $\Phi$ is a positive matrix with spectral radius $\rho^2$, from (31),

$$\Gamma_{ii}^{(t)} = \left(1, \left(I - \frac{\Phi}{\nu_i^2d^2}\right)^{-1}(\phi_i \circ \phi_i)\right) + O(\tau_i^{4\ell}) = \gamma_i + O(\tau_i^{4\ell}).$$

Putting all of the previous bounds together, recalling the definition of $\zeta^R_i$ from (11), we do find

$$\left\langle \frac{\zeta_i}{\|\zeta_i\|}, \phi_i \right\rangle = \left\langle \frac{S\Delta \xi_i}{\|S\Delta \xi_i\|}, \phi_i \right\rangle = \frac{1}{\sqrt{\gamma_i}} + O\left(\frac{2K^2d\sqrt{\log n} \sigma}{\delta_i}\right),$$

which is the statement of Theorem 3 upon diving both $\delta_i$ and $\sigma$ by $\nu_i^{2\ell}$. The same holds for the left eigenvectors by replacing $u_i$ by $\hat{u}_i$ in the proof.

**Appendix C. Parameter inequalities**

In this section, we collect a few simple inequalities that will be useful in the proof. First, since

$$\rho = \|Q\| \leq \sqrt{nm} \max_{xy} |Q_{xy}|,$$

we have $K \geq 1$. Further, using the variational formula for singular values,

$$\rho \geq \frac{1}{\sqrt{nm}} \langle 1, Q1 \rangle = \frac{1}{\sqrt{nm}} \sum_{xy} Q_{xy} = \sum_{x,y} M^2_{xy} = \|M\|_F^2 \geq \nu_1^2.$$

Hence

$$1 \leq K = \frac{L^2}{\rho} \leq \frac{L^2}{\nu_1^2}.$$

Next, we derive some entrywise bound on $\Phi := QQ^*$. For any $x, y$,

$$\Phi_{xy} = (QQ^*)_{xy} = \sum_z Q_{xz}Q_{yz} \leq \frac{K^2\rho^2}{n}. \quad (35)$$

Let $(v_k)$ be a set of orthonormal left singular vectors for $Q$ with singular values $\mu_1, \ldots, \mu_n$. For $t \geq 1$,

$$(\Phi^t)_{xy} = [(QQ^*)^t]_{xy} = \sum_k \mu_k^{2t} v_k(x)v_k(y) \quad (36)$$

$$\leq \rho^{2t-2} \sum_k \mu_k^2 \|v_k(x)||v_k(y)|| \leq \rho^{2t-2} \sqrt{(\Phi)_{xx}} \sqrt{(\Phi)_{yy}} \leq \frac{K^2\rho^{2t}}{n}, \quad (37)$$

$$\leq \rho^{2t-2} \sqrt{(\Phi)_{xx}} \sqrt{(\Phi)_{yy}} \leq \frac{K^2\rho^{2t}}{n}, \quad (38)$$
where the last line follows from Cauchy-Schwartz inequality and (35). Similarly, let \( \tilde{\Phi} = Q^*Q \), we have
\[
(\tilde{\Phi}^t)_{xy} \leq \frac{K^2 \rho^{2t}}{m}. \tag{39}
\]
And
\[
(\Phi^tQ)_{xy} \leq \frac{K^2 \rho^{2t+1}}{\sqrt{nm}}, \quad (\tilde{\Phi}^tQ^*)_{xy} \leq \frac{K^2 \rho^{2t+1}}{\sqrt{nm}}. \tag{40}
\]
Moreover, for any \( t \geq 1 \) and \( v, w \in \mathbb{R}^n \),
\[
\langle v, \Phi^t w \rangle = \sum_{xy} (\Phi^t)_{xy} v(x) w(y) \leq \frac{K^2 \rho^{2t}}{n} \|v\|_1 \|w\|_1. \tag{41}
\]

Appendix D. Local study of \( G \)

D.1. Preliminaries

Rooted graphs and trees

We begin with a rigorous definition for the objects considered in this section:

**Definition 11 (Labeled rooted graphs)** A labeled rooted graph is a triplet \( g^\star = (g, o, \iota) \) that consists in:

- a graph \( g = (V, E) \),
- a distinguished vertex (or root) \( o \in V \),
- a labeling function \( \iota : V \to \mathbb{N} \).

The set of labeled rooted graphs is denoted by \( G^\star \).

When the labeling function \( \iota \) is clear, we will use the shortcut \((g, o)\) for \( g^\star \). In particular, this will be the case whenever \( V \subseteq \mathbb{N} \) and \( \iota(x) = x \). The notion of induced subgraph extends naturally to rooted and/or labeled graphs, provided that the induced subgraph does contain the root \( o \).

An important object of study will be the bipartite Galton-Watson tree defined as follows:

**Definition 12 (Bipartite Galton-Watson tree)** The bipartite Galton-Watson tree \( (T, \rho) \) is a random tree \((T, o, \iota)\) defined as follows:

- the root of the tree (depth 0) has label \( \iota(o) = \rho \),
- the number of children of each vertex \( x \in T \) is independent, and is defined as follows:
  - if \( x \) has even depth \( 2h \), \( x \) has \( \text{Poi}(d^2) \) children at depth \( 2h + 1 \),
  - if \( x \) has odd depth \( 2h + 1 \), \( x \) has exactly one child at depth \( 2h + 2 \).
- each vertex \( x \) of even depth has a label \( \iota(x) \sim \text{Unif}([n]) \), and each vertex of odd depth has a label \( \iota(x) \sim \text{Unif}([m]) \).
Tangle-free neighbourhoods and graphs. For a rooted graph \((g, o)\), we denote by \((g, o)_t\) the subgraph induced by the vertices within distance \(t\) from \(o\). The boundary of this subgraph, spanned by the vertices at distance exactly \(t\) from \(o\), will be denoted by \(\partial(g, o)_t\). The notion of tangle-freeness, which will be crucial in our analysis, is defined as follows:

**Definition 13 (Tangle-freeness)** A rooted graph \((g, o)\) is said to be tangle-free if it contains at most one cycle, otherwise, it is called tangled. For a given \(t \in \mathbb{N}\), a graph \(g\) is said to be tangle-free if for any vertex \(x \in g\), the neighborhood \((g, x)_t\) is tangle-free.

D.2. Bounding the neighborhood sizes

The main issue in bounding the neighborhood sizes is that a typical left vertex \(x\) has approximately \(mp = d\sqrt{\alpha}\) neighbors, which diverges as \(\alpha \to \infty\). However, most of those neighbors have degree one and thus do not matter in the spectrum of \(B\)! We formalize this by defining the bipartite graph \(\tilde{G} = (V_1, \tilde{V}_2, \tilde{E})\) induced by \(V_1\) and the non-leaf vertices in \(V_2\):

\[\tilde{V}_2 := \{y \in V_2 : \deg(y) \geq 2\}.\]

We are then able to show the following proposition:

**Proposition 14** There exists constants \(c_0, c_1 > 0\) such that for all \(x \in [n]\) and \(s > 0\),

\[\mathbb{P}(\forall t \geq 0, \quad |\partial(\tilde{G}, x)_{2t}| \leq sd^{2t} \quad \text{and} \quad |\partial(\tilde{G}, x)_{2t+1}| \leq sd^{2t+2} \geq 1 - c_0e^{-c_1s}\]

**Proof** We define

\[\tilde{p} = \mathbb{P}(\text{Bin}(n, p) \geq 1) = 1 - (1 - p)^n, \quad \tilde{d}_1 = mp\tilde{p}, \quad \text{and} \quad \tilde{d}_2 = \tilde{d}_1(1 + np).\]

We shall actually show the following bound: for some constants \(c_0, c_1, c_2, c_3, c_4\), and \(s > 1\),

\[\mathbb{P}(\forall t \geq 0, \quad |\partial(\tilde{G}, x)_{2t}| \leq c_1s\tilde{d}_1^2 \quad \text{and} \quad |\partial(\tilde{G}, x)_{2t+1}| \leq c_2s\tilde{d}_1\tilde{d}_2 \geq 1 - c_3e^{-c_4s} \quad (42)\]

We first show how this implies Proposition 14: for \(s > 1\) the bound is vacuous as soon as \(t > c\log_d(n)\) for some constant \(c\), so we can always assume otherwise. Second, we have

\[\tilde{d}_1 = d^2 + O\left(d^3/\sqrt{\alpha}\right) \quad \text{and} \quad \tilde{d}_2 = d^2 + O\left(d^3/\sqrt{\alpha}\right),\]

and hence for the timescale considered for \(t\) we have \(\tilde{d}_2 \leq cd^{2t}\) under the assumption (9). Finally, we can adjust \(c_0\) and \(c_1\) such that \(c_0e^{-c_1} \geq 1\) and hence the bound also becomes vacuous when \(s \leq 1\).

We now move to prove (42). Define the following sequences:

\[\varepsilon_{2t} = \max\left(\tilde{d}_1^{-1}\tilde{d}_2^{-t+1}(2t), \sqrt{mp\tilde{d}_1^{-1/2}\tilde{d}_2^{-(t-1)/2}}\sqrt{2t}\right),\]

\[\varepsilon_{2t+1} = \tilde{d}_2^{-t/2}\sqrt{2t + 1} \quad \text{and} \quad f_t = \prod_{t'=1}^{t} (1 + \varepsilon_{t'}).\]

Due to Assumption (8), by definition of \(\varepsilon_t, f_t\), there exists some constants \(c_0, c_1, c_2\) such that

\[c_0 \leq f_t \leq c_1 \quad \text{and} \quad \varepsilon_t \leq c_1.\]
We also define the following upper-bound sequence:
\[ S_{2t} = f_{2t} \tilde{d}_2 \quad \text{and} \quad S_{2t+1} = f_{2t+1} \tilde{d}_1 \tilde{d}_2, \]
which satisfies the following recursion equation:
\[ S_{2t+1} = (1 + \varepsilon_{2t+1}) \tilde{d}_1 S_{2t} \quad \text{and} \quad S_{2t+2} = (1 + \varepsilon_{2t+2})(1 + np)S_{2t+1}. \]

We will also need the following lemma, which stems from a classical Chernoff bound:

**Lemma 15** Let \((Y_1, \ldots, Y_T)\) be i.i.d binomial random variables, with mean \(\mu\). Then, for any \(u \geq 0\),
\[ \mathbb{P} \left( \sum_{i=1}^{T} Y_i \geq (1 + u)\mu T \right) \leq e^{-T\mu \gamma(u)}, \]
where \(\gamma(u) = (1 + u) \log(1 + u) - u\). The function \(\gamma(u)\) is strictly increasing, and satisfies the inequality
\[ \gamma(u) \geq \min \left( \frac{u^2}{4}, u \right) \tag{43} \]

For a given vertex \(x \in \partial(G, x)_{2t}\), the number of neighbors of \(x\) in \(V_2 \setminus (\bar{G}, x)_{2t}\) is dominated by \(\text{Bin}(m, np)\). For any \(t \geq 0\), we can apply Lemma 15 with \(u = \varepsilon_{2t+1}\) to get
\[ \mathbb{P}(|\partial(G, x)_{2t+1}| \geq sS_{2t+1} \mid |\partial(G, x)_{2t}| \leq sS_{2t}) \leq \exp \left( -s f_{2t+1} \tilde{d}_1 \tilde{d}_2 \gamma(\varepsilon_{2t+1}) \right). \]

Equation (43) implies that when \(u \leq c_1, \gamma(u) \geq c_2 u^2\) for some constant \(u^2\), and hence
\[ \mathbb{P}(|\partial(G, x)_{2t+1}| \geq sS_{2t+1} \mid |\partial(G, x)_{2t}| \leq sS_{2t}) \leq \exp \left( -c s(2t + 1) \right). \]

Now, the number of neighbors of a vertex in \(\partial(G, x)_{2t+1}\) in \(V_1\) is dominated by a \(X \sim \text{Bin}(n, p)\) random variable, conditioned on \(X \geq 1\), which is itself dominated by a \(1 + \text{Bin}(n, p)\) variable (see, e.g., (Broman et al., 2011, Proposition 1.1)). This time, we apply Lemma 15 by noticing that
\[ \mathbb{P} \left( \sum_{i=1}^{T} (1 + Y_i) \geq (1 + u)(1 + \mu)T \right) = \mathbb{P} \left( \sum_{i=1}^{T} Y_i \geq \left( 1 + u + \frac{u}{\mu} \right) \mu T \right), \]
so that
\[ \mathbb{P}(|\partial(G, x)_{2t+2}| \geq sS_{2t+2} \mid |\partial(G, x)_{2t+1}| \leq sS_{2t+1}) \leq \exp \left( -snp f_{2t+1} \tilde{d}_1 \tilde{d}_2 \gamma(\varepsilon_{2t+2}/np) \right). \]

We can now notice that by the bound on \(\gamma\) in (43) and definition of \(\varepsilon_{2t+2}\), we have
\[ np \tilde{d}_1 \tilde{d}_2 \gamma(\varepsilon_{2t+2}/np) \geq c(2t + 2). \]

Hence, we have shown that for any \(t \geq 0\),
\[ \mathbb{P}(|\partial(G, x)_{t+1}| \geq sS_{t+1} \mid |\partial(G, x)_t| \leq sS_t) \leq \exp(-cs(t + 1)). \]
This implies that

\[ P(\exists t \geq 0, |\partial(\bar{G}, x_t)| \geq S_t) \leq \sum_{t \geq 0} \exp(-cs(t + 1)) = \frac{e^{-cs}}{1 - e^{-cs}}, \]

from which (42) ensues since \( s > 1 \).

With the same arguments as in (Bordenave et al., 2018, Lemma 29), this tail bound implies the following proposition:

**Proposition 16** With probability at least \( 1 - 1/n \), we have for all \( x \in [n] \),

\[ |(\bar{G}, x)_{2t}| \leq c \log(n)d^{2t} \quad \text{and} \quad |(\bar{G}, x)_{2t+1}| \leq c \log(n)d^{2t+2}. \]

Further, we have for all \( x \in V_1, p \geq 1 \),

\[ \mathbb{E}[|(\bar{G}, x)_{2t}|^p]^{1/p} \leq cpd^{2t}, \]

and

\[ \mathbb{E}\left[\max_{x \in [n], t \geq 0} |(\bar{G}, x)_{2t}|^p\right] \leq (cp)^p + (c \log(n))^p. \]

**D.3. Tree approximation**

We now couple the neighborhoods of \( G \) with an appropriately defined tree process. Central to this section is the following breadth-first exploration process of a vertex \( x \in [n] \):

- start with \( A_0 = \{x\} \)
- at step \( t \), take any vertex \( x_t \in A_t \) among those closest to \( x \)
  - if \( x_t \in V_2 \), we let \( N_t \) be the set of all neighbors of \( x_t \) in \( V_1 \setminus \bigcup_{s \leq t} A_s \)
  - if \( x \in V_1 \), \( N_t \) is the set of neighbors of \( x_t \) in \( V_2 \setminus \bigcup_{s \leq t} A_s \) that have a neighbor in \( V_1 \setminus \bigcup_{s \leq t} A_s \).
- update \( A_{t+1} = A_t \cup N_t \setminus \{x_t\} \) and continue.

We denote by \( \mathcal{F}_t \) the filtration adapted to this process and call an edge \( \{e_1, e_2\} \) a discovered edge if for some \( t \geq 0 \) we have \( x_t = e_1 \) and \( e_2 \in N_t \). We begin with a result about the tree-like properties of neighborhoods in \( G \):

**Proposition 17** For a given vertex \( x \), and \( h \geq 0 \), we have

\[ P((\bar{G}, x)_{2h} \text{ contains a cycle}) \leq \frac{cd^{2h+2}}{n}. \]

Further, with probability at least \( 1 - cd^{4h+4}/n \), the graph \( G \) is \( h \)-tangle-free: for every vertex \( x \), \( (\bar{G}, x)_{2h} \) contains at most one cycle.
Proof Consider the exploration process outlined above, and let \( \tau \) be the first time at which all the vertices in \((G, x)_h\) have been revealed. By construction, \( \tau \) is a stopping time for the filtration \( \mathcal{F}_t \), and given \( \mathcal{F}_\tau \), the discovered edges of \((G, x)_h\) form a spanning tree. Hence, there are two possibilities for a cycle in \((G, x)_h\):

- there is an edge between a vertex of \((G, x)_h \cap V_1 \) and \((G, x)_h \cap V_2\),
- there exists a vertex \( z \in V_2 \setminus (G, x)_h \) and two vertices \( x_1, x_2 \in (G, x)_h \cap V_1 \) such that \( z \) is connected to both \( x_1 \) and \( x_2 \).

The number of such events of the first type is stochastically dominated by a Bin\(|(G, x)_h|^2, p\) variable and the second by a Bin\(|(G, x)_h|^2, mp^2\). Further, given \( \mathcal{F}_\tau \), those two types of events are independent. Recall the following simple bounds:

\[
\mathbb{P}(\text{Bin}(N, q) \geq 1) \leq Nq \quad \text{and} \quad \mathbb{P}(\text{Bin}(N, q) \geq 2) \leq N^2q^2.
\]

Then,

\[
\mathbb{P}((G, x)_h \text{ contains a cycle } | \mathcal{F}_\tau) \leq |(G, x)_h|^2(mp^2 + p).
\]

and similarly

\[
\mathbb{P}((G, x)_h \text{ is tangled } | \mathcal{F}_\tau) \leq |(G, x)_h|^4(m^2p^4 + p^2).
\]

Taking expectations in both cases and using Proposition 16,

\[
\mathbb{P}((G, x)_{2h} \text{ contains a cycle}) \leq \frac{cd^{2h+2}}{n} \quad \text{and} \quad \mathbb{P}((G, x)_{2h} \text{ is tangled}) \leq \frac{cd^{4h+4}}{n^2}.
\]

The second statement of the proposition easily ensues from a union bound. \(\blacksquare\)

We now prove a coupling result between the neighborhoods of \( G \) and the Galton-Watson trees introduced in Definition 12. Recall that the total variation between two probability measures \( \mathbb{P}_1, \mathbb{P}_2 \) is defined as

\[
d_{TV}(\mathbb{P}_1, \mathbb{P}_2) = \min_{\pi \in \Pi(\mathbb{P}_1, \mathbb{P}_2)} \mathbb{P}(X_1 \neq X_2),
\]

where \( \Pi(\mathbb{P}_1, \mathbb{P}_2) \) is the set of all couplings of \( \mathbb{P}_1 \) and \( \mathbb{P}_2 \), i.e. joint distributions of random variables \( (X_1, X_2) \) such that \( X_1 \sim \mathbb{P}_1 \) and \( X_2 \sim \mathbb{P}_2 \).

**Proposition 18** There exists a constant \( c \geq 0 \) such that for any \( h \geq 0 \), \( x \in V_1 \), we have

\[
d_{TV}(\mathcal{L}((G, x)_h), \mathcal{L}((T, x)_h)) \leq c \log(n)^2 \frac{d^{4h+3}}{n} \wedge \frac{1}{\sqrt{\alpha}}.
\]

**Proof** We proceed in two steps: we first couple the unlabeled versions of the graphs and then move on to the labels. We shall use the following classical bounds for total variation distances:

\[
d_{TV}(\text{Bin}(n, p), \text{Poi}(np)) \leq p \quad \text{and} \quad d_{TV}(\text{Poi}(\lambda), \text{Poi}(\lambda')) \leq |\lambda - \lambda'|. \quad (44)
\]

Firstly, from Proposition 17, \((G, x)_h\) is a tree with probability \( 1 - cd^{2h+2}/n \), so we only need to couple the offspring distributions. Consider the exploration process \((x_t)_{t \geq 0}\) defined above; at each step \( t \), we let \( n_t \) (resp. \( m_t \)) be the number of vertices in \( V_1 \setminus \bigcup_{t' \leq t} A_{t'} \) (resp. \( V_2 \setminus \bigcup_{t' \leq t} A_{t'} \)). We aim to couple it with the same exploration process \((x'_t)_{t \geq 0}\) on \((T, x)\). Let \( \mathbb{P}_t \) (resp. \( \mathbb{Q}_t \)) be the offspring distribution of \( x_t \) (resp. \( x'_t \)). By definition, we have an explicit definition for \( \mathbb{P}_t \) and \( \mathbb{Q}_t \):
• if $x_t$ has even depth, $\mathbb{P}_t = \text{Bin}(m_t, p\tilde{m}_t)$ where $\tilde{p}_t = \mathbb{P} (\text{Bin}(n_t, p) \geq 1)$, and $\mathbb{Q}_t = \text{Poi}(d^2)$;

• if $x_t$ has odd depth, $\mathbb{P}_t = \text{Bin}(n_t, p)$ conditioned on being at least 1, and $\mathbb{Q}_t = 1$ a.s.

We begin with the second case, as it is the simplest. As before, $\mathbb{P}_t$ is dominated by a distribution equal to $1 + \text{Bin}(n, p)$, and hence

$$\mathbb{P}_t(|N_t| > 1) \leq \mathbb{P}(\text{Bin}(n, p) \geq 1) \leq np = d\alpha^{-1/2}.$$ 

We now move to the first case. We have

$$\tilde{p}_t = 1 - (1 - p)^n_t = pn_t + O(p^2n^2) = O(np).$$

Using the bounds in eq. (44),

$$d_{TV}(\mathbb{P}_t, \mathbb{Q}_t) \leq p\tilde{p}_t + |m_t p\tilde{p}_t - d^2| \leq \frac{cd^2}{m} + |m_t n_p^2 - d^2| + \frac{cd^2}{\sqrt{\alpha}}.$$ 

The middle term can be further decomposed by noticing that $nmp^2 = d^2$:

$$|m_t n_p^2 - d^2| \leq |m_t - m|n_p^2 + m|n_t - n|p^2 \leq \frac{2d^2|(G, x)_{2h}|}{n}$$

All in all, we showed that for any $t \geq 0$,

$$d_{TV}(\mathbb{P}_t, \mathbb{Q}_t) \leq cd^2 \left(\frac{|(G, x)_{2h}|}{n} \vee \frac{d}{\sqrt{\alpha}}\right).$$

Let $\hat{L}$ denote the unlabeled distribution of the graphs. By the high-probability bounds of Proposition 16, we obtain

$$d_{TV}(\hat{L}, (G, x)_{2h}) \leq \frac{cd^{2h+2}}{n} + \frac{1}{n} \text{ Prop. 16 does not hold}$$

$$+ c \log(n)^2d^{2h+2} \left(\frac{d^{2h}}{n} \vee d\alpha^{-1/2}\right)$$

$$\leq c' \log(n)^2d^{4h+3} \frac{n^\vee \sqrt{\alpha}}{n},$$

where in the third part of the first inequality, we apply (45) with a union bound over all vertices in $(G, x)_{2h}$. This bound is lower than the one in Proposition 18 since $d > 1$.

Consider now the distribution of the labels. Given a coupling between the unlabeled versions of $(G, x)_{2h}$ and $(T, x)_{2h}$, the labels of the even (resp. odd) depth vertices of $(G, x)_{2h}$ are obtained by sampling without replacement from $[n]$ (resp. $[m]$), whereas those of $(T, x)_{2h}$ are sampled with replacement. From (Freedman, 1977), the total variation distance between sampling with and without replacement $k$ elements from a population of $N$ elements is bounded above by $k^2/N$, and hence

$$d_{TV}(\mathbb{P}, \mathbb{Q}) \leq 2\frac{|(G, x)_{2h}|^2}{n},$$

where $\mathbb{P}, \mathbb{Q}$ are the labels distributions for $(G, x)_{2h}, (T, x)_{2h}$, respectively. A final application of Proposition 16 completes the proof. 

31
D.4. Quantitative bounds for local functionals

The bounds obtained in Proposition 18 can be viewed as a quantitative version of the local convergence of Benjamini and Schramm (Benjamini and Schramm, 2001), for possibly diverging depth. It is known that this coupling also implies weak convergence in the probability measure sense, i.e., the convergence of graph functionals. We now provide a quantitative version of this weak convergence, but only for a specific class of functionals, called local functionals:

**Definition 19 (Local functionals)** A functional $f : G_* \rightarrow \mathbb{R}$ is said to be $t$-local if $f(G,o)$ is only a function of $(G,o)_t$.

**Proposition 20** Let $f : G_* \rightarrow \mathbb{R}$ be a $2h$-local function for some $h \geq 0$, such that $f(g,o) \leq a |(g,o)_{2h}|^b$ for some $a, b$. Then with probability at least $1 - n^{-1}$,

$$\sum_{x \in V_1} f(G,x) - \sum_{x \in [n]} \mathbb{E}[f(T,x)] \leq ca \log(n)^{3/2+b} d^{2h(1+b)} \sqrt{n} \wedge \alpha^{1/4}.$$

**Proof** Notice that the bounds on the neighborhood sizes match exactly the ones of (Bordenave et al., 2022b), and hence we can borrow the following concentration result from it (see the proof of Theorem 12.5 in (Bordenave et al., 2022b)): with probability at least $1 - n^{-1}$,

$$\sum_{x \in V_1} f(G,x) - \sum_{x \in [n]} \mathbb{E}[f(G,x)] \leq ca \log(n)^{3/2+b} d^{2h(1+b)} \sqrt{n} \wedge \alpha^{1/4}. \quad (46)$$

On the other hand, we have for any $x \in V_1$

$$|\mathbb{E}f(T,x) - \mathbb{E}f(G,x)| = \mathbb{E}\left[|f(T,x) - f(G,x)| \mathbb{1}_{\mathcal{E}(x)}\right],$$

where $\mathcal{E}(x)$ denotes the event of the coupling between $(G,x)_{2h}$ and $(T,x)_{2h}$ fails. From the Cauchy-Schwarz inequality,

$$|\mathbb{E}f(T,x) - \mathbb{E}f(G,x)| \leq \sqrt{\mathbb{P}(\mathcal{E}(x))} \sqrt{\mathbb{E}[(f(T,x) - f(G,x))^2]} \leq \sqrt{d_{TV}(\mathcal{L}((G,x)_{2h}), \mathcal{L}((T,x)_{2h}))} \left( \sqrt{\mathbb{E}[|f(G,x)|^2]} + \sqrt{\mathbb{E}[|f(T,x)|^2]} \right).$$

Proposition 18 bounds the first factor. It is easy to check that the bounds for $|(G,x)_{2h}|$ in Proposition 16 also apply to $|(T,x)_{2h}|$, which yields

$$|\mathbb{E}f(T,x) - \mathbb{E}f(G,x)| \leq ca \log(n) d^{2h+1} \frac{d^{2h}}{\sqrt{n} \wedge \alpha^{1/4}}. \quad (47)$$

Combining (46) and (47) yields the desired bound. \hfill \blacksquare
Appendix E. From tree functionals to pseudo-eigenvectors

Since we now have a way to translate quantities from the Galton-Watson tree $T$ to the graph $G$, we aim to apply these results to specific functionals. We begin with a few preliminaries: given a functional $f : G_* \to \mathbb{R}$, define its Galton-Watson transform $\mathcal{F}$ as

$$\mathcal{F}(x) = E[f(T, x)].$$

This definition is consistent since from Definition 12, the distribution of a Galton-Watson tree only depends on its root label $x$. We shall also need the following definition: given a matrix $W \in \mathbb{R}^{n \times m}$, and a functional $f : G_* \to \mathbb{R}$, define $\partial W f$ as

$$\partial W f(g, o) = \sum_{o', o'' \in N_2(o)} W_{i(o) ; i(o')} W_{i(o') ; i(o'')} f(g \setminus \{o, o'\}, o'').$$

where $o', o''$ runs over the set of non-backtracking wedges starting from $o$. In particular, when $g$ is a rooted tree, $o'$ is uniquely determined by $o''$, and the graph $(g \setminus \{o, o'\}, o'')$ is the subtree rooted at $o''$.

The functionals we will consider are of the following form: for a vector $\phi$,

$$f_{\phi, t}(g, o) = \left(\frac{mn}{d^2}\right)^t \sum_{o_1, \ldots, o_{2t}} \left( \prod_{s=0}^{t-1} M_{i(o_{2s}) ; i(o_{2s+1})} M_{i(o_{2s+2}) ; i(o_{2s+1})} \right) \phi_{i(o_{2t})}$$

where the sum runs over all non-backtracking paths of length $2t$ starting from $o$. Then, the following proposition holds:

**Proposition 21** For any $t \geq 0$, and $i, j \in [r]$,

$$\mathcal{F}_{\phi_i, t} = \nu_i^{2t} \phi_i,$$

$$\mathcal{F}_{\phi_i, t} f_{\phi_j, t} = (\nu_i \nu_j)^{2t} \left( \sum_{s=0}^{t} \frac{\Phi^s}{(\nu_i \nu_j d)^{2s}} \right) (\phi_i \circ \phi_j)$$

and if $F_{\phi_i, t} = (f_{\phi_i, t+1} - \nu_i^2 f_{\phi_i, t})^2$, then

$$F_{\phi_i, t} = \frac{\Phi^t(\phi_i \circ \phi_i)}{d^{2t}}$$

Additionally, for any functional $f : G_* \to \mathbb{R}$, and any weight matrix $W$,

$$\partial W f = \frac{d^2(W^* W)}{mn} \mathcal{F}.$$

**Proof** We proceed by reduction to (Stephan and Massoulié, 2022). Let $\tilde{T}$ be an ordinary Galton-Watson tree with offspring distribution $\text{Poi}(d^2)$, such that each vertex has a random label $\text{Unif}([n])$, and define the random weights

$$\tilde{W}_{xy} = \frac{mn}{d^2} M_{xZ} M_{yZ}, \quad Z \sim \text{Unif}([m])$$
The equivalent tree functional $\tilde{f}_{\phi,t}$ is defined as

$$\tilde{f}_{\phi,t}(g,o) = \sum_{o_1,\ldots,o_t} \left( \prod_{s=0}^{t-1} \tilde{W}_{s(o_s),s(o_{s+1})} \right) \phi_{i(o_t)}.$$ 

It is easy to see that $f_{\phi,t}(T,x) \overset{d}{=} \tilde{f}_{\phi,t}(\tilde{T},x)$.

On the other hand, the process $\tilde{f}$ on $(\tilde{T},x)$ was already studied in depth in (Stephan and Massoulié, 2022). It corresponds to the signal matrix

$$\tilde{Q} = \frac{d^2 \mathbf{1}\mathbf{1}^\top}{n} \circ \left( \frac{mn}{d^2} \cdot \frac{MM^*}{m} \right) = P$$

and the variance matrix

$$\tilde{K} = \frac{d^2 \mathbf{1}\mathbf{1}^\top}{n} \circ \left( \frac{m^2 n^2 (M \circ M)(M \circ M)^*}{d^4} \right) = \frac{\Phi}{d^2}.$$ 

Proposition 21 then ensues from Propositions 6 and 7 in (Stephan and Massoulié, 2022).

### E.1. Pseudo-eigenvectors

We now make use of Proposition 20 to translate the expectations on $T$ to pseudo-eigenvectors of $B$.

This corresponds to Equations (32)-(33) of Theorem 10. We shall prove stronger versions of those bounds in the following two lemmas and show how they imply the desired inequalities.

**Lemma 22** Let $h = c \log_d(n)$. With probability at least $1 - c' d^4 n^{12c-1}$, for any $t \leq 3h$, we have

$$|\langle B^t \chi_i, \tilde{\chi}_j \rangle - \nu_i^{2t+2} \delta_{ij}| \leq c' \kappa^2 \theta_{2t+2}^2 d^{12h+4} \frac{\log(n)^{5/2}}{\sqrt{n} \wedge \alpha^{1/4}}.$$ 

**Proof** Consider the following functional:

$$f(g,o) = 1_{(g,o)\in T} \text{ is tangle-free } \phi_j(o) f_{\phi_i,t+1}(g,o),$$

where $f_{\phi_i,t}$ was defined in (48). It is easy to check that $f$ is $(2t+2)$-local, and when $(g,o)$ is $3h$-tangle-free (which happens with probability at least $1 - c'd^4 n^{12c-1}$ from Proposition 17), there are at most two non-backtracking paths between $o$ and any vertex of $(g,o)$, hence from (48),

$$|f(g,o)| \leq 2 \frac{\kappa^2}{n} \theta_{2t+2}^2 |(g,o)_{2t+2}|.$$ 

We are, therefore, in the setting of Proposition 20. It remains to compute the corresponding quantities on $(G,x)$ and $(T,x)$. First, by definition,

$$\sum_{x \in [n]} f(G,x) = \langle B^t \chi_i, \tilde{\chi}_j \rangle,$$
and by Proposition 21 we have
\[ \sum_{x \in [n]} f(T, x) = \sum_{x \in [n]} \phi_j(x) \nu_i^{2t+2} \phi_i(x) = \nu_i^{2t+2} \delta_{ij}, \]
which completes the proof.

Recall the definition of \( \Gamma_{ij}^{(t)} \) from (31). The following bounds hold:

**Lemma 23** Let \( h = c \log_d(n) \). With probability at least \( 1 - \epsilon d^4 n^{4c-1} \), for any \( t \leq h \), we have
\[
\left| \langle B^t \chi_i, B^t \chi_j \rangle - \nu_i^{2t} \nu_j^{2t} d^2 \Gamma_{ij}^{(t)} \right| \leq c d^6 \kappa^2 \log(n)^{7/2} \theta_2^4 \frac{d^6t}{\sqrt{n} \wedge \alpha^{1/4}},
\]
\[
\left| \langle (B^*)^t \chi_i, (B^*)^t \chi_j \rangle - \nu_i^{2t+2} \nu_j^{2t+2} \Gamma_{ij}^{(t+1)} - \hat{\delta}_{ij} \right| \leq c d^6 \kappa^2 \log(n)^{7/2} \theta_2^{4t+4} \frac{d^6t}{\sqrt{n} \wedge \alpha^{1/4}},
\]
\[
\left| \langle S \Delta B^t \chi_i, S \Delta B^t \chi_j \rangle - \nu_i^{2t+2} \nu_j^{2t+2} \Gamma_{ij}^{(t+1)} \right| \leq c d^6 \kappa^2 \log(n)^{7/2} \theta_2^{4t+4} \frac{d^6t}{\sqrt{n} \wedge \alpha^{1/4}}.
\]

**Proof** In the following, it will be useful to define the following functional, this time acting on wedge-rooted graphs:
\[
f_{\phi, t}(g, e) = \left( \frac{mn}{d^2} \right)^t \sum_{o_2 = e_3, \ldots, o_{2t}} \left( \prod_{s=1}^t M_i(o_{2s+1}) M_i(o_{2s+2}) \right) \phi_i(o_{2t+1}),
\]
where \( e \) is a wedge and the sum runs over all tuples \( (o_i) \) such that \( e_1, e_2, o_2, \ldots, o_{2t+1} \) is a non-backtracking path of length \( 2t + 2 \). First, let
\[
f(g, o) = \sum_{e : e_1 = o} f_{\phi, t}(g, e) f_{\phi, t}(g, e).
\]
Again, \( f \) is \((2t + 2)\)-local, and
\[
|f(g, o)| \leq 2 \frac{\kappa^2}{n} \theta_2^4 |(g, o)_{2t+2}|^2.
\]
Additionally, it is easy to check that
\[
\sum_{x \in V_1} f(G, x) = \langle B^t \chi_i, B^t \chi_j \rangle,
\]
and that, letting \( W \) be the all-one matrix,
\[
f(T, x) = \partial W f_{\phi, t} f_{\phi, t},
\]
hence using Proposition 21,
\[
\mathcal{T} = \frac{d^2 11^*}{n} \left( \nu_i \nu_j \right)^{2t} \left( \sum_{s=0}^t \frac{\Phi^s}{(\nu_i \nu_j d)^{2s}} \right) (\phi_i \circ \phi_j).
\]
Taking the scalar product of the above equation with 1, and using Proposition 20, concludes the proof of (50).

For (51), the parity-time symmetry (27) implies that

\[ \langle (B^*)^t \bar{\chi}_i, (B^*)^t \bar{\chi}_j \rangle = \langle \Delta B^t \chi_i, \Delta B^t \chi_j \rangle. \]

The corresponding graph functional is now given by

\[ f(g, o) = \frac{mn}{d^2} \sum_{e:e_1=0} M_{e_1e_2} M_{e_3e_2} f_{\phi_i, t}(g, e) f_{\phi_j, t}(g, e), \]

which yields

\[ \bar{f} = \Phi(\nu_i \nu_j)^2 t \left( \sum_{s=0}^{t} \frac{\Phi^s}{(\nu_i \nu_j d^2)^s} \right) (\phi_i \circ \phi_j) = (\nu_i \nu_j)^2 t^{2+2} \left( I_{ij}^{(t+1)} - \delta_{ij} \right). \]

The rest of the proof proceeds as above. Finally, the last equation corresponds to

\[ f(g, o) = f_{\phi_i, t}(g, o) f_{\phi_j, t}(g, o), \]

and proceeds identically. \[\blacksquare\]

Now we are ready to prove Equations (32)-(33) of Theorem 10.

**Proof [Proof of (32) and (33)]** Recall the definition of \( u_i, \hat{u}_i \) in (30). For any \( r_0 \times r_0 \) matrix \( M \), we have

\[ \|M\| \leq \|M\|_F \leq r_0 \max_{ij} |M_{ij}|. \]  \[\text{(54)}\]

The first statement in (32) follows from (54) and the entrywise bounds in (50) by taking \( c = \varepsilon \) and \( t = \ell \). The second statement in (32) follows in the same way due to (51). For any \( i, j \in [r_0] \),

\[ \langle u_i, \hat{u}_j \rangle = \frac{\langle B^t \chi_i, (B^*)^t \bar{\chi}_j \rangle}{\nu_i^2 \nu_j^2 t^{2+2}} = \frac{\langle B^{2t} \chi_i, \bar{\chi}_j \rangle}{\nu_i^2 \nu_j^2 t^{2+2}}. \]

Then the first statement of (33) follows from Lemma 22 by taking \( c = \varepsilon \) and \( t = 2\ell \). Similarly,

\[ \langle \hat{u}_i, B^t u_j \rangle = \frac{\langle B^{3\ell} \chi_j, \bar{\chi}_i \rangle}{\nu_i^{2\ell+2} \nu_j^{2\ell}}. \]

The second statement of (33) is proved by taking \( t = 3\ell \) in Lemma 22. This completes the proof. \[\blacksquare\]

Finally, we show a result on a pseudo-eigenvector property of \( \chi_i \) that shall be useful for the next step of the proof.

**Lemma 24** Let \( h = c \log_d(n) \). With probability at least 1 \( - c' d^4 n^{4c-1} \), for any \( t \leq h \),

\[ \|B^{t+1} \chi_i - \nu_i^2 B^t \chi_i\|^2 \leq K^2 d^2 \theta_1^4 + c' d^{12} \kappa^2 \theta_2^4 \log(n)^{7/2} \frac{d^{6t}}{\sqrt{n} \wedge \alpha^{1/4}}. \]
Proof. Recalling the definition of \( f \) in (53), we define
\[
  f(g, o) = \sum_{e, s=1} X e_{i, t+1} (g, e) - \nu^2 f \phi_i (g, e)) \]
Then \( f \) is \((2t + 4)\)-local, and satisfies
\[
  |f(g, o)| \leq 4 \kappa^2 \theta^4 t^2 \|
\]
On the other hand, by the same arguments as above, we have
\[
  f = d^2 \Phi^t (\phi_i \circ \phi_i),
\]
and using (41)
\[
  \left| \sum_{x \in [n]} f(x) \right| \leq K^2 d^2 \theta^4 t^2.
\]
Proposition 20 with a triangular inequality completes the proof.

Appendix F. Matrix expansion and norm bounds

We now move on to show (34). Let \( E_2(V) \) be the oriented wedge set on a complete bipartite graph with vertex sets \( V_1, V_2 \). We can extend the definition of \( B \) to be an operator on \( E_2(V) \) such that for any \( e = (e_1, e_2, e_3) \), \( f = (f_1, f_2, f_3) \) \( \in E_2(V) \),
\[
  B_{e, f} = \begin{cases} A_{f_1, f_2} A_{f_3, f_2} 1 \{ e \rightarrow f \} & \text{if } e, f \in E_2, \\
  0 & \text{otherwise.} \end{cases}
\]

For \( k \geq 0 \), we can define \( \Gamma^{2k+2}_{e, f} \) to be the set of non-backtracking walks of length \((2k + 2)\) from \( e \) to \( f \) denoted by \((\gamma_0, \cdots, \gamma_{2k+2}) \). We then have
\[
  B_{e, f}^k = \sum_{\gamma \in \Gamma^{2k+2}_{e, f}} X e_1 e_2 X e_3 e_2 \prod_{s=1}^k A_{\gamma_2 s, \gamma_{2s+2}, \gamma_{2s+1}}.
\]
A bipartite graph spanned by \( \gamma \) is given by all vertices and edges from \( \gamma \). We say \( \gamma \) is a tangle-free path if the bipartite graph \( G \) spanned by \( \gamma \) contains at most one cycle. Otherwise, we call \( \gamma \) a tangled path. A bipartite graph \( G \) is called \( \ell \)-tangle-free if for any \( x \in V_1 \), there is at most one cycle in the \( \ell \)-neighborhood of \( x \) in \( G \) denoted by \( (G, x)_\ell \).

For \( k \geq 0 \), let \( F^{2k+2}_{e, f} \subseteq \Gamma^{2k+2}_{e, f} \) be the subset of all tangle-free paths of length \((2k + 2)\). If the bipartite graph \( G \) corresponding to the biadjacency matrix \( X \) is \((2\ell + 2)\)-tangle free, then for all \( 1 \leq k \leq \ell \), we must have \( B^k = B^{(k)} \), with
\[
  B^{(k)}_{e, f} = \sum_{\gamma \in F^{2k+2}_{e, f}} X e_1 e_2 X e_3 e_2 \prod_{s=1}^k A_{\gamma_{2s}, \gamma_{2s+2}, \gamma_{2s+1}}.
\]
We introduce short-hand notations:

\[ A_{e_1 e_2} = A_{e_1} A_{e_2}, \quad X_{e_1 e_2} = X_{e_1} X_{e_2}, \quad M_{e_1 e_2} = M_{e_1} M_{e_2}. \]

And similarly,

\[ Q_{e_1 e_2} = Q_{e_1} Q_{e_2} = mn M_{e_1 e_2}^2. \]  

(56)

Define the corresponding centered random variables

\[ A_{e_1 e_2} = A_{e_1} A_{e_2} - M_{e_1} M_{e_2}, \quad X_{e_1 e_2} = X_{e_1} X_{e_2} - \frac{d^2}{mn}. \]

We then define \( B^{(k)} \), a centered version of \( B^{(k)} \), as

\[ B^{(k)} = \sum_{\gamma \in F_{2k+2}^{ef}} X_{e_1 e_2} \prod_{s=1}^{k} A_{\gamma_2 \gamma_{2s+2}, \gamma_{2s+1}}. \]

We also define

\[ B^{(0)} = \{ e = f \} X_{e_1 e_2}, \quad B^{(0)} = \{ e = f \} X_{e_1 e_2}. \]

The following telescoping sum formula holds for any real numbers \( a_s, b_s, 0 \leq s \leq \ell \):

\[ \prod_{s=0}^{\ell} a_s = \prod_{s=0}^{\ell} b_s + \sum_{t=0}^{\ell} b_t (a_t - b_t) \prod_{s=t+1}^{\ell} a_s. \]

Separating the cases \( t = 0 \) in the sum, we can decompose \( B^{(\ell)} \) as

\[ B^{(\ell)} = B^{(\ell)} + \frac{d^2}{mn} \sum_{\gamma \in F_{2\ell+2}^{ef}} \prod_{s=1}^{\ell} A_{\gamma_2 \gamma_{2s+2}, \gamma_{2s+1}} \]

\[ + \sum_{t=1}^{\ell} \sum_{\gamma \in F_{2t+2}^{ef}} X_{e_1 e_2} \prod_{s=1}^{t-1} A_{\gamma_2 \gamma_{2s+2}, \gamma_{2s+1}} M_{\gamma_2 t, \gamma_2 t+2, \gamma_2 t+1} \prod_{s=t+1}^{\ell} A_{\gamma_2 s, \gamma_2 s+2, \gamma_2 s+1}. \]

For \( 1 \leq t \leq \ell - 1 \), define \( F_{2t+2}^{ef} \subset \Gamma_{2t+2}^{ef} \) the set of non-backtracking tangled paths \( \gamma = (\gamma_0, \ldots, \gamma_{2\ell+2}) \) such that \( (\gamma_0, \ldots, \gamma_{2t+2}) \in F_{2t}^{2t} g, (\gamma_{2t+2}, \ldots, \gamma_{2\ell+2}) \in F_{g'}^{2\ell+2} \) for some edges \( g, g' \in E_2(V) \). For \( t = 0 \), \( F_{2t+2}^{ef} \subset \Gamma_{2t+2}^{ef} \) is the set of non-backtracking tangled paths \( \gamma = (\gamma_0, \ldots, \gamma_{2\ell+2}) \) such that \( (\gamma_0, \gamma_1) = (e_1, e_2) \), and \( (\gamma_{2t+2}, \ldots, \gamma_{2\ell+2}) \in F_{g'}^{2t} \) for some \( g \in E_2(V) \). Necessarily, \( g_1 = e_3 \). Similarly, \( F_{2t+2}^{ef} \subset \Gamma_{2t+2}^{ef} \) is the set of non-backtracking tangled paths \( \gamma = (\gamma_0, \ldots, \gamma_{2\ell+2}) \) such that \( (\gamma_0, \ldots, \gamma_{2\ell}) \in F_{e_3}^{2t} \) for some \( g \in E_2 \), and \( (\gamma_{2t+1}, \gamma_{2t+2}) = (f_2, f_3) \). Necessarily, \( g_3 = f_1 \).
Next, we introduce three matrices $H, H^{(1)}$ and $H^{(2)}$ such that for $e, f \in E_2(V)$,

$$H_{ef} = \frac{d^2}{mn} 1\{e \to f\} \quad (57)$$

$$H_{ef}^{(1)} = 1\{e \to f\} M_{f_1 f_3 f_2} \quad (58)$$

$$H_{ef}^{(2)} = \sum_{e \to g \to f} M_{e_1 g_2} A_{f_1 f_3 f_2}, \quad (59)$$

where the sum is over all $g \in E_2$ such that $e \to g \to f$ is a non-backtracking walk of length 6. Then the following decomposition holds:

$$B^{(\ell)} = B^{(\ell)} + H B^{(\ell-1)} + \sum_{t=1}^{\ell-1} B^{(t-1)} H^{(2)} B^{(\ell-t-1)} + B^{(\ell-1)} H^{(1)} - \sum_{t=0}^{\ell} R_{t}^{(\ell)},$$

where for $1 \leq t \leq \ell$,

$$(R_{t}^{(\ell)})_{ef} = \sum_{\gamma \in F_{t+2, e f}} \prod_{s=1}^{t-1} A_{\gamma_{2s+2, \gamma_{2s+1}}} M_{\gamma_{2s+2, \gamma_{2s+1}}} \prod_{s=t+1}^{\ell} A_{\gamma_{2s+2, \gamma_{2s+1}}}.$$

From (23), the following holds:

$$(TPS_{\Delta})_{ef} = \sum_{i,j} T_{ei} P_{ij} S_{jf} \Delta_{ff} = P_{e_3 f_1} \Delta_{ff} = \sum_{u \in V_2} M_{e_3 u} M_{f_1 u} A_{f_1 f_2} A_{f_3 f_2}.$$

And from the singular value decomposition of $P$ in (28), and the matrix relations from (25),

$$TPS_{\Delta} = \sum_{i=1}^{n} \nu_i^2 T_\phi \phi_i^* S_\Delta = \sum_{i=1}^{n} \nu_i^2 \chi_i \tilde{\chi}_i.$$ \quad (60)

Note that $H^{(2)}$ is close to $TPS_{\Delta}$. We define $\tilde{H}$ such that

$$H^{(2)} = \sum_{i=1}^{r} \nu_i^2 \chi_i \tilde{\chi}_i + \tilde{H}.$$ 

Therefore, the following decomposition holds:

$$B^{(\ell)} = B^{(\ell)} + H B^{(\ell-1)} + \sum_{t=1}^{\ell-1} \sum_{k=1}^{r} \nu_k^2 B^{(t-1)} \chi_k \tilde{\chi}_k^* B^{(\ell-t-1)} + \sum_{t=1}^{\ell-1} B^{(t-1)} \tilde{H} B^{(\ell-t-1)} + B^{(\ell-1)} H^{(1)} - \sum_{t=0}^{\ell} R_{t}^{(\ell)}.$$ \quad (61)

From (58), deterministically,

$$\|H^{(1)}\| \leq \sqrt{\|H^{(1)}\|_1 \|H^{(1)}\|_\infty} \leq nm \|M\|_\infty^2 = L^2.$$

With the decomposition in (61), the following estimate thus holds:
Lemma 25  For any unit vector \( x \in \mathbb{R}^{E_2(V)} \):

\[
\|B^\ell x\| \leq \|B^\ell\| + \|HB^\ell\| + \sum_{t=1}^{\ell-1} \sum_{i} \nu_i^2 \|B^{\ell-1} \chi_i\| |\langle \tilde{\chi}_i, B^{(\ell-t-1)} x \rangle| \\
+ \sum_{t=1}^{\ell-1} \|B^{(\ell-t)} \tilde{H}B^{(\ell-t-1)}\| + L^2 \|B^{(\ell-1)}\| + \sum_{t=0}^{\ell} \|R_t^{(\ell)}\|.
\]

The bulk estimates of (34) stem from Lemma 25 and the following proposition:

Proposition 26  Let \( \chi \) be any vector among \( \chi_1, \ldots, \chi_r \in \mathbb{C}^{E_2(V)} \). Let \( \ell \leq \lfloor \log_d(n) \rfloor \). There exists a constant \( C_1, c > 0 \) such that for \( n \geq C_1 K^{12} \), with probability at least \( 1 - cn^{-1/4} \), the following norm bounds hold for all \( 0 \leq k \leq \ell \):

\[
\|B^{(k)}\| \leq K^{11} d^{5/2} \log(n)^{10} \theta^{2k},
\]

(62)

\[
\|R_k^{(\ell)}\| \leq K^{20} d^{3/2} \log(n)^{28} \frac{L^{2\ell}}{\sqrt{mn}},
\]

(63)

\[
\|HB^{(k)}\| \leq d^5 \log(n)^8 \frac{L^{2k-2}}{mn},
\]

(64)

\[
\|B^{(k-1)} \tilde{H}B^{(\ell-k-1)}\| \leq K^{11} K_2^2 d^{3} \log(n)^{12} \theta^{2k}.
\]

(65)

And for all \( 1 \leq k \leq \ell - 1 \),

\[
\|B^{(k-1)} \tilde{H}B^{(\ell-k)}\| \leq K^{11} K_2^2 d^{3} \log(n)^{12} \frac{\theta^{2(k-1)} L^{2(\ell-k)}}{mn}.
\]

(66)

Before we prove Proposition 26 in the next section, we show how it implies eq. (34).

Proof [Proof of (34)] With the parity-time invariance from (27),

\[
\langle \tilde{\chi}_i, B^t w \rangle = \langle B^t \chi_i, J \Delta w \rangle,
\]

and we have \( \|J \Delta w\| \leq \frac{L^2}{d^2} \). Since \( w \) is orthogonal to all \( B^\ell \chi_i, i \in [r_0] \),

\[
|\nu_i^{-2t} \langle \tilde{\chi}_i, B^t w \rangle| = |\nu_i^{-2t} \langle \tilde{\chi}_i, B^t w \rangle - \nu_i^{-2\ell} \langle \tilde{\chi}_i, B^\ell w \rangle| \\
= \left| \sum_{s=t}^{\ell-1} \nu_i^{-2s} \langle \tilde{\chi}_i, B^s w \rangle - \nu_i^{-2(s+1)} \langle \tilde{\chi}_i, B^{s+1} w \rangle \right| \\
\leq \frac{L^2}{d^2} \sum_{s=t}^{\ell-1} \nu_i^{-2s-2} \|B^{s+1}\chi_i - \nu_i^2 B^s \chi_i\|.
\]

With Lemma 24,

\[
\|B^{s+1}\chi_i - \nu_i^2 B^s \chi_i\| \leq K d^2 + cK d \theta^{2s} \log(n)^{7/4} \frac{d^{3s}}{n^{1/4} \alpha^{1/8}}.
\]
Therefore since \( \nu_i \geq \theta \), for \( i \in [r_0] \),
\[
|\langle \hat{\chi}_i, B^t w \rangle| \leq K d \nu_i^{2t} \sum_{s=t}^{t-1} \left( \frac{\theta}{\nu_i} \right)^{2s+2} + \frac{cKd^6 \log^{7/4}(n)}{n^{1/4} \wedge \alpha^{1/8}} \nu_i^{2t} \sum_{s=t}^{t-1} \left( \frac{\theta}{\nu_i} \right)^{2s+2} \ d^{3s}
\]
\[
\leq \theta^{2t} \left( K d \log n + \frac{cKd^6 \log^{7/4}(n))d^{3t}}{n^{1/4} \wedge \alpha^{1/8}} \right). \tag{67}
\]

On the other hand, for \( i > r_0 \), \( \nu_i \leq \theta \), from (50),
\[
|\langle \hat{\chi}_i, B^t w \rangle| \leq \frac{L^2}{d^t} \| B^t \tilde{\chi}_i \| \leq L \nu_i^{2t} \sqrt{\Gamma_{ii}^{(t)}} + cd^2 \nu_i^{2t} \log(n)^{7/4} \frac{d^{3t}}{n^{1/4} \wedge \alpha^{1/8}}.
\]

From (31) and (41),
\[
\Gamma_{ii}^{(t)} \leq \sum_{s=0}^{t} \frac{K^2 \rho^{2s}}{(\nu_i^2 d)^{2s}} \leq K^2 \log(n) \theta^t \nu_i^{-4t},
\]
which implies for \( i > r_0 \),
\[
|\langle \hat{\chi}_i, B^t w \rangle| \leq \theta^{2t} \left( LK \log^{1/2}(n)) + cd^2 \nu_i^{2t} \log(n)^{7/4} \frac{d^{3t}}{n^{1/4} \wedge \alpha^{1/8}} \right). \tag{69}
\]

Let \( w \) be any unit vector orthogonal to all \( \tilde{u}_i \), \( i \in [r_0] \) defined in (30). By the tangle-free property in Proposition 17 and Lemma 25, with probability at least \( 1 - cn^{-1/4} \), we can apply the bounds in Proposition 26 as well as the two estimates (67) and (69) to conclude
\[
\| B^t w \| \leq crK^2 d^6 K^{20} \log^{14}(n) \theta^{2t},
\]
which is the first claim of (34). The proof of the second claim in (34) follows the same argument by considering the transpose of the decomposition in (61), as in the proof of (Stephan and Zhu, 2022, Lemma 26).

\section*{Appendix G. The trace method: proof of (62)}

We now show Proposition 26. We shall only show eq. (62) in the main text; other inequality are dealt with using similar methods, and relegated to the appendix. For any integer \( s \geq 0 \),
\[
\| B^{(k)} \|^{2s} \leq \text{tr} \left( B^{(k)} ( B^{(k)} )^* \right)^s = \sum_{(e_1, \ldots, e_{2s})} \prod_{i=1}^{2s} B^{(k)}_{e_{2i-1}, e_{2i}} B^{(k)}_{e_{2i+1}, e_{2i}}
\]
\[
= \sum_{\gamma \in W_{k,s}} \prod_{i=1}^{2s} X_{\gamma_{i,0} \gamma_{i,2} \gamma_{i,1}} \prod_{t=1}^{k} A_{\gamma_{t,2t} \gamma_{t,1} \gamma_{t,2t+1} \gamma_{t,2t+1}}, \tag{70}
\]
where \( W_{k,s} \) is the set of sequences of path \( (\gamma_1, \ldots, \gamma_{2s}) \) such that each \( \gamma_i = (\gamma_{i,0}, \ldots, \gamma_{i,2k+2}) \) is a non-backtracking tangle-free path of length \( 2k + 2 \), with boundary conditions that for all \( i \in [s] \),
\[
(\gamma_{2i, 2k}, \gamma_{2i, 2k+1}, \gamma_{2i, 2k+2}) = (\gamma_{2i-1, 2k}, \gamma_{2i-1, 2k+1}, \gamma_{2i-1, 2k+2})
\]
\[
(\gamma_{2i+1, 0}, \gamma_{2i+1, 1}, \gamma_{2i+1, 2}) = (\gamma_{2i, 0}, \gamma_{2i, 1}, \gamma_{2i, 2}) \tag{71}
\]


41
with the convention that $\gamma_{2s+1} = \gamma_1$. For each $\gamma \in W_{k,s}$, we associate with an undirected bipartite graph $G_\gamma = (V'_\gamma, V''_\gamma, E_\gamma)$ of visited vertices and edges. From the boundary condition, $G_\gamma$ is connected hence

$$|E_\gamma| - |V_\gamma| + 1 \geq 0.$$ 

Let $a(\gamma) = |E_\gamma|$, $v(\gamma) = |V_\gamma|$. We drop the dependence on $\gamma$ for ease of notation.

Now we consider the expectation of (70). It also depends on the degree profile of vertices in $V'_\gamma$. There are no vertices of degree 1 in $V'_\gamma$ by our construction of $G_\gamma$. Let $t_\gamma$ be the set of distinct triplets visited by $\gamma$ and let $m_e$ be the number of times a triplet $e$ is visited. If a triplet has an overlapped edge with another triplet, we call it a bad triplet; otherwise, we call it a good triplet. In addition, we call a vertex in $V'_\gamma$ bad if it is a middle point of a certain bad triplet.

Let $k_e$ be the number of times of a triplet $e$ is visited in $\gamma$, which carries the weight $X_e$ and denote $\tilde{m}_e = m_e - k_e$. Let $W_{k,s}$ be the subset of $W_{k,s}$ such that each good triplet was visited at least twice in $\gamma$. We can now write the expectation of (70) as

$$\sum_{\gamma \in W_{k,s}} \mathbb{E} \prod_{e \in T_\gamma} X_e^{k_e} A_e^{\tilde{m}_e}$$

$$= \sum_{\gamma \in W_{k,s}} \mathbb{E} \prod_{e \in T_\gamma, \text{good}} X_e^{k_e} A_e^{\tilde{m}_e} \prod_{e \in T_\gamma, \text{bad}} X_e^{k_e} A_e^{\tilde{m}_e}$$

$$\leq \sum_{\gamma \in W_{k,s}} \prod_{e \in T_\gamma} \left( \frac{|M_e|}{p^2} \right) \cdot \left( \prod_{e \in T_\gamma, \text{good}} p^2 \right) \cdot \mathbb{E} \prod_{e \in T_\gamma, \text{bad}} X_e^{m_e}$$

$$= \sum_{\gamma \in W_{k,s}} p^{-4ks} \prod_{e \in T_\gamma} |M_e| \tilde{m}_e \cdot \left( \prod_{e \in T_\gamma, \text{good}} p^2 \right) \cdot \mathbb{E} \prod_{e \in T_\gamma, \text{bad}} X_e^{m_e},$$

where in the last identity, we use that

$$\sum_{e \in T_\gamma} \tilde{m}_e = 2sk.$$

Notice that when $p \in [0, 1/2]$, deterministically, the following inequality holds:

$$|X_{e_1} X_{e_2} - p^2| \leq (1 + 4p)|X_{e_1} - p)(X_{e_2} - p)|.$$ (73)

Let $T_\gamma(v)$ be the collection of bad triplets with the same middle point $v$. Let $d(v)$ be the number of distinct edges connected to $v$ from bad triplets. For each $e \in E_\gamma$, with abuse of notation, let $m_e$ be the number of visits of $\gamma$ to the edge $e$. Since contributions of triplets with distinct middle points are independent, with (73), we can consider each contribution from $T_\gamma(v)$ as follows

$$\left| \mathbb{E} \prod_{e \in T_\gamma(v)} X_e^{m_e} \right| \leq (1 + 4p)^{\sum_{e \in T_\gamma(v)} m_e} \prod_{e \in E_\gamma(v)} \mathbb{E}|X_e - p|^{m_e}$$

$$\leq (1 + 4p)^{\sum_{e \in T_\gamma(v)} m_e} 2^{d(v)} p^{d(v)}.$$ (74)
Notice that $G_\gamma$ is a bipartite graph, and the imbalance between $|V_1^\gamma|$ and $|V_2^\gamma|$ is bounded by the genus plus the extra imbalance from the endpoints in each $\gamma_i$, so we have

$$||V_1^\gamma| - |V_2^\gamma|| \leq a - v + 1 + 2s.$$  \hspace{1cm} (75)

Let $v_i, v_{\geq i}$ be the number of vertices with degree $i$ (at least $i$, respectively) in $V_2^\gamma$. Then since there are no vertices of degree 1 in $V_2^\gamma$,$$v_2 + v_{\geq 3} = |V_2^\gamma|, \quad \text{and} \quad 2v_2 + \sum_{i\geq 3} iv_i \leq a.$$ Therefore

$$\sum_{v\in V_2^\gamma, \text{bad}} d(v) \leq \sum_{i\geq 3} iv_i \leq 3(\sum_{i\geq 3} (i-2)v_i) \leq 3(a-2|V_2^\gamma|) \leq 6(a-v+s+1).$$ \hspace{1cm} (76)

where the last inequality is from (75). Similarly,

$$v_{\geq 3} \leq a - 2|V_2^\gamma| \leq 2(a-v+2s).$$ \hspace{1cm} (77)

Since each $\gamma_i$ is tangle-free, each bad vertex can be visited at most twice for each $\gamma_i$. Therefore each bad vertex can be visited at most $4s$ times by $\gamma$, which implies

$$\sum_{e\in T_\gamma, \text{bad}} m_e \leq 4s \cdot v_{\geq 3} \leq 8s(a-v+2s).$$ \hspace{1cm} (78)

From (74), we obtain

$$\left| E \prod_{e\in T_\gamma, \text{bad}} X_e^{m_e} \right| \leq (1 + 4p)^{8s(a-v+2s)} 26(a-v+s+1)^s \sum_{v\in V_2^\gamma} d(v).$$ \hspace{1cm} (79)

Since

$$\sum_{v\in V_2^\gamma, \text{good}} \deg(v) + \sum_{v\in V_2^\gamma} d(v) = a,$$

with (79), we can simplify (72) to be

$$\sum_{\gamma\in W_{k,s}} E \left[ \prod_{e\in T_\gamma} X_e^{k_e} A_e^{m_e} \right] \leq \sum_{\gamma\in W_{k,s}} (1 + 4p)^{8s(a-v+2s)} 26(a-v+s+1)^s p^a 4ks \prod_{e\in E_\gamma} |M_e|^{\widehat{m}_e},$$ \hspace{1cm} (80)

where for each $e \in E_\gamma$, with abuse of notation, $\widehat{m}_e$ is defined as the number of visits of $\gamma$ to the edge $e$ excluding $(\gamma_i, 0, \gamma_{i+1}), (\gamma_i, 1, \gamma_{i+1}), i \in [2s].$

Define $W_{k,s}(v, a)$ to be the set of equivalence classes of paths in $W_{k,s}$ with $v$ distinct vertices and $s$ distinct edges. The following lemma holds.

**Lemma 27** Let $v, e$ be integers such that $a - v + 1 \geq 0$. Then

$$|W_{k,s}(v, a)| \leq (4(k+1)s)^{6s(a-v+1)+2s}.$$
**Proof** [Proof of Lemma 27] We can use the estimation from (Bordenave et al., 2018, Lemma 17) and notice that each graph is spanned by $2s$ concatenations of non-backtracking walks of length $2k + 2$. 

We now bound the contribution of paths in each equivalence class.

**Lemma 28** Let $\gamma \in \mathcal{W}_{k,s}$ such that $|V_\gamma| = v$, $|E_\gamma| = a$. We have

$$\sum_{\gamma' \sim \gamma} \prod_{e \in E_{\gamma'}} |M_e| \bar{m}_e \leq K^6(a - v + 12s + 2sk)\left(\sqrt{\frac{mn}{n}}\right)^{-4sk + v} \rho^{2sk} \left(\sqrt{\frac{m}{n}}\right)^{a - v}.$$ 

**Proof** [Proof of Lemma 28] Let $\tilde{\gamma} = (\tilde{\gamma}_1, \ldots, \tilde{\gamma}_{2s})$ be the sequence of paths from $\gamma \in \mathcal{W}_{k,s}$ such that $\tilde{\gamma}_i = (\gamma_i, 2sk_i + 2)$, i.e., we remove the initial two steps (the first triplet) in each $\gamma_i$, $1 \leq i \leq 2s$. From the boundary condition, the associated graph spanned by $\tilde{\gamma}$, denoted by $G_{\tilde{\gamma}}$, is connected and

$$|E_{\tilde{\gamma}}| = a, \quad |V_{\tilde{\gamma}}| = v.$$ 

Recall $\bar{m}_e$ is the number of visits of the edge $e \in E_{\tilde{\gamma}}$, we have

$$\sum_{e \in E_{\tilde{\gamma}}} \bar{m}_e = 4sk.$$ 

Let $H$ be the set of edges $e \in E_\gamma$ such that $\bar{m}_e = 1$ and denote $h = |H|$. Note that by our construction, any edge in $H$ is either one of the two initial or ending edges in $\tilde{\gamma}_i$, $i \in [2s]$, or from a bad triplet (note that it’s impossible to have both edges in a triplet belong to $H$, then the contribution is zero by independence). Therefore with (76),

$$h \leq 4s + \sum_{v \in V_{\gamma}^2, \text{bad}} d(v) \leq 10s + 6(a - v + 1). \quad (81)$$

We also have

$$\sum_{e \in E_\gamma \setminus H} \bar{m}_e = 4sk - h,$$

which implies

$$\sum_{\gamma' \sim \gamma} \prod_{e \in E_{\gamma'}} |M_e| \bar{m}_e \leq \left(\sqrt{\frac{K\rho}{mn}}\right)^{4sk - 2a + 2h} \sum_{\gamma' \sim \gamma} \prod_{e \in E_{\gamma'} \setminus H} \frac{Q_e}{\sqrt{mn}} \cdot \left(\sqrt{\frac{mn}{n}}\right)^{a - h - 4sk} \sum_{\gamma' \sim \gamma} \prod_{e \in E_{\gamma'} \setminus H} Q_e, \quad (82)$$

where $Q_e$ is defined in (5), and we use (6) to get $|M_e| \leq \sqrt{\frac{K\rho}{mn}}$. Next, we consider an upper bound on

$$\sum_{\gamma' \sim \gamma} \prod_{e \in E_{\gamma'} \setminus H} Q_e.$$
Let $G'_\tilde{\gamma}$ be the graph spanned by all edges in $E_{\tilde{\gamma}} \setminus H$. Let $t_i, t_{\geq i}$ be the number of vertices in $V'_{\tilde{\gamma}}$ with a degree equal to $i$ and at least $i$, respectively. We then have
\begin{align}
    v' &= t_1 + t_2 + t_{\geq 3} \geq v - h, \\
    t_1 + 2t_2 + 3t_{\geq 3} &\leq \sum_{k \geq 1} kt_k = 2(a - h).
\end{align}
(83)

The first inequality is because removing any edge in $H$ can delete at most one vertex from $G_{\tilde{\gamma}}$. Since degree-1 vertices can only appear at the endpoints of each $\tilde{\gamma}_i, i \leq 2s$, which are $\gamma_{i, 2}, \gamma_{i, 2k+2}, i \in [2s]$, we must have $t_1 \leq 2s$. Together with (83), we obtain that
\begin{equation}
    t_{\geq 3} \leq 2(a - v) + t_1 \leq 2(a - v) + 2s. 
\end{equation}
(84)

We reduce the graph $G'_\tilde{\gamma}$ to a multi-graph $\hat{G}_{\tilde{\gamma}}$ by gluing vertices of degree 2 while keeping the endpoints $\gamma_{i, 2}, \gamma_{i, 2k+2}, i \in [2s]$. Let $\hat{v}$ be the number of vertices in $\hat{G}_{\tilde{\gamma}}$. Next, we define the edge set $\hat{E}_{\tilde{\gamma}}$ for $\hat{G}_{\tilde{\gamma}}$. We partition edges in $E_{\tilde{\gamma}}'$ into a sequences of edges $\hat{e}_j = (e_{j,1}, \ldots, e_{j,q_j})$ where $e_{j,t} = (x_{j,t-1}, x_{j,t}) \in E_{\tilde{\gamma}}', x_{j,0} \in \hat{V}_{\tilde{\gamma}}$ and $x_{j,t} \notin \hat{V}_{\tilde{\gamma}}$ for $1 \leq t \leq q_j - 1$, and let $(x_{j,0}, x_{j,q_j})$ be an edge in $\hat{E}_{\tilde{\gamma}}$. This implies
\begin{equation}
    \sum_{j=1}^{\hat{a}} q_j = a - h.
\end{equation}

From the definition of $\hat{G}_{\tilde{\gamma}}$, the genus of the graph is preserved,
\begin{equation}
    \hat{a} - \hat{v} = a' - v' = a - h - v' \geq a - v - h. 
\end{equation}
(85)

and since $a' = a - h, v' \geq v - h$,
\begin{equation}
    \hat{a} - \hat{v} = a' - v' \leq a - v. 
\end{equation}
(86)

Moreover, since the number of degree-2 vertices in $\hat{G}_{\tilde{\gamma}}$ is bounded by the number of distinct endpoints in $\tilde{\gamma}_i, i \in [2s]$ , we have from (84),
\begin{equation}
    \hat{v} \leq t_1 + t_{\geq 3} + 2s \leq 2(a - v) + 6s.
\end{equation}
(87)

On the other hand, removing edges in $H$ does not increase the genus, so $a' - v' \leq a - v$. And with (87), it implies
\begin{equation}
    \hat{a} = \hat{a} - \hat{v} + \hat{v} = a' - v' + \hat{v} \leq a - v + \hat{v} \leq 3(a - v) + 6s.
\end{equation}
(88)

We classify the edges in $\hat{E}_{\tilde{\gamma}}$ according to endpoints $(x_{j,0}, x_{j,q_j})$ into 3 different groups $\hat{E}_1, \hat{E}_2,$ and $\hat{E}_3$: (1) both are in $V'_{\gamma}^1$, (2) both are in $V'_{\gamma}^2$, (3) one from $V'_{\gamma}^1$ and one from $V'_{\gamma}^2$. 45
Let $y_1, \ldots, y_v$ be the elements in $\hat{V}_\gamma$, where the first $\hat{v}_1$ vertices are from $\hat{V}_\gamma^1$, and let $a_j, b_j$ be the indices such that $x_{j,a} = y_{a_j}, x_{j,b} = y_{b_j}$. Then

$$\sum_{\gamma' : \gamma' \sim \gamma} \prod_{e \in E_{\hat{V}}_{\gamma}} Q_e \leq \sum_{(y_1, \ldots, y_v) \in [n]^{e_1} \times [m]^{e_2}} \prod_{e_j \in \hat{E}_1} (Q Q')^{q_j/2} \prod_{e_j \in \hat{E}_2} (Q')^{q_j/2} \prod_{e_j \in \hat{E}_3} \frac{K^2 \rho^j}{\sqrt{mn}}$$

$$\leq \sum_{(y_1, \ldots, y_v) \in [n]^{e_1} \times [m]^{e_2}} \prod_{e_j \in \hat{E}_1} \frac{K^2 \rho^j}{m} \prod_{e_j \in \hat{E}_2} \frac{K^2 \rho^j}{m} \prod_{e_j \in \hat{E}_3} \frac{K^2 \rho^j}{\sqrt{mn}}$$

$$= K^{2\hat{a}} \rho^{a-h} n \hat{v}_1 - |\hat{E}_1| - \frac{1}{2} |\hat{E}_2| - \frac{1}{2} |\hat{E}_3|$$

$$= (K^2)^{2\hat{a}} \rho^{a-h} (\sqrt{mn})^{\hat{v}_1 - \hat{v}} \left( \left( \frac{m}{n} \right)^{\hat{v}_2} - |\hat{E}_2| - |\hat{E}_3| - 2 |\hat{E}_2| - |\hat{E}_3| - 2 |\hat{E}_2| + 2 |\hat{E}_3| \right)$$

$$\leq (K^2)^{2\hat{a}} \rho^{a-h} (\sqrt{mn})^{\hat{v}_1 - \hat{v}} \left( \left( \frac{m}{n} \right)^{\hat{v}_2} - |\hat{E}_2| - |\hat{E}_3| - 2 |\hat{E}_2| + 2 |\hat{E}_3| \right)$$

where the first inequality is from (40), the second inequality is from (35) and (39), the third inequality is from (88) and (85), and the last one is from (86). Since all vertices in $\hat{V}_\gamma^2$ have degree at least 3,

$$-|\hat{E}_3| - 2 |\hat{E}_2| + 2 |\hat{E}_3| = \sum_{v \in \hat{V}_\gamma^2} (2 - \deg(v)) \leq 0.$$

We get

$$\sum_{\gamma' : \gamma' \sim \gamma} \prod_{e \in E_{\hat{V}}_{\gamma} \setminus H} Q_e \leq (K^2)^{3(a-v) + 6s} \rho^{a-h} (\sqrt{mn})^{v+h-a} \left( \frac{m}{n} \right)^{a-v}.$$

Therefore from (82),

$$\sum_{\gamma' : \gamma' \sim \gamma} \prod_{e \in E_{\hat{V}}_{\gamma} \setminus H} |M_e|^{\tilde{m}} \leq K^{6(a-v) + 12s + 2sk + h-a} (\sqrt{mn})^{-4sk+v} \rho^{2sk} \left( \frac{m}{n} \right)^{a-v}$$

$$\leq K^{12(a-v+1) + 22s + 2sk - a} (\sqrt{mn})^{-4sk+v} \rho^{2sk} \left( \frac{m}{n} \right)^{a-v},$$

where the last inequality is from (81) and the fact that $K \geq 1$. This completes the proof of Lemma 28.

We take

$$k \leq \lfloor \log n \rfloor, \quad s = \left\lfloor \frac{\log \left( \frac{n}{8(dvK)^2K^v} \right)}{8 \log \log n} \right\rfloor$$

(90)
and let \( g = a - v + 1 \). We have the first simple constraint on \( v, a \) as

\[
0 \leq v - 1 \leq a \leq 4(k + 1)s.
\]

Moreover, for any edge in \( G_\gamma \) not included in any bad triplets must be visited at least twice to have a nonzero contribution in (70). And for any \( \gamma \) with a nonzero contribution in (70), the number of edges visited only once is bounded by

\[
6(g + s) \text{ from (76). This implies } a \leq 2ks + 3g + 5s.
\]

From (70), (80) and (89),

\[
\mathbb{E}\|B^{(k)}\|^2 s \leq \sum_{a=1}^{4s(k+1)} \sum_{v=1} a \{a \leq 2ks + 3g + 5s\} |W_{k,s}(v, a)|
\cdot (1 + 4p)^{8s(a-v+2s)} 2^{6(a-v+s+1)} p^a - 4ks
\cdot K^{12(a-v+1)+22s+2sk-a} (\sqrt{mn})^{-4sk+v} \rho^{2sk} \left( \sqrt{\frac{m}{n}} \right)^{a-v}
\leq n(4(k+1)s)^{2s+2} (1 + 4p)^{16s^2} 2^{6s} K^{22s} \partial_1^{4sk}
\cdot \sum_{g=0}^{\infty} \sum_{a=1} (\frac{K}{d})^{2sk-a} \left( \frac{2K^{12} (1 + 4p)^{8s} (4(k + 1)s)^{6s}}{n} \right)^g.
\]

We now consider two cases:

(1) When \( K \geq d \), (91) can be further bounded by

\[
n(4(k+1)s)^{2s+3} (1 + 4p)^{16s^2} 2^{6s} K^{22s} \partial_1^{4sk} \cdot \left( \frac{K}{d} \right)^{2sk} \sum_{g=0}^{\infty} \left( \frac{2K^{12} (1 + 4p)^{8s} (4(k + 1)s)^{6s}}{n} \right)^g.
\]

(2) When \( K \leq d \), (91) is bounded by

\[
n(4(k+1)s)^{2s+3} (1 + 4p)^{16s^2} 2^{6s} K^{22s} \partial_1^{4sk} (d/K)^{5s} \sum_{g=0}^{\infty} \left( \frac{2d^3 K^9 (1 + 4p)^{8s} (4(k + 1)s)^{6s}}{n} \right)^g.
\]

Therefore in both cases, (91) is bounded by

\[
n(4(k+1)s)^{2s+3} (1 + 4p)^{16s^2} 2^{6s} K^{22s} \partial_1^{4sk} \left( 1 \lor \frac{d}{K} \right)^{5s}
\cdot \sum_{g=0}^{\infty} \left( \frac{2(d \lor K)^3 K^9 (1 + 4p)^{8s} (4(k + 1)s)^{6s}}{n} \right)^g.
\]

Since \( d \leq n^{1/12} \), from our choices of \( k \) and \( s \) in (90), for \( n \geq C_1 K^{16} \), we have

\[
(1 + 4p)^{8s} \leq 2, \quad \frac{d \lor K)^3 K^9 (4(k + 1)s)^{6s}}{n} \leq \frac{1}{8}, \quad n^{\frac{1}{52}} \leq \log(n)^5.
\]

which implies

\[
\mathbb{E}\|B^{(k)}\|^2 s \leq 2n(4(k+1)s)^{2s+3} 2^{2s} K^{22s} \partial_1^{4sk} \left( 1 \lor \frac{d}{K} \right)^{5s}.
\]
Then by Markov’s inequality and the assumption $C \geq 1$, with probability at least $1 - n^{-1/2}$, the following bound holds:

$$\|E(k)\| \leq K^{11}d^{5/2} \log(n)^{19/2} \theta^{2k}.$$ 

This finishes the proof for (62).

**Appendix H. Additional trace methods**

**H.1. Proof of (63) on $R_k^{(\ell)}$**

This is similar to the proof of (62). We only emphasize the main difference. For any $0 \leq k \leq \ell$, 

$$\|R_k^{(\ell)}\|^{2s} \leq \text{tr} \left[ R_k^{(\ell)} R_k^{(\ell)*} \right]^s = \sum_{\gamma \in T_{\ell,s,k}} 2s \prod_{i=1}^{2s} X_{\gamma_i,0} \gamma_i,2 \gamma_i,1 \prod_{t=1}^{k-1} A_{\gamma_i,2t \gamma_i,2t+2,\gamma_i,2t+1} M_{\gamma_i,2k \gamma_i,2k+2,\gamma_i,2k+1} \prod_{t=k+1}^{\ell} A_{\gamma_i,2t \gamma_i,2t+2,\gamma_i,2t+1}.$$ 

(92)

where $T_{\ell,s,k}$ is the set of all $(\gamma_1, \ldots, \gamma_{2s})$ such that for all $i$,

$$\gamma_i^1 = (\gamma_i,0,\gamma_i,1,\gamma_i,2,\ldots,\gamma_i,2k) \quad \text{and} \quad \gamma_i^2 = (\gamma_i,2k+2,\ldots,\gamma_i,2\ell+1)$$

are non-backtracking tangle-free and $\gamma_i$ is non-backtracking tangled with the same boundary condition as in (71).

For any $\gamma \in T_{\ell,s,k}$, denote $G_\gamma = (V_\gamma, E_\gamma)$ as the union of the graphs $G_{\gamma_i^j}$ for $i \in [2s]$ and $j = 1, 2$. Note that the edges $(\gamma_i,2k,\gamma_i,2k+1,\gamma_{2k+2})$ are not taken into account in $G_\gamma$. For any fixed $\gamma_i^1, \gamma_i^2, 1 \leq i \leq 2s$, there are at most $m^{2s}$ many ways to choose all $\gamma_i,2k, 1 \leq i \leq 2s$. Let $v = |V_\gamma|$ and $a = |E_\gamma|$. Since $\gamma_i$ is tangled, each connected component in $G_{\gamma_i}$ contains a cycle, therefore

$$v \leq a.$$ 

(93)

Define the set $\mathcal{T}_{\ell,s,k}(v,a)$ as the set of all equivalence classes of $T_{\ell,s,k}$ with given $v$ and $a$. The following lemma from (Bordenave et al., 2018) holds.

**Lemma 29** Let $v \leq a$. Then $|\mathcal{T}_{\ell,s,k}(v,a)| \leq (8(\ell + 1)s)^{12s(a-v+1)+8s}.

Taking expectation of (92) yields

$$\mathbb{E}\|R_k^{(\ell)}\|^{2s} \leq \frac{L^{4s}}{(mn)^s} \sum_{\gamma \in \mathcal{T}_{\ell,s,k}} \mathbb{E} \prod_{i=1}^{2s} X_{\gamma_i,0} \gamma_i,2 \gamma_i,1 \prod_{t=1}^{k-1} A_{\gamma_i,2t \gamma_i,2t+2,\gamma_i,2t+1} M_{\gamma_i,2k \gamma_i,2k+2,\gamma_i,2k+1} \prod_{t=k+1}^{\ell} A_{\gamma_i,2t \gamma_i,2t+2,\gamma_i,2t+1}.$$ 

(94)

where $\mathcal{T}_{\ell,s,k}$ is the subset of $T_{\ell,s,k}$ where each $\gamma \in \mathcal{T}_{\ell,s,k}$ has nonzero contribution in the expectation. For each $\gamma \in \mathcal{T}_{\ell,s,k}$,

$$\mathbb{E} \prod_{i=1}^{2s} X_{\gamma_i,0} \gamma_i,2 \gamma_i,1 \prod_{t=1}^{k-1} A_{\gamma_i,2t \gamma_i,2t+2,\gamma_i,2t+1} M_{\gamma_i,2k \gamma_i,2k+2,\gamma_i,2k+1} \prod_{t=k+1}^{\ell} A_{\gamma_i,2t \gamma_i,2t+2,\gamma_i,2t+1} = \mathbb{E} \prod_{e \in E_\gamma} X_e^{k_e} A_e^{\tilde{m}_e} A_e^{t_e},$$

48
where \(k_e, \tilde{m}_e, t_e\) is the number of visits of \(\gamma\) to a triplet \(e\) that carry the weight \(X_e, A_e,\) and \(A_e,\) respectively and let \(m_e = k_e + \tilde{m}_e,\) let \(T_\gamma\) be the set of all triplets from \(\gamma.\) We call a triplet \(bad\) if it overlaps another triplet. Separating good and bad triplets yields

\[
\mathbb{E} \prod_{e \in T_\gamma} X_e^{k_e} A_e^{\tilde{m}_e} A_e^{t_e} = \prod_{e \in T_\gamma} \left( \frac{|M_e|}{p^s} \right)^{\tilde{m}_e + t_e} \mathbb{E} \prod_{e \in T_\gamma \text{ good}} X_e^{m_e} X_e^{t_e} \cdot \mathbb{E} \prod_{e \in T_\gamma \text{ bad}} X_e^{m_e} X_e^{t_e} \leq p^{-4(\ell-1)s} \prod_{e \in T_\gamma} |M_e|^{\tilde{m}_e + t_e} \left( \prod_{e \in T_\gamma \text{ good}} p^2 \right) \mathbb{E} \prod_{e \in T_\gamma \text{ bad}} X_e^{m_e} X_e^{t_e}. \tag{95}\]

Let \(T_\gamma(v)\) be the collection of bad triplets with the same middle point \(v.\) Let \(d(v)\) be the number of distinct edges connected to \(v\) from bad triplets. For each \(e \in E_\gamma,\) with abuse of notation, let \(m_e, t_e\) be the corresponding number of visits of \(\gamma\) to the edge \(e.\) Applying inequality (73), we find

\[
\mathbb{E} \prod_{e \in T_\gamma(v)} X_e^{m_e} X_e^{t_e} \leq (1 + 4p)^{\sum_{e \in T_\gamma(v)} m_e} \prod_{e \in E_\gamma(v)} \mathbb{E}|X_e - p|^{m_e} X_e^{t_e} \leq (1 + 4p)^{\sum_{e \in T_\gamma(v)} m_e} 2^{d(v)} p^{d(v)}. \tag{96}\]

Let \(v_i, v_{\geq 3}\) be the number of vertices with degree \(i\) (at least \(i,\) respectively) in \(V^{2}_\gamma.\) Same as (76) and (77), we have

\[
\sum_{v \in V^{2}_\gamma, \text{bad}} d(v) \leq 6(a - v + s + 1) \tag{97}
\]

\[
v_{\geq 3} \leq 2(a - v + 2s).
\]

Since each \(\gamma^i\) is tangle-free, each bad vertex can be visited at most twice for each \(\gamma^i.\) Similar to (78), we obtain

\[
\sum_{e \in T_\gamma, \text{bad}} m_e \leq 8s \cdot v_{\geq 3} \leq 16s(a - v + 2s).
\]

With the three inequalities above, (96) and (95) imply the following bound for any \(\gamma \in T_{\ell, s, k}:\)

\[
\mathbb{E} \prod_{e \in T_\gamma} X_e^{k_e} A_e^{\tilde{m}_e} A_e^{t_e} \leq (1 + 4p)^{16s(a - v + 2s)} p^{2s(a - v + s + 1)} 4(\ell-1)s \prod_{e \in E_\gamma} |M_e|^{\tilde{m}_e + t_e}.
\]

Next, we bound the contribution of each equivalence class.

**Lemma 30** Let \(\gamma \in T_{\ell, s, k}(v, a).\) We have

\[
\sum_{\gamma': \gamma' \sim \gamma \in E_\gamma} |M_e|^{\tilde{m}_e + t_e} \leq K^{12(a - v + 1) + 2s(\ell - 1) + 40s - a} p^{2s(\ell - 1)} (\sqrt{mn})^{v - 4s(\ell - 1)} \left( \sqrt{m \over n} \right)^{a - v}. \tag{98}\]
Proof [Proof of Lemma 30] This is similar to the proof of Lemma 28. We only address the differences. Let \( \tilde{\gamma} = (\tilde{\gamma}_1, \ldots, \tilde{\gamma}_{2s}) \) be the sequence of paths from \( \gamma \in \Gamma_{t,s,k} \) with the initial two steps (the first triplet) removed in each \( \gamma_i, 1 \leq i \leq 2s \). From the boundary condition, the associated graph spanned by \( \tilde{\gamma} \), denoted by \( G_{\tilde{\gamma}} \), satisfies

\[
|E_{\tilde{\gamma}}| = a, \quad |V_{\tilde{\gamma}}| = v.
\]

Similarly, we have

\[
\sum_{e \in E_{\tilde{\gamma}}} \tilde{m}_e + t_e = 4(\ell - 1)s.
\]

Let \( H \) be the set of edges \( e \in E_{\gamma} \) such that \( \tilde{m}_e + t_e = 1 \) and denote \( h = |H| \). Therefore with (97),

\[
h \leq 4s + \sum_{v \in V_{\tilde{\gamma},bad}} d(v) \leq 10s + 6(a - v + 1).
\]

We also have

\[
\sum_{e \in E_{\tilde{\gamma}} \setminus H} \tilde{m}_e + t_e = 4(\ell - 1)s - h,
\]

which implies

\[
\sum_{\gamma' : \gamma' \sim \gamma \in E_{\tilde{\gamma}}} \prod_{e \in E_{\tilde{\gamma}}} |M_e|^{\tilde{m}_e + t_e} \leq (K\rho)^{2s(\ell - 1) + h - a} (\sqrt{mn})^{a - h - 4s(\ell - 1)} \sum_{\gamma' : \gamma' \sim \gamma \in E_{\tilde{\gamma}} \setminus H} \prod_{e \in E_{\tilde{\gamma}} \setminus H} Q_e, \tag{100}
\]

Let \( G'_{\tilde{\gamma}} \) be the graph spanned by all edges in \( E_{\tilde{\gamma}} \setminus H \). Let \( t_i, t_{\geq i} \) be the number of vertices in \( V_{\tilde{\gamma}}' \) with a degree equal to \( i \) and at least \( i \), respectively. We then have

\[
v' = t_1 + t_2 + t_{\geq 3} \geq v - h, \tag{101}
\]

\[
t_1 + 2t_2 + 3t_{\geq 3} \leq \sum_{k \geq 1} kt_k = 2(a - h).
\]

Since degree-1 vertices can only appear at the endpoints of each \( \tilde{\gamma}_i^j, i \leq 2s, j = 1, 2 \), we must have

\[
t_1 \leq 5s. \tag{102}
\]

Together with (101), we obtain that

\[
t_{\geq 3} \leq 2(a - v) + 5s. \tag{103}
\]

We reduce the graph \( G'_{\tilde{\gamma}} \) to a multi-graph \( \hat{G}_{\tilde{\gamma}} \) by gluing vertices of degree 2 except for the endpoints of \( \tilde{\gamma}_i^j, i \in [2s], j = 1, 2 \). Let \( \hat{v} \) be the number of vertices in \( \hat{G}_{\tilde{\gamma}} \). We partition edges in \( E_{\tilde{\gamma}}' \) into \( \hat{a} \) sequences of edges as before. Similar to the proof of (62),

\[
\sum_{j=1}^{\hat{a}} q_j = a - h, \quad \text{and} \quad a - v - h \leq \hat{a} - \hat{v} \leq a - v.
\]
Since the number of degree-2 vertices in \( \tilde{G}_\gamma \) is bounded by the number of distinct endpoints in \( \gamma_j^i, i \in [2s], j = 1, 2 \), we have \( t_2 \leq 5s \). And from (102) and (103),

\[
\begin{align*}
\hat{v} & \leq t_1 + t_{\geq 3} + 5s \leq 2(a - v) + 15s, \\
\hat{a} & \leq 3(a - v) + 15s.
\end{align*}
\] (104)

Following the same steps in the proof of (62), we get

\[
\sum_{\gamma' : \gamma' \sim \gamma \in E_{\mathcal{G}}} \prod_{e \in E} Q_e \leq (K^2)^{3(a - v) + 15s} \rho^{a - h}(\sqrt{mn})^{v + h - a} \left( \frac{m}{n} \right)^{a - v}.
\] (105)

Therefore from (99), (100), and (105),

\[
\sum_{\gamma' : \gamma' \sim \gamma \in E_{\mathcal{G}}} |M_e| \hat{m}_e + t_e \leq K^{12(a - v + 1) + 2s(\ell - 1) + 40s - a} \rho^{2s(\ell - 1)} (\sqrt{mn})^{v - 4s(\ell - 1)} \left( \frac{m}{n} \right)^{a - v}.
\]

This completes the proof of Lemma 30.

Now we take

\[
\ell \leq \lfloor \log n \rfloor, \quad s = \left\lfloor \frac{\log \left( \frac{n}{8(d + K)^2 M} \right)}{24 \log \log n} \right\rfloor
\] (106)

and let \( g = a - v + 1 \). From (93), we have the first simple constraint on \( v, a \) as \( 0 \leq v \leq a \leq 4(\ell + 1)s \). Any edge in \( \gamma_1^i, i \in [2s] \) not included in any bad triplets must be visited at least twice to have nonzero contribution in (94). The number of edges visited only once by \( \gamma_1^i, i \in [2s] \) is bounded by \( 6(g + s) \) from (97). This implies

\[
a \leq \frac{4ks - 6(g + s)}{2} + 6(g + s) + 2s(2\ell - 2k) = 2s(2\ell - k) + 3(g + s).
\] (107)

From (94),(98), (107), and Lemma 29,

\[
\mathbb{E}\| R_k^{(\ell)} \|^2 \leq \frac{L^{4s}}{(mn)^s} \sum_{a=1}^{4s(\ell+1)} \sum_{v=1}^{a} 1\{a \leq 2s(2\ell - k) + 3(g + s)\} (8(\ell + 1)s)^{12g + 8s}
\]

\[
\cdot (1 + 4p)^{16s(g+2s)} 2^{6s(g+s)} \rho^{a - 4(\ell - 1)s}
\]

\[
\cdot K^{12g + 2s(\ell - 1) + 40s - a} \rho^{2s(\ell - 1)} (\sqrt{mn})^{v - 4s(\ell - 1)} \left( \frac{m}{n} \right)^{a - v}
\]

\[
\leq L^{4s}(mn)^{-s} n(8(\ell + 1)s)^{8s + 1}(1 + 4p)^{32s^2} 2^{6s} K^{40s} \theta^{4s(\ell - 1)} \left( 1 + \frac{d}{K} \right)^{3s}
\]

\[
\cdot \sum_{g=0}^{\infty} \left( \frac{2^6(d + K)^2 K^g (1 + 4p)^{16s} (8(\ell + 1)s)^{12s}}{n} \right)^g.
\]
For \( n \geq C_1 K^{16} \) with a constant \( C_1 \) depending only on \( C \), from (106), we have
\[
(1 + 4p)^{16s} \leq 2, \quad \frac{(d \vee K)^3 K^9 (8(\ell + 1)s)^{12s}}{n} \leq \frac{1}{8}, \quad n^{\frac{3}{2}} \leq \log(n)^{13}.
\]
which implies
\[
E\| R_k^{(\ell)} \|^2 \leq L^{4s}(mn)^{-s}n(8(\ell + 1)s)^{8s + 1} 2^{2s} K^{4s} g^{4s(\ell - 1)} \left( 1 \vee \frac{d}{K} \right)^{3s}.
\]
Recall \( K, d \geq 1 \) and \( \theta \leq L \). Then by Markov’s inequality, with probability at least \( 1 - n^{-1/2} \), the following bound holds:
\[
\| R_k^{(\ell)} \| \leq \frac{L^{2\ell}}{\sqrt{mn}} K^{20} d^{3/2} \log(n)^{28}.
\]
This finishes the proof for (63).

**H.2. Proof of (64) on \( HB^{(k)} \)**

Recall the definition of \( H \) from (57). We have
\[
\| HB^{(k-1)} \|^2 \leq \text{tr}\left[ (HB^{(k-1)} (HB^{(k-1)})^*)^s \right] \\
\leq \left( \frac{d^2}{mn} \right)^{2s} \sum_{\gamma \in W_{k,s}} \prod_{i=1}^{s} X_{\gamma_1,1,2,\gamma_2,1-1,4,\gamma_2,1-3} \prod_{t=2}^{k-1} A_{\gamma_2,1,2,t-1,2,\gamma_2,2,t-3} \\
\cdot \prod_{t=0}^{k-1} A_{\gamma_2,2,t,2,\gamma_2,2,t+2,\gamma_2,2,t+1} X_{\gamma_2,2,k,2,\gamma_2,2,k+2,\gamma_2,2,k+1},
\]
where \( W_{k,s} \) is the set of paths defined in (71). Let \( G_\gamma \) be defined as before, and \( a = |E_\gamma|, v = |V_\gamma| \).
The set of edges that carries weight from \( H \) has cardinality at most \( 2s \) by the boundary conditions. Hence we have
\[
E\| HB^{(k-1)} \|^2 \leq \left( \frac{d^2}{mn} \right)^{2s} \sum_{\gamma \in W_{k,s}} \left( \frac{d}{\sqrt{mn}} \right)^{a-2s} \left( \frac{L^2}{d^2} \right)^{s(2k-4)} \\
= \sum_{\gamma \in W_{k,s}} \left( \frac{d}{\sqrt{mn}} \right)^{a+2s} \left( \frac{L}{d} \right)^{4s(k-2)}. 
\]
Recall the bound from Lemma 27 on \( W_{k,s}(v, a) \). Since each equivalence class contains at most \( n^{v_1} m^{v_2} \) many paths, we obtain
\[
E\| HB^{(k-1)} \|^2 \leq \sum_{a=1}^{4ks} \sum_{\ell=1}^{a+1} (4(k + 1)s)^{6s(a-v+1)+2s} n^{v_1} m^{v_2} \left( \frac{d}{\sqrt{mn}} \right)^{a+2s} \left( \frac{L}{d} \right)^{4s(k-2)}. 
\]
The imbalance between \( v_1 \) and \( v_2 \) comes from the number of cycles in \( G_\gamma \) and the total number of components of the paths (since on one path, the imbalance from \( v_1 - v_2 \) increases by at most 1, and the first path of length \( k \) introduces an imbalance of 1). We have the following upper bound that
\[
v_2 - v_1 \leq g - 1,
\]
which implies
\[ n^{r_1}m^{r_2} \leq \left( \sqrt{mn} \right)^{g-1}. \] (108)

Take \( s = \left\lfloor \frac{\log n}{12 \log \log n} \right\rfloor \). With (108),
\[ \mathbb{E}\|HB^{(k-1)}\|^{2s} \leq (4(k+1)s)^{2s+1}L^{4ks-8s}d^{10s}n(\sqrt{mn})^{-2s} \sum_{g=0}^{\infty} \left( \frac{(4(k+1)s)^{6s}}{n} \right)^{g} \]
\[ \leq 2n(\sqrt{mn})^{-2s}(4(k+1)s)^{2s+1}L^{4ks-8s}d^{10s}. \]

Applying Markov’s inequality yields (64).

H.3. Proof of (65) and (66)

The proof of (65) is similar to the proof of (62), and one can adapt the proof of (Bordenave et al., 2022b, Equation (16.21)). We skip the proof.

For (66), first notice that \( H_{\epsilon f}^{(2)} \) defined in (59) is equal to \( (TPS^*)_{\epsilon f} \) defined in (60) except for \( g \in E_2 \) such that \( (e, g, f) \) is a back-tracking path of length 6. This happens only when (1) \( g_2 = e_2 \), (2) \( g_2 = f_2 \), (3) \( f_1 = e_3 \). Therefore
\[ -\tilde{H}_{\epsilon f} = \left( M_{e_2}M_{f_2} + M_{e_3}M_{f_2} + 1\{f_1 = e_3\} \sum_{u \in V_2} M_{e_3u}^2 \right) A_{f_1f_3f_2} \]
\[ := M_{\epsilon f}^{(1)} + M_{\epsilon f}^{(2)} + M_{\epsilon f}^{(3)}. \]

Then we can write
\[ \|\tilde{H}B^{(k-1)}\| \leq \sum_{i=1}^{3} \|B^{(k-1)}M^{(i)}B^{(\ell-k-1)}\| \leq \sum_{i=1}^{3} \|B^{(k-1)}\| \|M^{(i)}B^{(\ell-k-1)}\|. \]

A similar argument of the bound (64) applied to each \( i \in [3] \) finishes the proof.

Appendix I. Proof of Theorem 7

For simplicity, in this section, we assume that \( \delta^{(i)}_{\rho,\ell} \) defined in (10) for each Unfold\( _i(T) \) satisfies \( \delta^{(i)}_{\rho,\ell} > c > 0 \) for all \( i, j \), and we place ourselves under the high-probability event of Corollary 6. Note that as \( d \to \infty \), \( L/d \ll \rho^{(i)}/\sqrt{d} \) for all \( i \), and as such \( \tau^{(i)} \leq \tau/\sqrt{d} \).

By a simple application of the triangle inequality,
\[ \|T - \tilde{T}\|_F \leq \sum_{j=1}^{r} \left\| \nu_j \bigotimes_{i=1}^{k} w^{(i)}_j - \hat{\nu}_j \bigotimes_{i=1}^{k} \hat{w}^{(i)}_j \right\|_F \]
\[ \leq \sum_{j=1}^{r} \left\| \nu_j - \hat{\nu}_j \right\| + \nu_j \left\| \bigotimes_{i=1}^{k} w^{(i)}_j - \bigotimes_{i=1}^{k} \hat{w}^{(i)}_j \right\|_F. \]
The first term is easily bounded through Corollary 6:

\[ |\nu_j - \hat{\nu}_j| \leq \frac{\nu_j^2 - \chi_j^{(1)}}{2 \min(\nu_j, \hat{\nu}_j)} \leq C_1 \nu_j \left( \frac{\tau}{d} \right)^\ell, \]

and hence by the Cauchy-Schwarz inequality

\[ \sum_{j=1}^r |\nu_j - \hat{\nu}_j| \leq C_2 \left( \frac{\tau}{d} \right)^\ell \|T\|_F. \]

For the second term, we also apply the Cauchy-Schwarz inequality to find

\[ \sum_{j=1}^r \left| \nu_j - \hat{\nu}_j \right| \leq \left\| \sum_{j=1}^r \left( \bigotimes_{i=1}^k w_j^{(i)} - \bigotimes_{i=1}^k \hat{w}_j^{(i)} \right) \right\|_F \leq \left( \sum_{j=1}^r \left\| \bigotimes_{i=1}^k w_j^{(i)} - \bigotimes_{i=1}^k \hat{w}_j^{(i)} \right\|_F^2 \right)^{1/2} \]

From Corollary 6, we get

\[ \langle w_j^{(i)}, \hat{w}_j^{(i)} \rangle \geq \frac{1}{\sqrt{\gamma_j^{(i)}}} - C_3 \left( \frac{\tau}{d} \right)^\ell, \]

where

\[ \gamma_j^{(i)} = \left( 1, \left( I - \frac{Q^{(i)}Q^{(i)*}}{\nu_j d^2} \right)^{-1} \langle w_j^{(i)}, w_j^{(i)} \rangle \right) \leq 1 + K^2 \frac{(\tau/d)^2}{1 - (\tau/d)^2}, \]

where the last inequality comes from Equation (41) and \( \langle 1, w_j^{(i)} \circ w_j^{(i)} \rangle = \|w_j^{(i)}\|^2 = 1. \) Now, it is possible to check that for \( a \geq 1 \) and \( x > 0 \) small enough,

\[ \frac{1}{\sqrt{1 + ax}} \geq 1 - ax \]

and applying this to the above bound yields

\[ \langle w_j^{(i)}, \hat{w}_j^{(i)} \rangle \geq 1 - \left( \frac{K\tau}{d} \right)^2 - C_3 \left( \frac{\tau}{d} \right)^\ell. \]

Finally, for \( a_1, \ldots, a_k \geq 0 \) such that \( \sum a_i \leq 1 \), we have \( \prod (1 - a_i) \geq 1 - \sum a_i \), and hence

\[ \sum_{j=1}^r \left( 2 - 2 \prod_{i=1}^k \langle w_j^{(i)}, \hat{w}_j^{(i)} \rangle \right) \leq kr \left( \left( \frac{K\tau}{d} \right)^2 + C_3 \left( \frac{\tau}{d} \right)^\ell \right). \]

Theorem 7 ensues from the bound \( \sqrt{a + b} \leq \sqrt{a} + \frac{b}{2\sqrt{a}}. \)

As in Bordenave et al. (2022b), we could improve the bound of Theorem 7 by considering better estimators of the \( w_i \), as well as an optimal shrinkage of the \( \hat{\nu}_i \). However, this only gains constant factors over the “naive” ones, hence we chose to keep a simpler exposition.
Appendix J. Proof of Theorem 8

We first treat the case where $x \in \{\pm 1\}^n$. By the natural homeomorphism between $(\{\pm 1\}, \times)$ and $(\mathbb{F}_2, \oplus)$, each equation of the form $\hat{x}_{i_1} \ldots \hat{x}_{i_k} = \hat{T}_{i_1 \ldots i_k}$ can be mapped to a linear equation in $\mathbb{F}_2$. This is known as the $k$-XORSAT problem, and the optimal sample complexity can be found in (Creignou and Daudé, 2003):

**Theorem 31 (Proposition 4.1 from (Creignou and Daudé, 2003))** Let $S$ be a random $k$-XORSAT problem, where each clause of the form $y_{i_1} \oplus \cdots \oplus y_{i_k} = 0$ is selected with probability $p$. Then, for $k \geq 3$, there exists a constant $c_k$ such that if $p > c_k n^{1-k}$, then with high probability, the only solution to $S$ is $x = 0$.

By considering $y_i = \hat{x}_i x_i$, when $p \geq c_k n^{1-k}$, the only solution to $\hat{x}_{i_1} \ldots \hat{x}_{i_k} = \hat{T}_{i_1 \ldots i_k}$ for all $i_1, \ldots, i_k$ is $\hat{x} = x$. Note that this system can be solved through Gaussian elimination in $\mathbb{F}_2$, which is a polynomial algorithm. Hence, the following holds:

**Proposition 32** If $p \geq c_k n^{1-k}$, and $T = x \otimes^k$ with $x \in \{\pm 1\}^n$, then there exists a polynomial-time algorithm that outputs an estimator $\hat{x}$ such that $\hat{x} = x$ with high probability.

Now, if $T = x \otimes^k$ for an arbitrary $x \in \mathbb{R}^n$, then $\text{sign}(T) = \text{sign}(x) \otimes^k$. Hence, applying Proposition 32 to $\text{sign}(T)$, we obtain an estimator $\hat{x} = \text{sign}(x)$. Therefore, we get

$$\langle \hat{x}, x \rangle = \langle \text{sign}(x), x \rangle = \|x\|_1.$$

Theorem 8 then ensues from the simple inequality

$$1 = \|x\|_2^2 = \sum x_i^2 \leq \|x\|_\infty \sum |x_i| = \|x\|_\infty \|x\|_1.$$