Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2020, Vol. 56, No. 4, 2971–2995
https://doi.org/10.1214/20-AIHP1065

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES

Spectral gap of sparse bistochastic matrices with exchangeable rows

Charles Bordenave[a], Yanqi Qiu[b] and Yiwei Zhang[c]

[a]*Institut de Mathématiques de Marseille, CNRS and Aix-Marseille University, 39 Rue Frédéric Joliot Curie, 13013 Marseille, France. E-mail: charles.bordenave@univ-amu.fr*

[b]*Institute of Mathematics and Hua Loo-Keng Key Laboratory of Mathematics, AMSS, Chinese Academy of Sciences, Beijing 100190, China and CNRS, Institut de Mathématiques de Toulouse and University of Toulouse III, Toulouse, France. E-mail: yanqi.qiu@amss.ac.cn*

[c]*School of Mathematics and Statistics, Center for Mathematical Sciences, Hubei Key Laboratory of Engineering Modeling and Scientific Computing, Huazhong University of Sciences and Technology, Wuhan 430074, China. E-mail: yiweizhang@hust.edu.cn*

Received 24 March 2019; revised 3 April 2020; accepted 25 April 2020

Abstract. We consider a random bistochastic matrix of size n of the form MQ where M is a uniformly distributed permutation matrix and Q is a given bistochastic matrix. Under sparsity and regularity assumptions on Q, we prove that the second largest eigenvalue of MQ is essentially bounded by the normalized Hilbert–Schmidt norm of Q when n grows large. We apply this result to random walks on random regular digraphs.

Résumé. Considérons une matrice bi-stochastique aléatoire de taille n et de la forme MQ avec M une matrice de permutation uniformément distribuée et Q une matrice bi-stochastique fixée. Sous des conditions de parcimonie et de régularité sur Q, on démontre que la deuxième plus grande valeur propre de MQ est essentiellement bornée par la norme de Hilbert–Schmidt normalisée de Q lorsque n est très grand. Ce résultat s'applique aux marches au hasard sur les graphes aléatoires dirigés réguliers.

MSC2020 subject classifications: 60B20; 60C05; 05C80

Keywords: Spectral gap; Random bistochastic matrices; High trace method; Tangled-free paths

1. Introduction

1.1. *Model and main result*

For $n \geq 1$ integer, let $[n] = \{1, \dots, n\}$. Let $Q \in M_n(\mathbb{C})$ be a bistochastic matrix of size n, that is, for any x, y in $[n]$, $Q_{xy} \geq 0$ and the constant vector $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^n$ is an eigenvector of Q and its transpose $Q^\intercal$:

$$Q\mathbf{1} = Q^\intercal \mathbf{1} = \mathbf{1}. \tag{1}$$

In probabilistic terms, Q is the transition matrix of a Markov chain on $[n]$ which admits the uniform measure as an invariant measure.

Let $\mathbb{S}_n$ be the symmetric group on n elements. We will denote by $|\cdot|$ the cardinal number of a set and the usual absolute value, $\mathbb{P}(\cdot)$ and $\mathbb{E}(\cdot)$ are the probability and expectation under the uniform measure on $\mathbb{S}_n$: for any subset $E \subset \mathbb{S}_n$,

$$\mathbb{P}(E) = \frac{|E|}{|\mathbb{S}_n|}.$$

CB is supported by French ANR grant ANR-16-CE40-0024-01. YQ is supported by National Natural Science Foundation of China grants NSFC Y7116335K1 and NSFC 11688101. YZ is supported by National Science Foundation of China grant NSFC 11701200, NSFC 11871262. YZ would also like to thank Department of mathematics at Southern University of Science and Technology, where part of the work was done, for their hospitality and support.

Let σ be a uniformly distributed random permutation in $\mathbb{S}_n$. We denote by M the $n \times n$ permutation matrix of σ. In matrix notation, for all $x, y \in [n]$,

$$M_{xy} = \mathbb{1}\big(\sigma(x) = y\big).$$

In this paper, we study the $n \times n$ random matrix

$$P = MQ, \tag{2}$$

or, in matrix notation, for all $x, y \in [n]$, $P_{xy} = Q_{\sigma(x)y}$. Then, P is the transition matrix of a Markov chain on $[n]$ where at each step, we compose with σ before performing a step according to Q. Note that P itself is bistochastic and thus the constant vector $\mathbf{1}$ is an eigenvector of P and its transpose $P^\intercal$ with eigenvalue 1. From Perron–Frobenius theorem, it follows that 1 is the largest eigenvalue of P. We order non-increasingly the moduli of the eigenvalues of P, $\lambda_i = \lambda_i(P)$,

$$1 = \lambda_1 \geq |\lambda_2| \geq \cdots \geq |\lambda_n|. \tag{3}$$

The spectral gap is defined as $1 - |\lambda_2|$. It measures the asymptotic mixing rate to equilibrium. For example, if P is aperiodic and irreducible, then for any probability measure π_0 on $[n]$,

$$\lim_{t\to\infty} \big\|\pi_0 P^t - \pi\big\|_{\mathrm{TV}}^{1/t} = |\lambda_2|.$$

where $\pi = \mathbf{1}/n$ is the invariant measure of P and, for a signed measure ν on $[n]$, $\|\nu\|_{\mathrm{TV}} = \frac{1}{2}\sum_x |\nu(x)|$ denotes the total variation norm (we refer to [12]).

Our main result is a sharp probabilistic upper bound on $|\lambda_2|$ which involves strikingly very few parameters of Q. For $A \in M_n(\mathbb{C})$, the normalized Hilbert–Schmidt norm is defined as

$$\|A\|_{\mathrm{HS}} = \sqrt{\frac{1}{n}\operatorname{tr}\big(AA^*\big)} = \sqrt{\frac{1}{n}\sum_{x,y}|A_{xy}|^2} = \sqrt{\frac{1}{n}\sum_{i=1}^n s_i(A)^2}, \tag{4}$$

where the scalars $s_i(A)$, denote the singular values of A (that is, the eigenvalues of $\sqrt{AA^\intercal}$ and $\sqrt{A^\intercal A}$).

The ℓ^1 to ℓ^∞ norm of $A \in M_n(\mathbb{C})$ is

$$\|A\|_{1\to\infty} = \max_{x,y}|A_{xy}|.$$

For some applications, we introduce a relaxation of this norm. It is defined, for $0 < \delta \leq 1$, as

$$\|A\|_{1\to\infty}^{(\delta)} = \inf_{\mathcal{E}\subset[n],|\mathcal{E}|<n^{1-\delta}} \max_{x\notin\mathcal{E},y}|A_{yx}|, \tag{5}$$

(note that this is not a norm for $\delta \neq 1$ and $\|A\|_{1\to\infty}^{(1)} = \|A\|_{1\to\infty}$). We also introduce a usual sparsity parameter of $A \in M_n(\mathbb{C})$, defined as

$$\|A\|_{1\to 0} = \max_x \big|\{y : A_{xy} \neq 0\}\big|, \tag{6}$$

(this is the ℓ^1 to ℓ^0 pseudo-norm for the pseudo-norm ℓ^0 on $\mathbb{C}^n$, $\|u\|_{\ell^0} = \sum_x \mathbb{1}(u_x \neq 0)$).

For the remainder of the text, we fix some $0 < \delta < 1$ and set the following notation

$$d := \big\|Q^\intercal Q\big\|_{1\to 0} \quad \text{and} \quad \rho := \|Q\|_{\mathrm{HS}} \vee \|Q\|_{1\to\infty}^{(\delta)}.$$

We will always assume that $d \geq 2$ (otherwise $d = 1$, Q itself is a permutation matrix and P and M have the same distribution). We observe that d and ρ are intrinsic parameters of P since $\|Q\|_{\mathrm{HS}} = \|P\|_{\mathrm{HS}}$, $\|Q\|_{1\to\infty}^{(\delta)} = \|P\|_{1\to\infty}^{(\delta)}$, $\|Q^\intercal Q\|_{1\to 0} = \|P^\intercal P\|_{1\to 0}$. Note also that the singular values of P and Q are equal. Our main result asserts that $|\lambda_2|$ is essentially bounded by ρ as long as d is not too large.

Theorem 1. *Let $n \geq 1$ be an integer and let σ be a uniformly distributed random permutation in $\mathbb{S}_n$. Let M be the permutation matrix of σ and $Q \in M_n(\mathbb{R})$ be a bistochastic matrix as above. Let $P = MQ$. The eigenvalues of P are denoted as in* (3). *Then for any $0 < c_0 < \delta \leq 1$, there exists a constant $c_1 > 0$ (depending only on δ, c_0) such that*

$$\mathbb{P}\big(|\lambda_2| \geq (1+\varepsilon)\rho\big) \leq n^{-c_0},$$

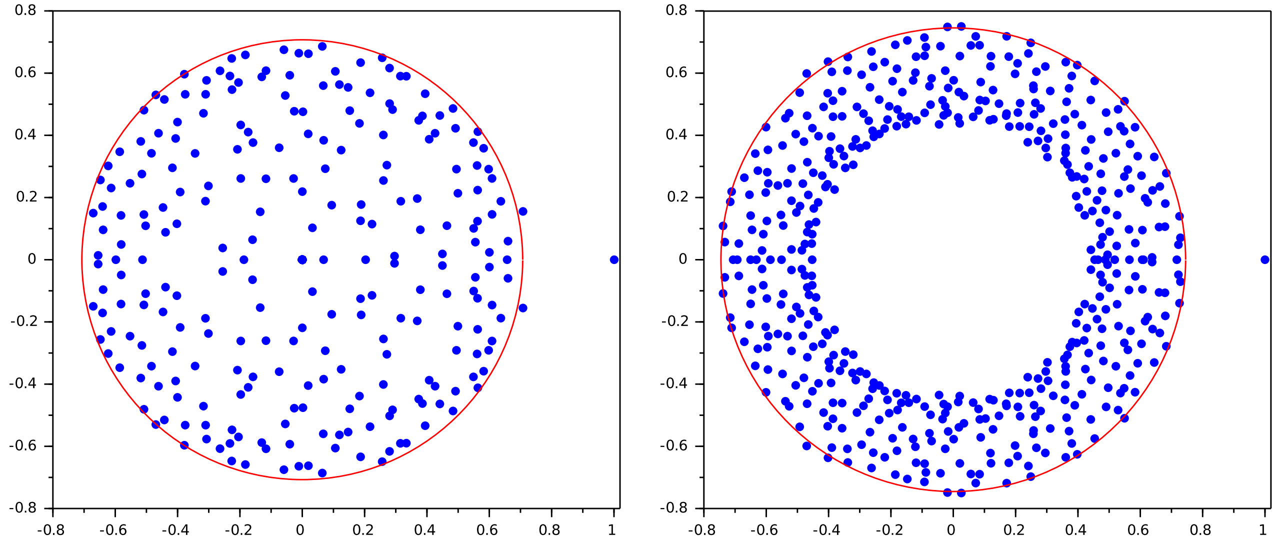

Fig. 1. Plot of the eigenvalues of P for a single realization of M when $n = 500$ and $Q = pI_n + (1-p)I_{n/2} \otimes D$ where I_n is the identity matrix of size n, D is the matrix of size 2 given by $D_{11} = D_{22} = 0$, $D_{21} = D_{12} = 1$ with $p = 1/2$ (left) and $p = 1/3$ (right). The circles in red have radii $\|Q\|_{\mathrm{HS}} = \sqrt{p^2 + (1-p)^2}$.

where

$$\varepsilon = c_1 \frac{\log d}{\sqrt{\log n}}.$$

See Figure 1 for numerical simulations. Theorem 1 implies that in many cases, the second largest eigenvalue of P is much smaller than the second largest eigenvalue of Q. Assume for example that Q is symmetric (in probabilistic term, Q is a reversible Markov chain) and that $\rho = \|Q\|_{\mathrm{HS}}$ (that is $\|Q\|^{(\delta)}_{1\to\infty} \le \|Q\|_{\mathrm{HS}}$). Then the eigenvalues of Q are real and their absolute values coincide with the singular values of Q. From (4), $\|Q\|_{\mathrm{HS}}$ is the ℓ^2-average of the eigenvalues of Q, the latter is typically much smaller than the second largest eigenvalue of Q in absolute value. Note also that the eigenvalues of M are all of modulus 1 and that, with probability tending to 1 as n goes to infinity, M is non irreducible. It follows that, even if the Markov chains Q and M have a small spectral gap (Q may even be non irreducible), the composed Markov chain $P = MQ$ has typically a large spectral gap.

The conclusion of Theorem 1 is especially interesting when $\rho = \|Q\|_{\mathrm{HS}}$. This is a condition on the inhomogeneity of the matrix Q. Indeed, observe that

$$\max_y Q_{yx} \le \sqrt{\sum_y Q_{yx}^2}.$$

Assume that the right-hand side of the above inequality does not depend on x. Then we find that $\|Q\|_{1\to\infty} \le \|Q\|_{\mathrm{HS}}$ and $\rho = \|Q\|_{\mathrm{HS}}$. The latter condition holds for example if Q is a transition matrix of simple random walk on the simple regular graph.

We remark that the order n^{-c_0} in Theorem 1 cannot be improved significantly when Q admits an invariant subspace of small dimension spanned by vectors of the canonical basis $(e_x)_{x\in[n]}$. More precisely, assume for example that $H = \mathrm{span}(e_1, \dots, e_k)$ is the invariant subspace of Q for some fixed integer $1 \le k \le n/2$. Consider the event $\sigma([k]) = [k]$. It is not hard to check that this event has probability $1/\binom{n}{k} \ge 1/n^k$. On this event, H and its orthogonal $H^\perp$ are both invariant by Q. Hence, on this event, $\lambda_1 = \lambda_2 = 1$ and

$$\mathbb{P}\big(|\lambda_2| = 1\big) \ge n^{-k}.$$

Similarly, if $\delta = 0$ (that is, $\rho = \|Q\|_{\mathrm{HS}}$), the conclusion of the theorem may be wrong. Assume for example that Q is a bistochastic matrix such that the subset

$$S = \big\{x \in [n] : Q_{y_x x} = 1 \text{ for some } y_x \in [n]\big\}$$

is of positive proportion in $[n]$. Then the probability that for at least one of such $x \in S$, we have $\sigma(x) = y_x$ is uniformly lower bounded in n. On the latter event, $\lambda_2 = 1$ since $P_{xx} = 1$.

We expect that when $\rho = \|Q\|_{\mathrm{HS}}$ and $d = \exp(o(\sqrt{\log n}))$, the conclusion of Theorem 1 is sharp. Namely, we conjecture that for any $\varepsilon > 0$, $|\lambda_2| \geq (1-\varepsilon)\rho$ with probability tending to 1 as n goes to infinity. In the next subsection, we will discuss some examples where the conjecture is true. There is an indirect evidence supporting this conjecture when we replace the random permutation matrices by other random unitary matrices. Let U be a random unitary matrix of size n sampled according to the Haar measure on the unitary group. Under mild assumptions on Q, it is known that the spectral radius of $UQ/\|Q\|_{\mathrm{HS}}$ converge in probability to 1, see [9,10,16] and, for the connection to free probability [11,15]. More generally, from these references, we might also guess an asymptotic formula for the empirical distribution of the eigenvalues of $P/\|Q\|_{\mathrm{HS}}$.

Theorem 1 is related to the recent work by Coste [6]. There, the author studies the spectral gap of the transition matrix of simple random walk on a random digraph. With our notation, it corresponds to the second eigenvalue of a Markovian matrix of size m, proportional to n, of the form $ASB^{\intercal}$, where S is uniformly distributed in $\mathbb{S}_n$ and A, B are specific matrices in $M_{m,n}(\mathbb{C})$ such that $A\mathbf{1}_n = \mathbf{1}_m$ and $B^{\intercal}\mathbf{1}_m = \mathbf{1}_n$. In some cases treated in [6], the upper bound on $|\lambda_2|$ is also given by $(1+o(1))\|B^{\intercal}A\|_{\mathrm{HS}}$. Our two results are thus of the same nature even if they are not directly comparable.

We remark finally that Theorem 1 can be extended to some extend beyond the uniform measure on $\mathbb{S}_n$, see Remark 2 below, and beyond bistochastic matrices, see Remark 3 (for examples to matrices Q such that $\mathbf{1}$ is a common eigenvector of Q and $Q^{\intercal}$).

1.2. *Random walks on random digraphs*

In this section, we state some immediate consequences of Theorem 1.

A *digraph* $G = (V, E)$ is the pair formed by a countable vertex set V and a set of oriented edges $E \subset V \times V$. If $e = (u, v) \in E$ then e is an incoming edge of v and an outgoing edge of u. For $r \in \mathbb{N}$, we say that G is r-regular if any vertex has exactly r incoming and r outgoing edges. If the set E is symmetric then G can be interpreted as an undirected graph.

Theorem 2. *Let $n \geq 1$ and $r \geq 2$ be integers and Q be the transition matrix of a simple random walk on a r-regular digraph $G = (V, E)$ with $V = [n]$. Let σ be a uniformly distributed permutation in $\mathbb{S}_n$ and let M be its permutation matrix. Let $P = MQ$ be as in (2) with eigenvalues denoted as in (3). For any $0 < c_0 < 1$, there exists $c_1 > 0$ (depending only on c_0) such that the conclusion of Theorem 1 holds with $\rho = 1/\sqrt{r}$ and $d = r^2$.*

In the above theorem, the matrix P is the transition matrix of the simple random walk on the random digraph $G^\sigma = (V, E^\sigma)$ where $E^\sigma = \{(\sigma^{-1}(x), x') : (x, x') \in E\}$. Note that G^σ will have many weak cycles of length 4 if G has many weak cycles of length 4 where a *weak cycle* of length $k \geq 1$ is a sequence $(v_0, \ldots, v_k)$ in V such that $v_0 = v_k$ and for each $t \in [k]$, either $(v_{t-1}, v_t) \in E$ or $(v_t, v_{t-1}) \in E$ (in words: it is a cycle in the undirected graph associated to G).

Theorem 2 can be applied to uniformly sampled r-regular digraphs.

Corollary 1. *Let $n \geq 1$ and $r \geq 2$ be integers. Let P be sampled uniformly over bistochastic matrices of size $n \times n$ with entries in $\{0, 1/r\}$ and with eigenvalues as in (3). Then for any $0 < c_0 < 1$, there exists $c_1 > 0$ (depending only on c_0) such that the conclusion of Theorem 1 holds with $\rho = 1/\sqrt{r}$ and $d = r^2$.*

For $r \geq 2$ uniformly bounded in n, Corollary 1 is contained in [6, Corollary 1.2]. There is a converse of Corollary 1 in some range of the degree r. It is a consequence of the main results in [5,13] that, if $r \leq n - (\log n)^{96}$ and $r \to \infty$, then, for any $\varepsilon > 0$, with probability tending to 1 as n goes to ∞, $|\lambda_2| \geq (1-\varepsilon)\rho$. Hence, if $r \to \infty$ and $r = \exp(o(\sqrt{\log n}))$, $|\lambda_2|/\rho$ converges in probability to 1 as $n \to \infty$.

Let us give another application of Theorem 1. From Birkhoff-von Neumann Theorem, the set of bistochastic matrices is the convex hull of permutation matrices. We thus have the decomposition

$$Q = \sum_{i=1}^{r} p_i M_i, \tag{7}$$

where M_i are permutations matrices and $(p_1, \ldots, p_r)$ is a probability vector. This decomposition is not unique in general. Our next result asserts that if Q admits such decomposition with r not too large and matrices M_i which have few common non-zeros entries then the second largest eigenvalues of P is at most $(1+o(1))\sqrt{\sum_i p_i^2}$.

Theorem 3. *Let $n \geq 1$ and $r \geq 2$ be integers, $p = (p_1, \dots, p_r)$ be a probability vector and $\sigma_1, \dots, \sigma_r$ be permutations in $\mathbb{S}_n$ with associated permutation matrices $M_1, \dots, M_r$. Assume that Q is given by (7). We set $S = \{x \in [n] : \exists i \neq j, \sigma_i(x) = \sigma_j(x)\}$. Let σ be a uniformly distributed permutation in $\mathbb{S}_n$ and let M be its permutation matrix. Let $P = MQ$ be as in (2) with eigenvalues denoted as in (3). For any $0 < c_0 < \delta \leq 1$, there exists a constant $c_1 > 0$ (depending only on δ, c_0) such that if $|S| \leq n^{1-\delta}$, then the conclusion of Theorem 1 holds with $\rho = \sqrt{\sum_i p_i^2}$ and $d = r^2$.*

In Theorem 3, assume that $S = \varnothing$. Then $G = (V, E)$ and $G^\sigma = (V, E^\sigma)$ with $V = [n]$, $E = \{(x, \sigma_i(x)) : x \in V, i \in [r]\}$ and $E^\sigma = \{(\sigma^{-1}(x), \sigma_i(x)) : x \in V, i \in [r]\}$ are r-regular digraphs. The transition matrices Q and P correspond to anisotropic random walks on G and G^σ. Interestingly, the scalar $\sqrt{\sum_i p_i^2}$ is the spectral radius of the anisotropic random walk on the infinite homogeneous directed tree, see the monograph [7].

Corollary 2. *Let $n \geq 1$ and $r \geq 2$ be integers, $p = (p_1, \dots, p_r)$ be a probability vector and $\sigma_1, \dots, \sigma_r$ be independent and uniformly distributed permutations in $\mathbb{S}_n$ with associated permutation matrices $M_1, \dots, M_r$. Set*

$$P = \sum_{i=1}^{r} p_i M_i$$

with eigenvalue denoted as in (3). For any $0 < c_0 < 1$, there exists a constant $c_1 > 0$ (depending only on c_0) such that the conclusion of Theorem 1 holds with $\rho = \sqrt{\sum_i p_i^2}$ and $d = r^2$.

Consider the setting of Corollary 2 in the case $p_i = 1/r$ for all $i \in [r]$. Then $\rho = 1/\sqrt{r}$. It follows from the main result in [1] that if, for some $c > 0$, $(\log n)^{12} \leq r \leq cn$ then for any $\varepsilon > 0$, $|\lambda_2| \geq (1 - \varepsilon)\rho$ with probability tending to 1 as n goes to infinity. Hence, in the regime $(\log n)^{12} \leq r \leq \exp(o(\sqrt{\log n}))$, $|\lambda_2|/\rho$ converges in probability to 1 as n goes to infinity.

1.3. *Strategy of proof of Theorem 1*

The proof of Theorem 1 will follow the strategy developed in [2,3,14] to study the spectral gap of non-backtracking operators of random graphs. This approach is quite general and it is designed to compute the top eigenvalues of a matrix defined on a sparse random graph such that the successive powers of the matrix count some weighted paths with few cycles.

Let us summarize the strategy of proof and its caveats. We will fix an integer ℓ of order $\log n$. Since (1) also holds for P, it is immediate to check that

$$|\lambda_2|^\ell \leq \left\| \left(P^\ell\right)_{|\mathbf{1}^\perp} \right\| := \max_{\langle v, \mathbf{1}\rangle = 0} \frac{\|P^\ell v\|_2}{\|v\|_2}. \tag{8}$$

Our main result is an upper bound for the operator norm of P^ℓ on $\mathbf{1}^\perp$. By adjusting the constants c_0, c_1, Theorem 1 is an immediate consequence of (8) and the following result applied to $\ell \sim (c_0/3)\log n/\log d$.

Theorem 4. *For any $0 < c_0 < \delta \leq 1$, there exists a constant $c_1 > 0$ such that, for any integer $\ell \geq 1$,*

$$\mathbb{P}\left(\left\| \left(P^\ell\right)_{|\mathbf{1}^\intercal} \right\| \geq e^{c_1\sqrt{\log n}} \rho^\ell\right) \leq d^{\ell + 50\sqrt{\log n}} n^{-c_0}.$$

To prove Theorem 4, it would seem natural to introduce the matrix $\underline{P} = \underline{M} Q$ where

$$\underline{M} = M - \frac{1}{n} \cdot \mathbf{1} \otimes \mathbf{1} = M - \mathbb{E}M, \tag{9}$$

and

$$\mathbf{1} \otimes \mathbf{1} = \mathbf{1}\mathbf{1}^\intercal.$$

Indeed, from (1),

$$\left\| \left(P^\ell\right)_{|\mathbf{1}^\intercal} \right\| = \left\| (\underline{P})^\ell \right\|.$$

A usual route would then be estimating the operator norm $\|(\underline{P})^\ell\|$ thanks to the high trace method. That is, we use for any real random matrix B and integer $m \geq 1$,

$$\mathbb{E}\|B\|^{2m} = \mathbb{E}\|BB^\intercal\|^m \leq \mathbb{E}\,\mathrm{tr}\big[(BB^\intercal)^m\big]. \tag{10}$$

Our problem requires to use the above inequality with $B = (\underline{P})^\ell$ and $\ell m \gg \log n$. However, as explained above, due to the potential presence of low dimensional invariant subspaces in P, the event $\lambda_2 = 1$ has probability at least n^{-c} and hence $\mathbb{E}\|(\underline{P})^\ell\|^{2m} \geq n^{-c}$, which may be much larger than $\rho(1+\varepsilon)^{2\ell m}$ for ε small enough, in the regime $\ell m \gg \log n$.

To circumvent this difficulty, we have to remove beforehand some events. We will then use the crucial fact that with high probability the random matrix M is *free of ℓ-tangles* with the matrix Q, where a tangle is a path of length ℓ which contains at least two cyles in a graph associated to the non-zero entries of $P = MQ$ and Q or meet the subset $\mathcal{E} \subset [n]$ (see Definition 2 below for a precise definition). For an intuition, recall that the matrix $P^h = (QM)^h$ is the h-steps transition of the Markov chain. Now every multiplication by M sends a point $x \in [n]$ to a random point $\sigma(x)$, and since Q has few non zero entries, as long as h is not too large, for most starting points, it is unlikely that there exists trajectories which come back at their starting point after h steps.

On this event, we will have the matrix identity

$$P^\ell = P^{(\ell)},$$

where $P^{(\ell)}$ is a matrix where the contribution of all tangles will vanish at once (see (13) below). Thanks to basic linear algebra, we will then project the matrix $P^{(\ell)}$ on the orthogonal of the vector $\mathbf{1}$ and give a deterministic upper bound of $\|(P^\ell)_{|\mathbf{1}^\intercal}\|$ in terms of the operator norms of new matrices which will be expressed as weighted paths of length at most ℓ.

In the remainder of the proof, we will use the high trace method to upper bound the operator norms of these new matrices: if A is such matrix, we will use (10) for some integer m of order $\sqrt{\log n}$. By construction, the expression on the right-hand side of (10) is then an expected contribution of some weighted paths of lengths $2m\ell$ of order $\ell\sqrt{\log n}$.

The study of the expected contribution of weighted paths in (10) will have a probabilistic and a combinatorial part. The necessary probabilistic computations on the random permutation are gathered in Section 3. In Section 4, we will use these computations together with combinatorial upper bounds on directed paths to deduce sharp enough bounds on our operator norms. The success of this step will essentially rely on the fact that the contributions of tangles vanish in $P^{(\ell)}$. Finally, in Section 5, we gather all ingredients to conclude.

In the remainder of the paper, we let $\mathcal{E}$ be a fixed subset of $[n]$ of cardinality at most $n^{1-\delta}$ which achieves the minimum in (5) for $A = Q$.

2. Path decomposition

In this section, we fix $\sigma \in \mathbb{S}_n$ with permutation matrix M and a positive integer ℓ. Our aim is to derive a deterministic upper bound on the norm of $(P^\ell)_{\mathbf{1}^\perp}$ defined in (8) (in forthcoming Lemma 1) when M and Q satisfy a property which will be called ℓ-tangle-free. This can be studied by an expansion of paths in the graph. To this end, we introduce some definition.

Definition 1. A *path of length k* is a sequence $\gamma = (x_1, y_1, x_2, \dots, x_k, y_k, x_{k+1})$, with $x_t, y_t \in [n]$ and $Q_{y_t x_{t+1}} > 0$. The set of paths of length k is denoted by Γ^k. If $x, y \in [n]$, we denote by Γ^k_{xy} paths in Γ^k such that $x_1 = x$, $x_{k+1} = y$.

A *subpath* of γ is a path of the form $(x_s, y_s, \dots, y_t, x_{t+1})$ with $1 \leq s \leq t \leq k$, or, if $x_i = x_j$ for some $1 \leq i < j \leq n$, a path of the form $(x_s, y_s, \dots, x_i, y_j, \dots, x_{t+1})$ with $1 \leq s \leq i < j \leq t \leq k$.

We will use the convention that a product over an empty set is equal to 1 and the sum over an empty set is 0. By construction, for integer $k \geq 0$, from (2) we find that

$$\big(P^k\big)_{xy} = \sum_{\gamma \in \Gamma^k_{xy}} \prod_{t=1}^{k} M_{x_t y_t} Q_{y_t x_{t+1}}, \tag{11}$$

where the sum is over all paths of length k from x to y. Note that, in the above expression for P^k, only the summand depends on the permutation σ. Observe that $\underline{M}$ defined in (9) is the orthogonal projection of M on $\mathbf{1}^\perp$. The matrix $(\underline{P})^k = (\underline{M}Q)^k$ can similarly be written as

$$\big((\underline{P})^k\big)_{xy} = \sum_{\gamma \in \Gamma^k_{xy}} \prod_{t=1}^{k} \underline{M}_{x_t y_t} Q_{y_t x_{t+1}}.$$

As pointed in Introduction, the matrix $(\underline{P})^k$ is orthogonal projection of P^k on $\mathbf{1}^\perp$ but it is not suited for our probabilistic analysis.

We will now introduce the central definition of *tangled paths*. Recall that $\mathcal{E} \subset [n]$ is a fixed set of cardinality at most $n^{1-\delta}$ which achieves the minimum in (5).

Definition 2. Fix the integer $h := \lceil 20\sqrt{\log n}\rceil$.

- A *coincidence* is a path $(x_1, y_1, \dots, x_t, y_t, x_{t+1})$ with $(x_1, \dots, x_t)$ **pairwise distinct** such that $((Q^\intercal Q)^h)_{x_1 x_{t+1}} > 0$.
- An $\mathcal{E}$*-coincidence* is a path $(x_1, y_1, \dots, x_t, y_t, x_{t+1})$ with $(x_1, \dots, x_t)$ **pairwise distinct** such that $x_1 = x_{t+1}$ is in $\mathcal{E}$.
- A path γ is tangle-free if it contains (as subpaths) at most one coincidence, no $\mathcal{E}$-coincidence. It is tangled otherwise. The subsets of tangle-free paths in Γ^k and Γ^k_{xy} will be denoted by F^k and F^k_{xy} respectively.
- The pair (M, Q) is ℓ-tangle-free if for any $k \in [\ell]$ and $\gamma = (x_1, y_1, x_2, \dots, x_k, y_k, x_{k+1}) \in \Gamma^k \backslash F^k$, we have

$$\prod_{t=1}^{k} M_{x_t y_t} = 0.$$

Importantly, note that the definition of paths, coincidences and tangles do not depend on σ, they depend only on the non-zero entries of Q. For example, the set Γ^k does not depend on the permutation matrix M. Observe also that the condition $((Q^\intercal Q)^h)_{xx'} > 0$ is equivalent to the existence of an integer $0 \le k \le h$ and sequences $(x_0, \dots, x_k)$, $(y_1, \dots, y_k)$ such that $x_0 = x$, $x_k = x'$, $(x_0, \dots, x_k)$ pairwise distinct and for any $s \in [k]$, $\min(Q_{y_s x_{s-1}}, Q_{y_s, x_s}) > 0$.

Remark 1. Note that by our definition, a path following multiple times the same cycle may not be tangled. For example, assume that $x_1, \dots, x_t$ are points in $[n]\backslash\mathcal{E}$ such that there does not exist an integer $0 \le s < h$ and $i \neq j$ with $Q^s_{x_i, x_j} > 0$. Then the following path

$$\gamma = (x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4, x_2, y_2, x_3, y_3, x_4, y_4, x_2, y_2, x_3, y_3, x_4, y_4, x_5)$$

is tangle-free. Note however that if one of the x_j's in $\mathcal{E}$ then the path is tangled.

If the pair (M, Q) is ℓ-tangle-free then by definition, for any $k \in [\ell]$ and for any γ in $\Gamma^k_{xy}\backslash F^k_{xy}$, the summand on the right-hand side of (11) is zero. Therefore,

$$P^k = P^{(k)}, \tag{12}$$

where $P^{(k)}$ is defined by the following formula

$$\big(P^{(k)}\big)_{xy} := \sum_{\gamma \in F^k_{xy}} \prod_{t=1}^{k} M_{x_t y_t} Q_{y_t x_{t+1}}. \tag{13}$$

For $k \in [\ell]$, we define similarly the matrix $\underline{P}^{(k)}$ by

$$\big(\underline{P}^{(k)}\big)_{xy} = \sum_{\gamma \in F^k_{xy}} \prod_{t=1}^{k} (\underline{M})_{x_t y_t} Q_{y_t x_{t+1}}. \tag{14}$$

Note that it is not necessarily true that even if the pair (M, Q) is ℓ-tangle-free that $(\underline{P})^\ell = \underline{P}^{(\ell)}$. Nevertheless, we may still express $P^{(\ell)}v$ in terms of $\underline{P}^{(\ell)}v$ for all $v \in \mathbf{1}^\perp$ at the cost of adding an explicit error term. We start with the following telescopic sum decomposition:

$$\big(P^{(\ell)}\big)_{xy} = \big(\underline{P}^{(\ell)}\big)_{xy} + \sum_{\gamma \in F^\ell_{xy}} \sum_{k=1}^{\ell} \prod_{t=1}^{k-1} (\underline{M}_{x_t y_t}) Q_{y_t x_{t+1}} \cdot \frac{Q_{y_k x_{k+1}}}{n} \cdot \prod_{t=k+1}^{\ell} M_{x_t y_t} Q_{y_t x_{t+1}}, \tag{15}$$

which is a consequence of the identity,

$$\prod_{t=1}^{\ell} a_t = \prod_{t=1}^{\ell} b_t + \sum_{k=1}^{\ell} \prod_{t=1}^{k-1} b_t \cdot (a_k - b_k) \cdot \prod_{t=k+1}^{\ell} a_t.$$

We now rewrite (15) as a sum of matrix products for lower powers of $\underline{P}^{(k)}$ and $P^{(k)}$ up to some remainder terms. For $k \in [\ell]$, let $T^{\ell,k}$ denote the set of paths $\gamma = (x_1, y_1, \dots, y_\ell, x_{\ell+1})$ such that (i) $\gamma' = (x_1, y_1, \dots, y_{k-1}, x_k) \in F^{k-1}$, (ii) $\gamma'' = (x_{k+1}, y_{k+1}, \dots, y_\ell, x_{\ell+1}) \in F^{\ell-k}$, (iii) γ is tangled. We have the following picture:

$$\gamma = (\gamma', y_k, \gamma'') = (\underbrace{x_1, y_1, \dots, y_{k-1}, x_k}_{\gamma' \in F^{k-1}}, y_k, \underbrace{x_{k+1}, y_{k+1}, \dots, y_\ell, x_{\ell+1}}_{\gamma'' \in F^{\ell-k}}).$$

Then, if $T^{\ell,k}_{xy}$ is the subset of $\gamma \in T^{\ell,k}$ such that $x_1 = x$ and $x_{\ell+1} = y$, we set

$$\big(R_k^{(\ell)}\big)_{xy} = \sum_{\gamma \in T^{\ell,k}_{xy}} \prod_{t=1}^{k-1} (\underline{M}_{x_t y_t}) Q_{y_t x_{t+1}} \cdot Q_{y_k x_{k+1}} \cdot \prod_{t=k+1}^{\ell} M_{x_t y_t} Q_{y_t x_{t+1}}. \tag{16}$$

Let us rewrite (15) as

$$\big(P^{(\ell)}\big)_{xy} = \big(\underline{P}^{(\ell)}\big)_{xy} + \frac{1}{n} \sum_{k=1}^{\ell} \underbrace{\sum_{\gamma \in F^\ell_{xy}} \prod_{t=1}^{k-1} (\underline{M}_{x_t y_t}) Q_{y_t x_{t+1}} \cdot Q_{y_k x_{k+1}} \cdot \prod_{t=k+1}^{\ell} M_{x_t y_t} Q_{y_t x_{t+1}}}_{\text{denoted by } S(k,x,y)}.$$

For fixed $k \in [\ell]$, let us rewrite the summand $S(k, x, y)$. Using the following equality,

$$F^\ell_{xy} \bigsqcup T^{\ell,k}_{xy} = \bigsqcup_{x_k \in [n]} \bigsqcup_{x_{k+1} \in [n]} \bigsqcup_{y_k \in [n] : Q_{y_k x_{k+1}} > 0} \big\{(\gamma', y_k, \gamma'') | \gamma' \in F^{k-1}_{x x_k}, \gamma'' \in F^{\ell-k}_{x_{k+1} y}\big\},$$

and using the definition (16) for $(R_k^{(\ell)})_{xy}$, we obtain that

$$\begin{aligned} S(k,x,y) &= \sum_{x_k \in [n]} \sum_{x_{k+1} \in [n]} \sum_{y_k \in [n]} \sum_{\gamma' \in F^{k-1}_{x x_k}} \sum_{\gamma'' \in F^{\ell-k}_{x_{k+1} y}} \prod_{t=1}^{k-1} (\underline{M}_{x_t y_t}) Q_{y_t x_{t+1}} \cdot Q_{y_k x_{k+1}} \cdot \prod_{t=k+1}^{\ell} M_{x_t y_t} Q_{y_t x_{t+1}} - \big(R_k^{(\ell)}\big)_{xy} \\ &= \sum_{x_k \in [n]} \sum_{x_{k+1} \in [n]} \sum_{y_k \in [n]} \big(\underline{P}^{(k-1)}\big)_{x x_k} \cdot Q_{y_k x_{k+1}} \cdot \big(P^{(\ell-k)}\big)_{x_{k+1} y} - \big(R_k^{(\ell)}\big)_{xy} \\ &= \sum_{x_k \in [n]} \sum_{x_{k+1} \in [n]} \big(\underline{P}^{(k-1)}\big)_{x x_k} \cdot (\mathbf{1} \otimes \mathbf{1} \cdot Q)_{x_k x_{k+1}} \cdot \big(P^{(\ell-k)}\big)_{x_{k+1} y} - \big(R_k^{(\ell)}\big)_{xy} \\ &= \big(\underline{P}^{(k-1)} (\mathbf{1} \otimes \mathbf{1}) P^{(\ell-k)}\big)_{xy} - \big(R_k^{(\ell)}\big)_{xy}, \end{aligned}$$

where at the last line we have used that Q is bistochastic: $\mathbf{1} \otimes \mathbf{1} \cdot Q = \mathbf{1} \otimes (\mathbf{1}^\intercal Q) = \mathbf{1} \otimes \mathbf{1}$. Therefore,

$$P^{(\ell)} = \underline{P}^{(\ell)} + \frac{1}{n} \sum_{k=1}^{\ell} \underline{P}^{(k-1)} (\mathbf{1} \otimes \mathbf{1}) P^{(\ell-k)} - \frac{1}{n} \sum_{k=1}^{\ell} R_k^{(\ell)},$$

where we have set $P^{(0)} = \underline{P}^{(0)} = I$. Observe that if (M, Q) is ℓ-tangle-free, then (12) and P being bistochastic imply

$$\mathbf{1}^\intercal P^{(\ell-k)} = \mathbf{1}^\intercal P^{\ell-k} = \mathbf{1}^\intercal.$$

Hence, if (M, Q) is ℓ-tangle-free and $\langle v, \mathbf{1} \rangle = 0$, $\|v\|_2 = 1$, we find

$$\big\| P^\ell v \big\|_2 \le \big\| \underline{P}^{(\ell)} \big\| + \frac{1}{n} \sum_{k=1}^{\ell} \big\| R_k^{(\ell)} \big\|.$$

We mention here that the method used for the proof of the above inequality first appeared in [14] and was further developed in [2,3,6].

We arrive at the following lemma.

Lemma 1. *Let $\ell \geq 1$ be an integer and $\sigma \in \mathbb{S}_n$ with permutation matrix M be such that the pair (M, Q) is ℓ-tangle-free. Then,*

$$\left\| \left(P^\ell\right)_{|\mathbf{1}^\intercal} \right\| \leq \left\| \underline{P}^{(\ell)} \right\| + \frac{1}{n} \sum_{k=1}^{\ell} \left\| R_k^{(\ell)} \right\|.$$

3. Computations on random permutation

In this section, we check that if σ is uniformly distributed on $\mathbb{S}_n$ then, with high probability the pair (M, Q) is ℓ-tangle-free provided that ℓ is not too large. We will then state a proposition on the expected product of entries of the permutation matrix $\underline{M}$. Recall that $h = \lceil 20\sqrt{\log n} \rceil$ was defined in Definition 2.

Lemma 2. *There exists $c > 0$ such that for any integer $\ell \geq 1$, the pair (M, Q) is ℓ-tangle-free with probability at least $1 - c\ell d^{\ell+2h} n^{-\delta}$.*

Proof. We may assume without loss of generality that $\ell \leq n/2$ (otherwise the content of the lemma is empty). Let us say that a path $\gamma = (x_1, y_1, \dots, y_k, x_{k+1})$ *occurs* if for any $t \in [k]$, $M_{x_t y_t} = 1$ (that is $\sigma(x_t) = y_t$). If the pair (M, Q) is ℓ-tangled then at least one of the three following paths occurs for some integers with $1 \leq k + k' \leq \ell$, $1 \leq i \leq k + k'$ and $1 \leq k \leq j \leq k + k' + 1$:

($I_{k,k',i}$) There exists a path $(x_1, y_1, \dots, x_{k+k'+1})$, where all x_t's are pairwise distinct except possibly $x_1 = x_{k+1}$ and $x_i = x_{k+k'+1}$ such that $(x_1, y_1, \dots, x_{k+1})$ and $(x_i, y_i, \dots, x_{k+k'+1})$ are distinct coincidences.
($I'_{k,k',j}$) There exists a path $(x_1, y_1, \dots, x_{k+k'+1})$ where all x_t's are pairwise distinct except possibly $x_1 = x_{k+k'+1}$ such that $(x_k, y_k, \dots, x_j)$ is a coincidence and $(x_1, y_1, \dots x_{k+k'+1})$ is a coincidence.
(II_k) There exists a path $(x_1, y_1, \dots, y_k, x_{k+1})$ which is an $\mathcal{E}$-coincidence.

The configuration $I_{k,k',i}$ describes the situation when γ has two consecutive coincidences, $I'_{k,k',j}$ accounts for the possibility that one coincidence is contained in another. II_k describes the possibility of a closed cycle containing an element in $\mathcal{E}$.

Let us bound the probability of the two different configurations. Recall that if $\{a_1, \dots, a_t\}$ and $\{b_1, \dots, b_t\}$ are two subsets of cardinal t then

$$\mathbb{P}\big(\sigma(a_1) = b_1, \dots, \sigma(a_t) = b_t\big) = \frac{1}{(n)_t}. \tag{17}$$

where $(n)_t = n(n-1)\cdots(n-t+1)$.

Let us start with $I_{k,k',i}$. Then, there are $(n)_{k+k'-1}$ choices for (x_j), $j \notin \{k+1, k+k'+1\}$, at most d^h choices for x_{k+1} and $x_{k+k'+1}$ and $\|Q^\intercal\|_{1\to 0}^{k+k'} \leq d^{k+k'}$ choices for the y_t's (since $Q_{y_t x_{t+1}} > 0$ by the definition of a path). We apply (17) with $t = k + k'$ and $a_s = x_s$, $b_s = y_s$, we arrive at

$$\mathbb{P}(I_{k,k',i}) \leq \frac{d^{k+k'} d^{2h} (n)_{k+k'-1}}{(n)_{k+k'}} \leq 2\frac{d^{k+k'+2h}}{n},$$

(where the last inequality uses $\ell \leq n/2$).

The same argument gives

$$\mathbb{P}\big(I'_{k,k',j}\big) \leq \frac{d^{k+k'} d^{2h} (n)_{k+k'-1}}{(n)_{k+k'}} \leq 2\frac{d^{k+k'+2h}}{n}.$$

Similarly, for II_k there are at most $|\mathcal{E}|$ choices for x_1, $(n)_{k-1}$ choices for (x_j), $j \notin \{1\}$ and d^k choices for the y_t's. From (17), we get

$$\mathbb{P}(II_k) \leq \frac{d^{k+h} |\mathcal{E}| (n)_{k-1}}{(n)_k} \leq 2\frac{d^k |\mathcal{E}|}{n} \leq 2\frac{d^k}{n^\delta},$$

(where we have used the assumption that $|\mathcal{E}| \leq n^{1-\delta}$). $\square$

Let $\mathbf{x} = (x_1, \dots, x_k), \mathbf{y} = (y_1, \dots, y_k) \in [n]^k$. We are interested in estimating for $0 \le k_0 \le k$,

$$\mathbb{E} \prod_{t=1}^{k_0} \underline{M}_{x_t y_t} \prod_{t=k_0+1}^{k} M_{x_t y_t}.$$

To this end, the *arcs* of $(\mathbf{x}, \mathbf{y})$ is defined as

$$A_{\mathbf{xy}} = \{(x_t, y_t) : t \in [k]\}.$$

The cardinal of $A_{\mathbf{xy}}$ is at most k. The *multiplicity* of $e \in A_{\mathbf{xy}}$ is $m_e = \sum_{t=1}^{k} \mathbb{1}((x_t, y_t) = e)$. An arc $e = (x, y)$ is *consistent*, if $\{t : (x_t, y_t) = (x, y)\} = \{t : x_t = x\} = \{t : y_t = y\}$. It is inconsistent otherwise. The following proposition is proved in [2, Proposition 27].

Proposition 1. *There exists a constant $c > 0$ such that for any $\mathbf{x} = (x_1, \dots, x_k), \mathbf{y} = (y_1, \dots, y_k) \in [n]^k$ with $2k \le \sqrt{n}$ and any $k_0 \le k$, we have*

$$\left| \mathbb{E} \prod_{t=1}^{k_0} \underline{M}_{x_t y_t} \prod_{t=k_0+1}^{k} M_{x_t y_t} \right| \le c 2^b \left(\frac{1}{n}\right)^a \left(\frac{3k}{\sqrt{n}}\right)^{a_1},$$

where $a = |A_{\mathbf{xy}}|$, b is the number of inconsistent arcs of $(\mathbf{x}, \mathbf{y})$ and a_1 is the number of $1 \le t \le k_0$ such that (x_t, y_t) is consistent and has multiplicity 1 in $A_{\mathbf{xy}}$.

4. Expected high trace method

In this section, we use the expected high trace method to derive upper bounds on the operator norms of $\underline{P}^{(\ell)}$ and $R_k^{(\ell)}$ defined respectively by (14) and (16). Before starting the formal proof, let us give an overview of the arguments.

The expected high trace method was introduced by Füredi and Komlós [8] in random matrix theory. It starts from the two elementary observations: (i) for any matrix $A \in M_n(\mathbb{R})$, $\|A\| \le \operatorname{tr}((AA^\intercal)^k)^{1/(2k)}$ and these two terms are equivalent if $k \gg \log(n)$ and (ii) if A is a random matrix, we may use to a great benefit the linearity of the trace and the expectation to estimate this last expression:

$$\mathbb{E} \operatorname{tr}\{(AA^\intercal)^k\} = \sum_{x_1, \dots, x_{2m}} \mathbb{E} \prod_{i=1}^{m} (A_{x_{2i-1} x_{2i}} A_{x_{2i+1} x_{2i}}), \tag{18}$$

where $x_{2m+1} = x_1$ and the sum runs over all x_i's in $[n]$. The core of the argument is thus to capture which sequences $(x_1, \dots, x_{2m})$ will dominate the sum on the right-hand side of (18). Drawing a parallel with statistical physics, there is the usual tension between high energy terms (sequences $(x_1, \dots, x_{2m})$ such that $\mathbb{E} \prod_{i=1}^{m} A_{x_{2i-1} x_{2i}} A_{x_{2i+1} x_{2i}}$ is large) and the entropic contribution (contributions of typical sequences $(x_1, \dots, x_{2m})$). In our context for $A = \underline{P}^{(\ell)}$, the right-hand side of (18) will depend on both the randomness of the permutation M and the structure of the matrix Q. In the definition of $\rho = \|Q\|_{\mathrm{HS}} \vee \|Q\|_{1\to\infty}^{(\delta)}$, high energy terms are responsible of the factor $\|Q\|_{1\to\infty}^{(\delta)}$ while the entropic contribution is captured by the factor $\|Q\|_{\mathrm{HS}}$.

The usual strategy to estimate the right-hand side of (18) is to interpret the sequence $(x_1, \dots, x_{2m})$ as a path on the complete graph on $[n]$ and try to bound $\mathbb{E} \prod_{i=1}^{m} A_{x_{2i-1} x_{2i}} A_{x_{2i+1} x_{2i}}$ in terms of simple graph invariants such as the number of vertices. The goal is then to count how many paths give rise to a graph with these invariants. The arguments are thus quite technical and with a strong combinatorial flavor. In the seminal paper of Füredi and Komlós [8] it was sufficient to keep track of the number of vertices. Our proof follows roughly the same strategy than [2–4,6,14] where similar random matrices are studied and where it was possible to keep track of the number of vertices and the number of edges thanks to an analog of our tangle-free assumption. In the present paper however, the argument is substantially more involved because we have to take into account some features of the deterministic matrix Q. Due to this unusual situation, the graph, denoted below by K_γ, that we will associate to a path $(x_1, \dots, x_{2m})$ will depend on the matrix Q and there will be three graph invariants: the number of vertices, the number of edges and a new parameter p which will somewhat quantify how much the path interferes with the matrix Q.

4.1. *Operator norm of* $\underline{P}^{(\ell)}$

In this paragraph, we prove the following proposition.

Proposition 2. *Assume* $d \le \exp(\sqrt{\log n})$. *For any* $c_0 > 0$, *there exists* $c_1 > 0$ (*depending on* c_0, δ) *such that for any integer* $1 \le \ell \le \log n$, *with probability at least* $1 - n^{-c_0}$,

$$\|\underline{P}^{(\ell)}\| \le e^{c_1\sqrt{\log n}}\rho^{\ell}.$$

Recall the number h defined in Definition 2. Let m be a positive integer so that

$$6m < h. \tag{19}$$

With the convention that $x_{2m+1} = x_1$, we find from (14),

$$\begin{aligned}
\|\underline{P}^{(\ell)}\|^{2m} &= \|\underline{P}^{(\ell)}\underline{P}^{(\ell)\mathsf{T}}\|^{m} \le \operatorname{tr}\{(\underline{P}^{(\ell)}\underline{P}^{(\ell)\mathsf{T}})^{m}\} \\
&= \sum_{x_1,\dots,x_{2m}}\prod_{i=1}^{m}(\underline{P}^{(\ell)})_{x_{2i-1},x_{2i}}(\underline{P}^{(\ell)})_{x_{2i+1},x_{2i}} \\
&= \sum_{x_1,\dots,x_{2m}}\prod_{i=1}^{m}\Bigg[\sum_{\gamma_{2i-1}\in F^{\ell}_{x_{2i-1},x_{2i}}}\prod_{t=1}^{\ell}(\underline{M})_{x_{2i-1,t}y_{2i-1,t}}Q_{y_{2i-1,t}x_{2i-1,t+1}}\Bigg] \\
&\quad\times\Bigg[\sum_{\gamma_{2i}\in F^{\ell}_{x_{2i+1},x_{2i}}}\prod_{t=1}^{\ell}(\underline{M})_{x_{2i,t}y_{2i,t}}Q_{y_{2i,t}x_{2i,t+1}}\Bigg] \\
&= \sum_{x_1,\dots,x_{2m}}\sum_{\substack{\gamma_1,\dots,\gamma_{2m}\\ \gamma_{2i-1}\in F^{\ell}_{x_{2i-1},x_{2i}},\\ \gamma_{2i}\in F^{\ell}_{x_{2i+1},x_{2i}}}}\prod_{i=1}^{m}\prod_{t=1}^{\ell}(\underline{M})_{x_{2i-1,t}y_{2i-1,t}}Q_{y_{2i-1,t}x_{2i-1,t+1}}\prod_{t=1}^{\ell}(\underline{M})_{x_{2i,t}y_{2i,t}}Q_{y_{2i,t}x_{2i,t+1}},
\end{aligned}$$

where we used the notation $\gamma_i = (x_{i,1}, y_{i,1}, \dots, y_{i,\ell}, x_{i,\ell+1}) \in F^{\ell}$.

Now, we define $W_{\ell,m}$ as the set of $\gamma = (\gamma_1, \dots, \gamma_{2m})$ such that $\gamma_i = (x_{i,1}, y_{i,1}, \dots, y_{i,\ell}, x_{i,\ell+1}) \in F^{\ell}$ and for all $i \in [m]$,

$$x_{2i,1} = x_{2i+1,1} \quad \text{and} \quad x_{2i-1,\ell+1} = x_{2i,\ell+1}, \tag{20}$$

with the convention that $x_{2m+1,1} = x_{1,1}$. Using this notation, we obtain

$$\|\underline{P}^{(\ell)}\|^{2m} \le \sum_{\gamma\in W_{\ell,m}}\prod_{i=1}^{2m}\prod_{t=1}^{\ell}(\underline{M})_{x_{i,t}y_{i,t}}Q_{y_{i,t}x_{i,t+1}}. \tag{21}$$

Our goal is to estimate the expectation of the above expression thanks to Proposition 1 and a counting argument which will rely crucially on the fact that an element $\gamma \in W_{\ell,m}$ is composed of $2m$ tangle-free paths, $(\gamma_1, \dots, \gamma_{2m})$.

We will count the elements in $W_{\ell,m}$ in terms of a measure of the size of their support. For $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{2m}) \in W_{\ell,m}$, we define $X_\gamma = \{x_{i,t} : i \in [2m], t \in [\ell]\}$ and $Y_\gamma = \{y_{i,t} : i \in [2m], t \in [\ell]\}$. We then consider the graph K_γ with vertex set X_γ and, for any x, x' in K_γ, $\{x, x'\}$ is an edge of K_γ if and only if

$$(Q^{\mathsf{T}}Q)_{xx'} > 0.$$

(That is, there exists $y \in [n]$ such that $\min(Q_{yx}, Q_{yx'}) > 0$). The graph K_γ induces an equivalence relation on X_γ, where each equivalence class is a connected component of K_γ. We set

$$\mathrm{cc}(x) := \text{ the equivalence class of } x.$$

(Note that cc depends implicitly on X_γ). By definition, for any $x' \in \mathrm{cc}(x)$ with $x' \neq x$, there exists a sequence $(x_0, x_1, \dots, x_k)$ of distinct points in X_γ such that

$$x_0 = x, x_k = x' \quad \text{and} \quad \left(Q^\top Q\right)_{x_{t-1}x_t} > 0 \quad \text{for any } t \in [k].$$

The *arcs* of $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{2m}) \in W_{\ell,m}$, denoted by A_γ, is the set of distinct pairs $(x_{i,t}, y_{i,t})$. We define $W_{\ell,m}(s,a,p)$ as the set of $\gamma \in W_{\ell,m}$ with $s = |X_\gamma|$, $a = |A_\gamma|$ and $s - p$ connected components in K_γ. Then taking the expectation in (21), we may write

$$\mathbb{E}\left\|\underline{P}^{(\ell)}\right\|^{2m} \leq \mathbb{E} \sum_{\gamma \in W_{\ell,m}} \prod_{i=1}^{2m} \prod_{t=1}^{\ell} \underline{M}_{x_{i,t} y_{i,t}} Q_{y_{i,t} x_{i,t+1}} = \sum_{s,a,p} \sum_{\gamma \in W_{\ell,m}(s,a,p)} \mu(\gamma) q(\gamma),$$

where for $\gamma \in W_{\ell,m}$, we have defined

$$\mu(\gamma) := \mathbb{E} \prod_{i=1}^{2m} \prod_{t=1}^{\ell} \underline{M}_{x_{i,t}, y_{i,t}} \quad \text{and} \quad q(\gamma) = \prod_{i=1}^{2m} \prod_{t=1}^{\ell} Q_{y_{i,t} x_{i,t+1}}. \tag{22}$$

To estimate the above sum, we decompose further $W_{\ell,m}(s,a,p)$ into equivalence classes as follows. For $\gamma, \gamma' \in W_{\ell,m}(s,a,p)$, let us say $\gamma \sim \gamma'$ if there exist a pair of permutations α and β in S_n such that the image of K_γ by α is $K_{\gamma'}$ and for any (i,t), $x'_{i,t} = \alpha(x_{i,t})$, $y'_{i,t} = \beta(y_{i,t})$ (where $\gamma' = (\gamma'_1, \gamma'_2, \dots, \gamma'_{2m})$ with $\gamma'_i = (x'_{i,1}, y'_{i,1}, \dots, y'_{i,\ell}, x'_{i,\ell+1})$). We define $\mathcal{W}_{\ell,m}(s,a,p)$ as the set of equivalence classes. An element in $\mathcal{W}_{\ell,m}(s,a,p)$ is unlabeled in the language of combinatorics.

We notice that $\mu(\gamma) = \mu(\gamma')$ if $\gamma \sim \gamma'$ and we obtain the bound,

$$\mathbb{E}\left\|\underline{P}^{(\ell)}\right\|^{2m} \leq \sum_{s,a,p} \left|\mathcal{W}(s,a,p)\right| \max_{\gamma \in W(s,a,p)} \left(\left|\mu(\gamma)\right| \sum_{\substack{\gamma' \in W_{\ell,m}(s,a,p): \\ \gamma' \sim \gamma}} q\left(\gamma'\right) \right). \tag{23}$$

To estimate the right-hand side of the above inequality, we will establish three results respectively. In Lemma 3, for a given triple (s,a,p), we give a combinatorial bound on the number $|\mathcal{W}_{\ell,m}(s,a,p)|$. In Lemma 5, we produce a bound for the sum of $q(\gamma)$ over all γ in a single equivalence class. In Lemma 6, we estimate the probabilistic weight $\mu(\gamma)$. Importantly all these bound are expressed in terms of the common genus $g := a + p - s + 1$ of elements in $W_{\ell,m}(s,a,p)$. To conclude the proof of Proposition 2, it will remain to combine the three lemmas in expression (23).

Lemma 3. *If $g := a + p - s + 1 < 0$ or $2g + 2m > p$, then $W_{\ell,m}(s,a,p)$ is empty. Otherwise, we have*

$$\left|\mathcal{W}_{\ell,m}(s,a,p)\right| \leq 2^{4mp}\left(as^2\ell\right)^{2m(g+3)}.$$

We start with an important lemma on the size of the connected components of K_γ. It is based on the assumption that each $\gamma \in W_{\ell,m}$ is made of $2m$ tangle-free paths and that m is not too large.

Lemma 4. *Let $\gamma \in W_{\ell,m}$. Then for any $x \in X_\gamma$, $\mathrm{cc}(x)$ has at most $4m$ elements.*

Proof. The proof is by contradiction. Assume that there exist $x \in X_\gamma$ and $k \geq 2$ such that $2km + 1 \leq |\mathrm{cc}(x)| \leq 2(k+1)m$. Then, from the pigeonhole principle, there exists $i \in [2m]$ such that γ_i visits at least $k+1$ distinct vertices in $\mathrm{cc}(x)$. That is, there exist $1 \leq t_1 < \cdots < t_{k+1} \leq \ell$ such that $z_s := x_{i,t_s}$ are distinct vertices in $\mathrm{cc}(x)$.

Let $B(x,r)$ denote the ball of radius r in the graph K_γ around x. By definition, $B(x,r)$ is contained in the set of $x' \in X_\gamma$ such that $((Q^\top Q)^r)_{xx'} > 0$. We now claim that there exists a pair (s_1, s_2) with $1 \leq s_1 < s_2 \leq k+1$ such that for any $(s,s') \neq (s_1,s_2)$, with $1 \leq s < s' \leq k+1$, we have $B(z_s, h/2) \cap B(z_{s'}, h/2) = \varnothing$. Indeed, otherwise, we could find distinct $s_1 < s_2$ and $s_3 < s_4$ such that the distance between $z_{s_{2p-1}}$ and $z_{s_{2p}}$ is at most h with $p \in \{1,2\}$. In particular, $((Q^\top Q)^h)_{z_{s_{2p-1}}, z_{s_{2p}}} > 0$ and this contradicts the assumption that γ_i is tangle-free.

It follows also that for any $1 \leq s \leq k+1$, $B(z_s, h/2)$ contains at least $h/2$ vertices. Indeed, since $k \geq 2$, we may consider $s' \neq s$ such that $\{s,s'\} \neq \{s_1,s_2\}$. Then from what precedes, the distance between z_s and $z_{s'}$ is at least h (recall that h is even). In particular, the first $h/2$ vertices on the shortest path from z_s to $z_{s'}$ are in $B(z_s, h/2)$. We deduce that for any s,

$$\left|B\left(z_s, \frac{h}{2}\right)\right| \geq \frac{h}{2}.$$

So finally, since $B(z_s, h/2) \cap B(z_{s'}, h/2)$ is empty for all unordered pairs $\{s, s'\}$ with s, s', s_2, pairwise distinct, we have proved that

$$\left|\bigcup_{s=1}^{k+1} B\left(z_s, \frac{h}{2}\right)\right| \geq \sum_{s \neq s_2} \left|B\left(z_s, \frac{h}{2}\right)\right| \geq \frac{kh}{2}.$$

On the other end, $\bigcup_{s=1}^{k+1} B(z_s, h/2)$ is contained in $\mathrm{cc}(x)$. Using that $|\mathrm{cc}(x)| \leq 2(k+1)m$, we deduce that

$$\frac{kh}{2} \leq 2(k+1)m.$$

Hence, since $k \geq 2$,

$$h \leq 4m + \frac{4}{k}m \leq 6m.$$

It contradicts (19). $\square$

Proof of Lemma 3. The proof of Lemma 3 follows closely from [3, Lemma 17] and [2, Lemma 16]. Some new arguments are however introduced to deal with the parameter p. In order to upper bound $|\mathcal{W}_{\ell,m}(s, a, p)|$, we need to find an efficient way to encode the paths $\gamma \in \mathcal{W}_{\ell,m}(s, a, p)$ (that is, find an injective map from $\mathcal{W}_{\ell,m}(s, a, p)$ to a larger set whose cardinality is easier to be upper bounded).

If $\gamma \in W_{\ell,m}$, $i \in [2m]$, $t \in [\ell]$, we set $\gamma_{i,t} = (x_{i,t}, y_{i,t}, x_{i,t+1})$. We shall explore the sequence $(\gamma_{i,t})$ in lexicographic order denoted by $\preceq$ (that is $(i,t) \preceq (i+1, t')$ and $(i,t) \preceq (i, t+1)$). We think of the index (i,t) as a time. We define $(i,t)^-$ as the largest index smaller than (i,t): $(i,t)^- = (i, t-1)$ if $t \geq 2$, $(i,1)^- = (i-1, \ell)$ if $i \geq 2$ and, by convention, $(1,1)^- = (1,0)$.

We now define a relevant information on γ which characterizes its equivalence class. For $y \in Y_\gamma$, we define $\bar{y}$ as the order of apparition of y in the sequence $(y_{i,t})_{i\in[2m],t\in[\ell]}$. Similarly, for $x \in X_\gamma$, $\bar{x}$ is the order of apparition of x in $(x_{i,t})_{i\in[2m],t\in[\ell]}$ and $\bar{\mathrm{cc}}(x)$ is the order of apparition of $\mathrm{cc}(x)$ among the connected components of K_γ. Finally, if $x \in X_\gamma$, we set $\vec{x} = (\bar{x}, s_x)$, where s_x is the set of $\bar{x}'$ with $x' \in X_\gamma$ such that $\bar{x}' < \bar{x}$ and $(Q^\intercal Q)_{xx'} > 0$. For example $\bar{x}_{1,1} = \bar{y}_{1,1} = \bar{\mathrm{cc}}_\gamma(x_{1,1}) = 1$ and $\vec{x}_{1,1} = (1, \varnothing)$. If $x_{1,2} \neq x_{1,1}$ and $(Q^\intercal Q)_{x_{1,1}x_{1,2}} > 0$, we would have $\vec{x}_{1,2} = (2, \{1\})$. Finally, we set $\bar{\gamma}_{i,t} = (\vec{x}_{i,t}, \bar{y}_{i,t}, \vec{x}_{i,t+1})$. By construction, if the sequence $(\bar{\gamma}_{i,t})_{i\in[2m],t\in[\ell]}$ is known then the equivalence class of γ can be determined unambiguously. We thus need to find an encoding of this sequence $(\bar{\gamma}_{i,t})_{i\in[2m],t\in[\ell]}$.

To this end, we start by building a sequence of non-decreasing directed forests which will allow us to find this compact representation of $\gamma \in W_{\ell,m}(s, a, p)$. We set $V_\gamma = [s-p]$, V_γ will be thought as the set of connected components of K_γ ordered by the order of their apparition (since $\gamma \in W_{\ell,m}(s, a, p)$, there are $s-p$ such connected components). We consider the colored directed graph $\Gamma = (V_\gamma, E_\gamma)$ on the vertex set V_γ defined as follows. For each time (i,t), we put the directed edge $e_{i,t} := (\bar{\mathrm{cc}}(x_{i,t}), \bar{\mathrm{cc}}(x_{i,t+1}))$ in E_γ whose *color* is defined as the pair $(\bar{x}_{i,t}, \bar{y}_{i,t})$ (note that Γ may have loop edges of the form (c,c) or multiple edges of the form (c,c') if c is connected to c' by distinct colored edges). By definition, we have $|E_\gamma| = a$. By (20), the graph Γ is weakly connected, that is, after forgetting the direction of the edges of Γ, it becomes a connected undirected graph. Hence the genus of Γ is non-negative:

$$0 \leq g = |E_\gamma| - |V_\gamma| + 1 = a - (s-p) + 1 = a - s + p + 1. \tag{24}$$

This already implies the first claim of the lemma.

We define $\Gamma_{i,t}$ as the subgraph of Γ spanned by the edges $e_{j,s}$ with $(j,s) \preceq (i,t)$. We have $\Gamma_{2m,\ell} = \Gamma$. We now inductively define a spanning forest of $\Gamma_{i,t}$ as follows. $T_{1,0}$ has no edge and a vertex set $\{1\}$. We say that (i,t) is a *first time* if adding the edge $e_{i,t}$ to $T_{(i,t)^-}$ does not create a (weak) cycle. Then, if (i,t) is a first time, we add to $T_{(i,t)^-}$ the edge $e_{i,t}$. It gives $T_{i,t}$. If (i,t) is not a first time, we set $T_{i,t} = T_{(i,t)^-}$. By construction, $T_{i,t}$ is a spanning forest of $\Gamma_{i,t}$. We set $T = T_{2m,\ell}$. Due to (20), we have the following observations.

- If i is odd, $T_{i,t}$ is weakly connected for all $t \in [\ell]$;
- If i is even, $T_{i,t}$ has at most two (weak) connected components for all $t \in [\ell - 1]$ and $T_{i,\ell}$ is weakly connected.

In particular, $T = T_{2m,\ell}$ is a spanning tree of Γ viewed as an undirected graph.

For each even i, we define the *merging time* (i, t_i) as the smallest time (i, t) such that $T_{i,t}$ is weakly connected. Note that the merging time will be a first time if $t_i \geq 2$.

The edges of $\Gamma \backslash T$ will be called *excess edges*. The genus g of Γ defined by (24) is also the number of excess edges:

$$|\Gamma \setminus T| = |E_\gamma| - |V_\gamma| + 1.$$

We call (i, t) an *important time* if the visited edge $e_{i,t}$ is an excess edge.

By construction, the path γ_i can be decomposed by the successive repetition of

(1) a sequence of first times (possibly empty);
(2) an important time or the merging time;
(3) a path using the colored edges of the forest defined so far (possibly empty).

Recall that there is at most one path between two vertices of an oriented forest. Hence, in step (3), it is sufficient to know the starting and ending point to recover the path followed.

We can now build a first encoding of the sequence $(\bar{\gamma}_{i,t})_{i\in[2m],t\in[\ell]}$. Assume that the sequence $(\bar{\gamma}_{j,s})_{(j,s)\prec(i,t)}$ is known and that we seen so far u vertices in X_γ and v elements in Y_γ. Then, we observe that if (i, t) is a first time and not the merging time, $\bar{\gamma}_{i,t}$ is fully determined:

– if $t \geq 2$ or $t = 1$ and i odd, $\vec{x}_{i,t} = \vec{x}_{(i,t)_-+1}$, $\vec{x}_{i,t+1} = (u + 1, \varnothing)$ and $\bar{y}_{i,t} = v + 1$,
– if $t = 1$ and i even, $\vec{x}_{i,1} = (u + 1, \varnothing)$, $\vec{x}_{i,2} = (u + 2, \varnothing)$ and $\bar{y}_{i,1} = v + 1$.

Indeed, if $t \geq 2$ or $t = 1$ and i odd, we have $\vec{x}_{i,t} = \vec{x}_{(i,t)_-+1}$ by (20). Also, since (i, t) is a first time and not the merging time, $\mathrm{cc}(x_{i,t+1})$ has not been seen before. In particular, $x_{i,t+1}$ has not been seen before and for any $(j, s) \prec (i, t)$, $(Q^\intercal Q)_{x_{j,s}x_{i,t+1}} = 0$. It follows that $\vec{x}_{i,t+1} = (u + 1, \varnothing)$. Moreover, if we had $y_{i,t} = y_{j,s}$ for some $(j, s) \prec (i, t)$, then, by definition, $Q_{y_{j,s}x_{j,s+1}} > 0$ and $Q_{y_{j,s}x_{i,t+1}} = Q_{y_{i,t}x_{i,t+1}} > 0$. In particular, $(Q^\intercal Q)_{x_{j,s+1}x_{i,t+1}} > 0$, this contradicts that $\mathrm{cc}(x_{i,t+1})$ has not been seen before. We deduce that $\bar{y}_{i,t} = v + 1$. The case $t = 1$ and i even is similar.

If (i, t) is an important time, we mark the time (i, t) by the vector $(\bar{y}_{i,t}, \bar{x}_{i,t+1}, \bar{x}_{i,\tau})$, where (i, τ) is the next step outside $T_{i,t}$ (by convention, if the path γ_i remains on the forest, we set $\tau = \ell + 1$). By construction, (i, τ) is also the next first, important or merging time. Note that $x_{i,t+1}$ or $x_{i,\tau}$ could be seen for the first time (then by construction, $x_{i,t+1}$ or $x_{i,\tau}$ would belong to a connected component which has already been seen). If this is the case, we replace $\bar{x}_{i,t+1}$ or $\bar{x}_{i,\tau}$ by $\vec{x}_{i,t+1}$ or $\vec{x}_{i,\tau}$ and we call this extra mark the *connected component mark*. Similarly if (i, t) is the merging time, we mark the time (i, t) by the *merging time mark* $(\bar{y}_{i,t}, \bar{x}_{i,t+1}, \bar{x}_{i,\tau})$, where (i, τ) is the next step outside $T_{i,t}$. Again, if $x_{i,t+1}$ or $x_{i,\tau}$ are seen for the first time, we replace $\bar{x}_{i,t+1}$ or $\bar{x}_{i,\tau}$ by the connected component mark. It gives rise to our first encoding of the sequence $(\bar{\gamma}_{i,t})_{i\in[2m],t\in[\ell]}$.

Observe that $p = \sum_{i=1}^{s-p}(l_i - 1)$ where l_i is the size of the i-th connected component. Hence p is equal to the number of connected component marks and it is upper bounded by the twice the number of excess edges plus the number of merging times:

$$p \leq 2(g + m).$$

It proves the second statement of the lemma.

The issue with this first encoding is that the number of important times may be large. This is where the hypothesis that each path γ_i is tangle-free comes into play, more precisely, by Lemma 4 and (19), the path γ_i can visit at most one distinct cycle of Γ (since the diameter of a connected graph is at most its number of vertices).

We are going to partition important times into three categories, namely *short cycling*, *long cycling* and *superfluous* times. For each i, consider the smallest time (i, t_0) such that $\mathrm{cc}(x_{i,t_0+1}) \in \{\mathrm{cc}(x_{i,1}), \dots, \mathrm{cc}(x_{i,t_0})\}$. Let $1 \leq \sigma \leq t_0$ be such that $\mathrm{cc}(x_{i,t_0+1}) = \mathrm{cc}(x_{i,\sigma})$. By assumption, $C_i = (\bar{\mathrm{cc}}(x_{i,\sigma}), \dots, \bar{\mathrm{cc}}(x_{i,t_0+1}))$ will be the unique cycle of Γ visited by γ_i. The last important time $(i, t) \preceq (i, t_0)$ will be called the *short cycling* time. We denote by $(i, \hat{t})$ the smallest time $(i, \hat{t}) \succeq (i, \sigma)$ such that $\bar{\mathrm{cc}}(x_{i,\hat{t}+1})$ is not in C_i (by convention $\hat{t} = \ell + 1$ if γ_i remains on C_i). If $\hat{t} > t_0 + 2$, this means that the cycle C_i has been visited several times from time $(i, t_0 + 1)$ to time $(i, \hat{t})$. We modify the mark of the short cycling time as $(\bar{y}_{i,t}, \bar{x}_{i,t+1}, \sigma, \hat{t}, \bar{x}_{i,\tau})$, where (i, τ), $\tau \geq \hat{t}$, is the next step outside $T_{i,t}$ (it is the next first or important time after $(i, \hat{t})$, by convention $\tau = \ell + 1$ if the path remain on the tree). Important times (i, t') with $1 \leq t' < t$ or $\tau \leq t' \leq \ell$ are called long cycling times. The other important times are called superfluous. The key observation is that for each $i \in [2m]$, the number of long cycling times in γ_i is bounded by $g - 1$ (since there is at most one cycle, no edge of Γ can be seen by γ_i twice outside the time interval between $(i, t + 1)$ and (i, τ), the -1 coming from the fact that the short cycling time is an important time).

We now have our second encoding. We can reconstruct the sequence $(\bar{\gamma}_{i,t})_{i\in[2m],t\in[\ell]}$ from the positions of the merging times, the long cycling and the short cycling times and their respective marks. For each i, there are at most 1 short cycling time, 1 merging time and $g-1$ long cycling times. There are at most $\ell^{2m(g+1)}$ ways to position them. By Lemma 4, for any x, the number of x' such that $(Q^\intercal Q)_{xx'} > 0$ is at most $4m$. Hence, there are at most 2^{4m} possibilities for a connected component mark. Also, note that $|Y_\gamma| \le a$ for any $\gamma \in W_{\ell,m}(s,a,p)$. Thus, there are at most as^2 different possible marks for a long cycling time and $as^2\ell^2$ marks for a short cycling time. Finally, for even i, there are also at most as^2 possibilities for the merging time mark. We deduce that

$$\begin{aligned}\left|\mathcal{W}_{\ell,m}(s,a,p)\right| &\le \ell^{2m(g+1)}\left(2^{4m}\right)^p\left(as^2\right)^m\left(as^2\right)^{2m(g-1)}\left(as^2\ell^2\right)^{2m} \\ &\le \ell^{2m(g+3)}2^{4mp}\left(as^2\right)^{2m(g+1)}.\end{aligned}$$

We find the last statement of the lemma. □

We now estimate of the sum of $q(\gamma)$ over all γ in a single equivalence class. Recall the notion of multiplicity defined above Proposition 1, the multiplicity of an arc $(x,y) \in A_\gamma$ is the number of times (i,t) such that $(x_{i,t}, y_{i,t}) = (x,y)$.

Lemma 5. *Assume further that $m \le \frac{\delta}{8}\frac{\log n}{\log d}$. Then, there exists a constant $c > 0$ (depending on δ) such that for any $\gamma \in W_{\ell,m}(s,a,p)$,*

$$\sum_{\substack{\gamma' \in W_{\ell,m}(s,a,p): \\ \gamma' \sim \gamma}} q\left(\gamma'\right) \le c d^{2g+2(m-1)+a_1+p} n^{s-p} \rho^{2\ell m},$$

where $g = a - s + p + 1$ and a_1 is the number of arcs of A_γ with multiplicity one.

Proof. The proof relies on a decomposition of the product $q(\gamma)$ over edges in the graph $\Gamma = (V_\gamma, E_\gamma)$ defined in the Lemma 3. Let $e = (u,v)$ be an edge of Γ with color $(\bar{x}, \bar{y})$ and multiplicity $k = k(e)$. Let us define the out-degree $b = b(e)$ as the number of distinct elements $\bar{x}_{i,t+1}$ such that $(\bar{x}_{i,t}, \bar{y}_{i,t}) = (\bar{x}, \bar{y})$ (in words, b is the number of distinct elements in the v-th connected component which are visited immediately after a visit of $(\bar{x}, \bar{y})$). Now, the product $q(\gamma)$ can be decomposed as

$$q(\gamma) = \prod_{e \in E_\gamma} Q_{yx_1}^{k_1} \cdots Q_{yx_d}^{k_b}, \tag{25}$$

where $e = (u,v)$ is a generic edge as above and $k_1 + \cdots + k_b = k$, $k_j \ge 1$ and $x_1, \dots, x_b$ are in the v-th connected component of γ.

We thus have the upper bound

$$\sum_{\gamma':\gamma' \sim \gamma} q\left(\gamma'\right) \le \sum_{\star} \prod_{e \in E_\gamma} \left(\sum_y Q_{yx_1'}^{k_1} \cdots Q_{yx_d'}^{k_d}\right), \tag{26}$$

where the first sum $\sum_\star$ is over all possible choices for the elements in $X_{\gamma'}$.

To help the reader, let us first assume that $\|Q\|_{1\to\infty}^{(\delta)} = \|Q\|_{1\to\infty}$ (for example if $\delta = 1$). Then $\rho = \|Q\|_{\mathrm{HS}} \vee \|Q\|_{1\to\infty}$. If $e = (u,v)$ is a generic edge as above, then

$$\sum_y Q_{yx_1}^{k_1} \cdots Q_{yx_b}^{k_b} \le \left\|Q^\intercal\right\|_{1\to 0} \|Q\|_{1\to\infty}^k \le d\rho^k, \tag{27}$$

where we have used

$$\left\|Q^\intercal\right\|_{1\to 0} \le \left\|Q^\intercal Q\right\|_{1\to 0} = d.$$

Besides, if $b = 1$ and $k \ge 2$, we also have the bound

$$\sum_y Q_{yx_1}^k \le \sum_y Q_{yx_1}^2 \|Q\|_{1\to\infty}^{k-2} \le \rho^{k-2} \sum_y Q_{yx_1}^2. \tag{28}$$

We now partition the edges $e = (u,v)$ with color $(\bar{x}, \bar{y})$, multiplicity m and in-degree d in E_γ in three sets, E_1 is the set of edges of multiplicity $k = 1$. E_{21} is the set of edges such that $k \ge 2$ and the v-th connected component is a singleton.

Finally E_{22} is the set of edges such that $k \geq 2$ and the v-th connected component has at least two elements. Note that any edge $e \in E_1 \cup E_{21}$ has out-degree $b = 1$ and by definition $a_1 = |E_1|$. If e is in $E_1 \cup E_{22}$, we use (27), if e is in E_{21}, we use (28). For any $\gamma' \in W_{\ell,m}(s,a,p)$, $\gamma' \sim \gamma$, we arrive at

$$q(\gamma') \leq \prod_{e \in E_1 \cup E_{22}} (d\rho^k) \prod_{e \in E_{21}} \left(\rho^{k-2} \sum_y Q^2_{yx'_1} \right), \tag{29}$$

where in the second product, if $e = (u,v) \in E_{21}$, $x'_1 \in X_{\gamma'}$ is the unique element in the v-th connected component of γ'.

We may now estimate the (26). There are at most $n^{s-p} d^p$ choices for the different elements in $X_{\gamma'}$. The term n^{s-p} accounts for the possibilities of the first element in each of $s-p$ connected components. The term $d^p = \|Q^\intercal Q\|^p_{1\to 0}$ is an upper bound on the choices for the remaining p elements in the connected components (we add the elements one by one in each connected component in an order which preserves connectivity and we use that for any x there at most $\|Q^\intercal Q\|_{1\to 0}$ other x' such that $(Q^\intercal Q)_{xx'} > 0$). In (29), if e is in E_{21}, we may sum over all $x'_1 \in [n]$ (the possibilities for the unique vertex in the v-th connected component), we get

$$\begin{aligned} \sum_{\gamma' : \gamma' \sim \gamma} q(\gamma') &\leq n^{s-p} d^p \prod_{e \in E_1 \cup E_{22}} (d\rho^k) \prod_{e \in E_{21}} (\rho^{k-2} \|Q\|^2_{\mathrm{HS}}) \\ &= n^{s-p} d^{p + a_1 + |E_{22}|} \rho^{2\ell m}, \end{aligned} \tag{30}$$

where we have used that the sum of the multiplicities is equal to $2\ell m$.

It remains to give an upper bound on $|E_{22}|$. To this end, let s_k (respectively $s_{\geq k}$) be the set of vertices of Γ of in-degree k (respectively $\geq k$). We have

$$s_0 + s_1 + s_{\geq 2} = s - p \quad \text{and} \quad s_1 + 2 s_{\geq 2} \leq \sum_k k s_k = a.$$

Subtracting to the right-hand side, twice the left hand side,

$$s_1 \geq 2(s-p) - a - 2s_0 \geq a - 2g - 2m + 2.$$

Indeed, at the last step the bound $s_0 \leq m$ follows from the observation that only a vertex $u \in V_\gamma$ such that $u = \bar{\mathrm{c}}\mathrm{c}(x_{j,1})$ for some $1 \leq j \leq 2m$ can be of in-degree 0. We observe also that $s_1 \leq a_1 + |E_{12}|$ (vertices of in-degree 1 are in bijection with their unique incoming edge, which cannot be in E_{22}). In particular,

$$|E_{22}| = a - a_1 - |E_{12}| \leq a - s_1 \leq 2g + 2m - 2. \tag{31}$$

It concludes the proof when $\|Q\|^{(\delta)}_{1\to\infty} = \|Q\|_{1\to\infty}$.

In the general case, the bounds (27)-(28) remain valid except when x_j or x belong to $\mathcal{E}$. To deal with this case, we first observe the inequality

$$1 = \left(\sum_y Q_{yx} \right)^2 \leq \|Q^\intercal\|_{1\to 0} \sum_y Q^2_{yx} \leq d \sum_y Q^2_{yx}.$$

Summing over x, it implies that

$$\frac{1}{\sqrt{d}} \leq \|Q\|_{\mathrm{HS}} \leq \rho. \tag{32}$$

Hence, in (27)-(28) when x_j or x belong to $\mathcal{E}$, we may use the inequality $Q_{yx} \leq 1 \leq \sqrt{d}\rho$. With the argument leading to (29), we obtain for any $\gamma' \in W_{\ell,m}(s,a,p)$, $\gamma' \sim \gamma$,

$$q(\gamma') \leq d^{u/2} \prod_{e \in E_1 \cup E_{22}} (d\rho^k) \prod_{e \in E_{21} : x'_1 \notin \mathcal{E}} \left(\rho^{k-2} \sum_y Q^2_{yx'_1} \right) \prod_{e \in E_{21} : x'_1 \in \mathcal{E}} \rho^k, \tag{33}$$

where $u = u_{\gamma'}$ is the number of times (i,t), $i \in [2m], t \in [\ell]$ such that $x'_{i,t+1} \in \mathcal{E}$ and Now, for any $\gamma' \in W_{\ell,m}(s,a,p)$ with $\gamma' \sim \gamma$, let $r = r_{\gamma'}$ be the number of connected components which contain at least one element in $\mathcal{E}$. We claim that the number $u_{\gamma'}$ defined in (33) satisfies

$$u \leq 4mr.$$

Indeed, since γ_i is tangle-free for each $i \in [2m]$, γ_i visits at most once each element in $\mathcal{E}$ (to avoid a $\mathcal{E}$-coincidence) and at most 2 distinct elements in each connected components (to avoid two or more than two coincidences). Hence, for each $i \in [2m]$, the number of $t \in [\ell]$ such that $x'_{i,t+1} \in \mathcal{E}$ is at most $2r$. It gives the claimed bound.

We thus deduce from (33) that

$$q(\gamma') \le d^{2mr} \prod_{e \in E_1 \cup E_{22}} (d\rho^k) \prod_{e \in E_{21}: x'_1 \notin \mathcal{E}} \left(\rho^{k-2} \sum_y Q^2_{yx'_1} \right) \prod_{e \in E_{21}: x'_1 \in \mathcal{E}} \rho^k. \tag{34}$$

Now, in view of (34), we should upper bound the number of $\gamma' \in W_{\ell,m}(s,a,p)$, $\gamma' \sim \gamma$ such that $r_{\gamma'} = r$. A rough upper bound is given by

$$\binom{s-p}{r} n^{s-p-r} \left(|\mathcal{E}| d^{4m} \right)^r d^p \le n^{s-p} d^p \left(s d^{4m} n^{-\delta} \right)^r.$$

Indeed, on the left hand side, the binomial term bounds the number of choices for the connected components which contain at least one element in $\mathcal{E}$. As pointed above, the term d^p bounds the possibilities for all but the first element in each connected component. Finally the term $|\mathcal{E}| d^{4m}$ is an upper bound for the number of possibilities of the first element of a connected element which contains an element in $\mathcal{E}$ (by Lemma 4, for any such element, say x_0, there exists a sequence $(x_0, \dots, x_{4m})$ such that $x_{4m} \in \mathcal{E}$ and $(Q^\intercal Q)_{x_{s-1}x_s} > 0$ for all $s \in [4m]$).

Hence, from (34), the argument leading to (30) gives the upper bound

$$\sum_{\gamma': \gamma' \sim \gamma} q(\gamma') \le n^{s-p} d^{p+a_1+|E_{22}|} \rho^{2\ell m} \sum_{r=0}^{s-p} \left(s d^{6m} n^{-\delta} \right)^r.$$

We have $s \le 2\ell m \le 10 \lceil \log n \rceil^{3/2}$ from (19). Hence the assumption $d^{8m} \le n^\delta$ implies that $(s d^{6m} n^{-\delta}) \le 1/2$ for all n large enough. It follows that, for all n large enough, the above geometric series is bounded by 2 and

$$\sum_{\gamma': \gamma' \sim \gamma} q(\gamma') \le 2 n^{s-p} d^{p+a_1+|E_{22}|} \rho^{2\ell m}.$$

From (31), it concludes the proof. □

Recall the definition (22) of $\mu(\gamma)$ of the average contribution of γ in (21). Our final lemma will use Proposition 1 to estimate this average contribution.

Lemma 6. *There is a constant $c > 0$ such that, if $\gamma \in W_{\ell,m}(s,a,p)$, $g = a - s + p + 1$ and a_1 is the number of arcs in A_γ which are visited exactly once in γ, then we have*

$$\left| \mu(\gamma) \right| \le c^{m+g} n^{-a} \left(\frac{6\ell m}{\sqrt{n}} \right)^{(a_1 - 4g - 2m + 2p)_+}.$$

Moreover, $a_1 \ge 2(a - \ell m)$.

Proof. Let $A_1 \subset A_\gamma$ be the set of $e = (x,y)$ which are visited exactly once in γ, that is such that

$$\sum_{i=1}^{2m} \sum_{t=1}^{\ell} \mathbb{1}\left(e = (x_{i,t}, y_{i,t}) \right) = 1.$$

Let A'_1 be the subset of A_1 of consistent arcs and let A_* the set of inconsistent arcs (recall the definition above Proposition 1). We have

$$\left| A'_1 \right| + |A_*| \ge |A_1|.$$

Set $a'_1 = |A'_1|$ and $a_{\ge 2} = |A_\gamma \setminus A_1|$. That is, $a_{\ge 2}$ is the number of $e \in A_\gamma$ which are visited at least twice. We have

$$a_1 + a_{\ge 2} = a \quad \text{and} \quad a_1 + 2a_{\ge 2} \le 2\ell m.$$

Therefore,

$$a_1 \geq 2(a - \ell m).$$

It gives the second claim. Using the terminology of the proof of Lemma 3, a new inconsistent arc can appear after leaving the forest constructed so far, at a first visit of an excess edge, or at the merging time (i even) of γ_i, $i \in [2m]$. Every such step can create 2 inconsistent arcs. A step outside the forest constructed so far is preceded by the visit of a new excess edge. Hence, if $b = |A_*|$, then

$$b \leq 4g + 2m$$

and

$$a_1' \geq a_1 - b.$$

The bound on b can be slightly improved. As already pointed in the proof of Lemma 3, $p = \sum_{i=1}^{s-p}(l_i - 1)$ where l_i is the size of the i-th connected component. The first visit to any element in the connected component beyond the first will be a new excess edge but it will not create an inconsistent arc. It follows that $b \leq 4g + 2m - 2p$ and $a_1' \geq a_1 - 4g - 2m + 2p$. It remains to apply Proposition 1. □

All ingredients have been gathered to prove Proposition 2.

Proof of Proposition 2. We define

$$m = \left\lceil \frac{\delta}{10}\sqrt{\log n} \right\rceil. \tag{35}$$

From (23) and Markov inequality, it suffices to prove that for some $c > 0$,

$$S = \sum_{s,a,p} \left|\mathcal{W}(s,a,p)\right| \max_{\gamma \in W(s,a,p)} \left(\left|\mu(\gamma)\right| \sum_{\substack{\gamma' \in W_{\ell,m}(s,a,p): \\ \gamma' \sim \gamma}} q\left(\gamma'\right) \right) \leq n e^{cm^2}, \tag{36}$$

where $\ell' = \ell + 1 + 1/m$ and $\mu(\gamma)$ was defined in (22).

Let $\gamma \in W_{\ell,m}(s,a,p)$ with a_1 arcs of multiplicity one. Set $g = g(s,a,p) = a - s + p - 1$, by Lemma 5 and Lemma 6,

$$\left|\mu(\gamma)\right| \sum_{\substack{\gamma' \in W_{\ell,m}(s,a,p): \\ \gamma' \sim \gamma}} q\left(\gamma'\right) \leq c d^{2g+2(m-1)+a_1+p} n^{s-p} \rho^{2\ell m} c^{m+g} n^{-a} \left(\frac{6\ell m}{\sqrt{n}} \right)^{(a_1 - 4g - 2m + 2p)_+}.$$

Since $d \geq 1$, we have $d^{a_1} \leq d^{4g+2m-2p} d^{(a_1-4g-2m+2p)_+}$. Using $a_1 \geq 2(a - \ell m)$, we deduce the following upper bound, for some new constant $c > 1$,

$$\left|\mu(\gamma)\right| \sum_{\substack{\gamma' \in W_{\ell,m}(s,a,p): \\ \gamma' \sim \gamma}} q\left(\gamma'\right) \leq (cd)^{6g+4m} n^{-g+1} \rho^{2\ell m} \left(\frac{(6d\ell m)^2}{n} \right)^{(a-(\ell+1)m-2g+p)_+}.$$

For ease of notation, we set

$$\varepsilon = \frac{(6d\ell m)^2}{n} = o(1),$$

where we have used that $d \leq \exp(\sqrt{\log n})$ and $\ell m = O(\log n)^{3/2}$. Now by Lemma 3, since $a \leq 2\ell m$, $s \leq 2\ell m + 1 \leq 3\ell m$, for some new constant $c > 1$ changing from line to line, we arrive at

$$\begin{aligned} S &\leq n\rho^{2\ell m} \sum_{s,a,p: g(s,a,p) \geq 0, p \leq 2g(s,a,p)+2m} 2^{4mp} \left(as^2\ell\right)^{2m(g+3)} (cd)^{6g+4m} n^{-g} \varepsilon^{(s-\ell' m - g)_+} \\ &\leq n(c\ell m)^{24m} (cd)^{4m} \rho^{2\ell m} \sum_{s,g,p: g \geq 0, p \leq 2g+2m} 2^{4mp} (c\ell m)^{8mg} d^{6g} n^{-g} \varepsilon^{(s-\ell' m - g)_+}, \end{aligned}$$

where at the last line, we have performed the change of variable $a \to g = a + p - s + 1$. Then, we may sum over p, using $(\log n)^c = e^{o(m)}$ and $d \le e^{10m/\delta}$, we get for some new constant $c > 0$,

$$S \le ne^{cm^2}\rho^{2\ell m}\sum_{s,g\ge 0}\left(\frac{L}{n}\right)^g \varepsilon^{(s-\ell' m-g)_+},$$

where we have set $L = (c\ell m)^{8m}d^6$. We decompose the above sum as follows

$$S \le S_1 + S_2 + S_3,$$

where S_1 is the sum over $\{1 \le s \le \ell' m, g \ge 0\}$, S_2 over $\{\ell' m < s, 0 \le g \le s - \ell' m\}$, and S_3 over $\{\ell' m < s, g > s - \ell' m\}$. We start with the first term:

$$S_1 = ne^{cm^2}\rho^{2\ell m}\sum_{s=1}^{\ell' m}\sum_{g=0}^{\infty}\left(\frac{L}{n}\right)^g.$$

For our choice of m in (35), for some $c > 0$ and n large enough,

$$\frac{L}{n} = \frac{e^{c(\log\log n)\sqrt{\log n}}}{n} \le \frac{1}{2}.$$

In particular, for n large enough, the above geometric series converges:

$$S_1 \le 2ne^{cm^2}\rho^{2\ell m}\sum_{s=1}^{\ell' m} \le ne^{c' m^2}\rho^{2\ell m}.$$

Adjusting the value of c', the right-hand side of (36) is an upper bound for S_1. Similarly, since $L/(\varepsilon n) \ge 2$, we find

$$\begin{aligned}
S_2 &\le ne^{cm^2}\rho^{2\ell m}\sum_{s=\ell' m+1}^{\infty}\varepsilon^{s-\ell' m}\sum_{g=0}^{s-\ell' m}\left(\frac{L}{\varepsilon n}\right)^g \\
&\le 2ne^{cm^2}\rho^{2\ell m}\sum_{s=\ell' m+1}^{\infty}\varepsilon^{s-\ell' m}\left(\frac{L}{\varepsilon n}\right)^{s-\ell' m} \\
&= 2ne^{cm^2}\rho^{2\ell m}\sum_{k=1}^{\infty}\left(\frac{L}{n}\right)^k.
\end{aligned}$$

Again, for n large enough, the geometric series are convergent and the right-hand side of (36) is an upper bound for S_2. Finally, for n large enough,

$$\begin{aligned}
S_3 &\le ne^{cm^2}\rho^{2\ell m}\sum_{s=\ell' m+1}^{\infty}\sum_{g=s-\ell' m+1}^{\infty}\left(\frac{L}{n}\right)^g \\
&\le ne^{cm^2}\rho^{2\ell m}\sum_{s=\ell' m+1}^{\infty}2\left(\frac{L}{n}\right)^{s-\ell' m+1} \\
&= 2ne^{cm^2}\rho^{2\ell m}\sum_{k=0}^{\infty}\left(\frac{L}{n}\right)^k.
\end{aligned}$$

For n large enough, the right-hand side of (36) is an upper bound for S_3. It concludes the proof. □

4.2. *Operator norm of $R_k^{(\ell)}$*

We now adapt the above subsection for the treatment of $R_k^{(\ell)}$. A rougher bound will suffice for our purposes.

Proposition 3. *Assume $d \le \exp(\sqrt{\log n})$. For any $c_0 > 0$, there exists $c_1 > 0$ (depending on c_0) such that with probability at least $1 - n^{-c_0}$, for all integers $1 \le k \le \ell \le \log n$,*

$$\big\| R_k^{(\ell)} \big\| \le e^{c_1 \sqrt{\log n}}.$$

To help the reader, we use the same notation than in the Section 4.1, we add a prime exponent to our objects when the definition differs from the corresponding definition in Section 4.1.

We fix for some positive integer m such that

$$12m < h. \tag{37}$$

We use the inequality

$$\big\| R_k^{(\ell)} \big\|^{2m} \le \operatorname{tr}\big\{ \big(R_k^{(\ell)} R_k^{(\ell)\mathsf{T}} \big) \big\}.$$

We may expand the trace. To this end, we define $W'_{\ell,m}$ as the set of $\gamma = (\gamma_1, \dots, \gamma_{2m})$ such that $\gamma_i = (x_{i,1}, y_{i,1}, \dots, y_{i,\ell}, x_{i,\ell+1}) \in T^{\ell,k}$ and such that for all $i \in [m]$, the boundary condition (20) holds. Using this notation, the computation leading to (21) gives

$$\big\| R_k^{(\ell)} \big\|^{2m} \le \sum_{\gamma \in W'_{\ell,m}} \prod_{i=1}^{2m} \prod_{t=1}^{k-1} (\underline{M}_{x_{i,t} y_{i,t}}) Q_{y_{i,t} x_{i,t+1}} \cdot Q_{y_{i,k} x_{i,k+1}} \cdot \prod_{t=k+1}^{\ell} M_{x_{i,t} y_{i,t}} Q_{y_{i,t} x_{i,t+1}}. \tag{38}$$

We set

$$\gamma'_i = (x_{i,1}, y_{i,1}, \dots, y_{i,k-1}, x_{i,k}) \quad \text{and} \quad \gamma''_i = (x_{i,k+1}, y_{i,k+1}, \dots, y_{i,\ell}, x_{i,\ell+1}).$$

By construction γ'_i and γ''_i are tangle-free paths.

As in Section 4.1, for $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{2m}) \in W'_{\ell,m}$, we define $X_\gamma = \{x_{i,t} : i \in [2m], t \in [\ell]\}$ and $Y_\gamma = \{y_{i,t} : i \in [2m], t \in [\ell]\}$. We consider the same graph K_γ with vertex set X_γ and, for any x, x' in K_γ, $\{x, x'\}$ is an edge of K_γ if and only if $(Q^\mathsf{T} Q)_{xx'} > 0$. We denote by $\operatorname{cc}(x)$ the connected component of $x \in X_\gamma$ in K_γ. The *arcs* of $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{2m}) \in W'_{\ell,m}$, denoted by A'_γ, is the set of distinct pairs $(x_{i,t}, y_{i,t})$ with $t \ne k$. We define $W'_{\ell,m}(s,a,p)$ as the set of $\gamma \in W_{\ell,m}$ with $s = |X_\gamma|$, $a = |A'_\gamma|$ and $s - p$ connected components in K_γ. We take the expectation in (38) and write

$$\mathbb{E} \big\| R_k^{(\ell)} \big\|^{2m} \le \sum_{s,a,p} \sum_{\gamma \in W'_{\ell,m}(s,a,p)} \mu'(\gamma) q(\gamma),$$

where for $\gamma \in W'_{\ell,m}$, we have defined

$$\mu'(\gamma) := \mathbb{E} \prod_{i=1}^{2m} \prod_{t=1}^{k-1} \underline{M}_{x_{i,t}, y_{i,t}} \prod_{t=k+1}^{\ell} M_{x_{i,t}} \quad \text{and} \quad q(\gamma) = \prod_{i=1}^{2m} \prod_{t=1}^{\ell} Q_{y_{i,t} x_{i,t+1}}. \tag{39}$$

We decompose further $W'_{\ell,m}(s,a,p)$ into equivalence classes as follows. For $\gamma, \gamma' \in W'_{\ell,m}(s,a,p)$, let us say $\gamma \sim \gamma'$ if there exist a pair of permutations α and β in S_n such that the image of K_γ by α is $K_{\gamma'}$ and for any (i,t), $x'_{i,t} = \alpha(x_{i,t})$, $y'_{i,t} = \beta(y_{i,t})$ (where $\gamma' = (\gamma'_1, \gamma'_2, \dots, \gamma'_{2m})$ with $\gamma'_i = (x'_{i,1}, y'_{i,1}, \dots, y'_{i,\ell}, x'_{i,\ell+1})$). We define $\mathcal{W}'_{\ell,m}(s,a,p)$ as the set of equivalence classes. Since $\mu(\gamma) = \mu(\gamma')$ if $\gamma \sim \gamma'$, we obtain the bound,

$$\mathbb{E} \big\| R_k^{(\ell)} \big\|^{2m} \le \sum_{s,a,p} \big| \mathcal{W}'(s,a,p) \big| \max_{\gamma \in W'(s,a,p)} \Bigg(\big| \mu'(\gamma) \big| \sum_{\substack{\gamma' \in W_{\ell,m}(s,a,p): \\ \gamma' \sim \gamma}} q(\gamma') \Bigg). \tag{40}$$

We start by bounding the cardinality of $\mathcal{W}'_{\ell,m}(s,a,p)$.

Lemma 7. *If $g' := a + p - s < 0$ or $2g' + 10m > p$, then $W'_{\ell,m}(s,a,p)$ is empty. Otherwise, we have*

$$\big| \mathcal{W}'_{\ell,m}(s,a,p) \big| \le 2^{4mp} \big((a+2m)^2 s^2 \ell \big)^{4m(g'+4)}.$$

We have the following analog of Lemma 4.

Lemma 8. *Let $\gamma \in W'_{\ell,m}$. Then for any $x \in X_\gamma$, $\mathrm{cc}(x)$ has at most $8m$ elements.*

Proof. We repeat the proof of Lemma 4, we use this time that γ is composed of $4m$ tangle-free paths: γ'_i, γ''_i, for $i \in [2m]$. By contradiction, we assume that there exist $x \in X_\gamma$ and $k \geq 2$ such that $4km + 1 \leq |\mathrm{cc}(x)| \leq 4(k+1)m$. Then, from the pigeonhole principle, there exists $i \in [2m]$ and $\varepsilon \in \{','' \}$ such that γ_i^ε visits at least $k+1$ distinct vertices in $\mathrm{cc}(x)$. We then repeat verbatim the proof of Lemma 4 and use (37). $\square$

Proof of Lemma 7. We repeat the proof of Lemma 3. If $\gamma \in W'_{\ell,m}$, $i \in [2m]$, $t \in [\ell]$, we set $\gamma_{i,t} = (x_{i,t}, y_{i,t}, x_{i,t+1})$. We shall explore the sequence $(\gamma_{i,t})$ in lexicographic order denoted by $\preceq$ (that is $(i,t) \preceq (i+1,t')$ and $(i,t) \preceq (i,t+1)$). We think of the index (i,t) as a time. We define $(i,t)^-$ as the largest index smaller than (i,t) and, by convention, $(1,1)^- = (1,0)$.

As in Lemma 3, for $y \in Y_\gamma$, we define $\bar{y}$ as the order of apparition of y in the sequence $(y_{i,t})_{i\in[2m],t\in[\ell]}$. Similarly, for $x \in X_\gamma$, $\bar{x}$ is the order of apparition of x in $(x_{i,t})_{i\in[2m],t\in[\ell]}$ and $\bar{\mathrm{c}}\mathrm{c}(x)$ is the order of apparition of $\mathrm{cc}(x)$ among the connected components of K_γ. Finally, if $x \in X_\gamma$, we set $\vec{x} = (\bar{x}, s_x)$, where s_x is the set of $\bar{x}'$ with $x' \in X_\gamma$ such that $\bar{x}' < \bar{x}$ and $(Q^\intercal Q)_{xx'} > 0$. Finally, we set $\bar{\gamma}_{i,t} = (\vec{x}_{i,t}, \bar{y}_{i,t}, \vec{x}_{i,t+1})$. By construction, if the sequence $(\bar{\gamma}_{i,t})_{i\in[2m],t\in[\ell]}$ is known then the equivalence class of γ can be determined unambiguously. We thus need to find an encoding of this sequence $(\bar{\gamma}_{i,t})_{i\in[2m],t\in[\ell]}$.

We set $V_\gamma = [s-p]$ and consider the colored directed graph $\Gamma' = (V_\gamma, E'_\gamma)$ on the vertex set V_γ defined as follows. For each time (i,t), with $t \neq k$, we put the directed edge $e_{i,t} := (\bar{\mathrm{c}}\mathrm{c}(x_{i,t}), \bar{\mathrm{c}}\mathrm{c}(x_{i,t+1}))$ in E'_γ whose *color* is defined as the pair $(\bar{x}_{i,t}, \bar{y}_{i,t})$. By definition, we have $|E'_\gamma| = a$. Let $\bar{\Gamma}'$ be the associated undirected graph (that is the undirected graph obtained by forgetting the direction of the edges of Γ'). We observe that each connected component of $\bar{\Gamma}'$ contains at least a cycle. Indeed, by assumption γ_i is tangled while γ'_i and γ''_i is tangle-free. Hence if the image of the paths of γ'_i and γ'_{ii} on $\bar{\Gamma}'$ do not intersect then each one contains a distinct cycle. Otherwise, the images of the paths intersect, then they are in the same connected component of $\bar{\Gamma}'$ and their union has at least two distinct cycles. Hence the number of edges of Γ' is at least the number of vertices:

$$0 \leq g' = |E_\gamma| - |V_\gamma| = a - s + p.$$

This is the first claim of the lemma.

We define $\Gamma'_{i,t}$ as the subgraph of Γ' spanned by the edges $e_{j,s}$ with $(j,s) \preceq (i,t)$. We have $\Gamma'_{2m,\ell} = \Gamma'$. As in Lemma 3, we now inductively define a spanning forest $T_{i,t}$ of $\Gamma'_{i,t}$ as follows. $T_{1,0}$ has no edge and a vertex set $\{1\}$. We say that (i,t) is a *first time* if adding the edge $e_{i,t}$ to $T_{(i,t)^-}$ does not create a (weak) cycle. Then, if (i,t) is a first time, we add to $T_{(i,t)^-}$ the edge $e_{i,t}$. It gives $T_{i,t}$. If (i,t) is not a first time, we set $T_{i,t} = T_{(i,t)^-}$. We set $T = T_{2m,\ell}$.

For each even i, we define the *first merging time* (i,t'_i) as the smallest time (i,t) with $1 \leq t \leq k-1$ such that $T_{i,t}$ and $T_{(i,1)^-}$ have the same number of connected components. If this time does not exist, we set $t'_i = k$. Similarly, for each i, the *second merging time* (i,t''_i) is the smallest time (i,t) with $k \leq t \leq \ell$ such that $T_{i,t}$ and $T_{(i,k)^-}$ have the same number of connected components. If this time does not exist, we set $t''_i = \ell + 1$. If i is even then by (20), we have $t''_i \leq \ell$.

Note that the merging time will be a first time if $t_i \geq 2$.

The edges of $\Gamma' \backslash T$ will be called *excess edges*. We call (i,t) an *important time* if the visited edge $e_{i,t}$ is an excess edge. The total number of excess edges is $|E_\gamma| - |V_\gamma| + N_\gamma = g' + N_\gamma$ where $1 \leq N_\gamma \leq 2m$ is the number of connected components of $\bar{\Gamma}'$. However, since each connected component has at least a cycle, in each connected component of T, there are at most $g' + 1$ excess edges.

By construction, the path γ'_i or γ''_i can be decomposed by the successive repetition of

(1) a sequence of first times (possibly empty);
(2) an important time or the merging time;
(3) a path using the colored edges of the forest defined so far (possibly empty).

We build a first encoding of the sequence $(\bar{\gamma}_{i,t})_{i\in[2m],t\in[\ell]}$ as follows. If (i,t) is an important time, we mark the time (i,t) by the vector $(\bar{y}_{i,t}, \bar{x}_{i,t+1}, \bar{x}_{i,\tau})$, where (i,τ) is the next step outside $T_{i,t}$ (by convention, if the path γ_i remains on the forest, we set $\tau = \ell + 1$). By construction, (i,τ) is also the next first, important or merging time. Note that $x_{i,t+1}$ or $x_{i,\tau}$ could be seen for the first time (then by construction, $x_{i,t+1}$ or $x_{i,\tau}$ would belong to a connected component which has already been seen). If this is the case, we replace $\bar{x}_{i,t+1}$ or $\bar{x}_{i,\tau}$ by $\vec{x}_{i,t+1}$ or $\vec{x}_{i,\tau}$ and we call this extra mark the *connected component mark*. Similarly if (i,t) is a first merging time, we mark the time (i,t) by the *first merging time mark* $(\bar{y}_{i,t}, \bar{x}_{i,t+1}, \bar{x}_{i,\tau})$, where (i,τ) is the next step outside $T_{i,t}$. Similarly, the *second merging time mark* is $(\bar{y}_{i,k}, \bar{y}_{i,t}, \bar{x}_{i,t+1}, \bar{x}_{i,\tau})$. Again, if $x_{i,t+1}$ or $x_{i,\tau}$ are seen for the first time, we replace $\bar{x}_{i,t+1}$ or $\bar{x}_{i,\tau}$ by the connected component mark. Arguing as in the proof of Lemma 3, it gives a first encoding of the sequence $(\bar{\gamma}_{i,t})_{i\in[2m],t\in[\ell]}$.

Observe that $p = \sum_{i=1}^{s-p}(l_i - 1)$ where l_i is the size of the i-th connected component of K_γ. Hence p is equal to the number of connected component marks and it is upper bounded by twice the number of excess edges plus the number of merging times:

$$p \le 2\big(g' + N_\gamma + 3m\big) \le 2g' + 10m.$$

It proves the second statement of the lemma.

Arguing as in the proof of Lemma 3, to improve on the first encoding we use the hypothesis that each path γ_i' or γ_i'' is tangle-free. We partition important times into three categories *short cycling*, *long cycling* and *superfluous* times. For each i and $\varepsilon \in \{', ''\}$, consider the smallest time (i, t_0) such that $\mathrm{cc}(x_{i,t_0+1}) \in \{\mathrm{cc}(x_{i,1}), \dots, \mathrm{cc}(x_{i,t_0})\}$. Let $1 \le \sigma \le t_0$ be such that $\mathrm{cc}(x_{i,t_0+1}) = \mathrm{cc}(x_{i,\sigma})$. By assumption, $C_i = (\bar{\mathrm{cc}}(x_{i,\sigma}), \dots, \bar{\mathrm{cc}}(x_{i,t_0+1}))$ will be the unique cycle of Γ' visited by γ_i^ε. The last important time $(i,t) \preceq (i,t_0)$ will be called the *short cycling* time. We denote by $(i,\hat{t})$ the smallest time $(i,\hat{t}) \succeq (i,\sigma)$ such that $\bar{\mathrm{cc}}(x_{i,\hat{t}+1})$ is not in C_i (by convention $\hat{t} = \ell + 1$ if γ_i^ε remains on C_i). We modify the mark of the short cycling time as $(\bar{y}_{i,t}, \bar{x}_{i,t+1}, \sigma, \hat{t}, \bar{x}_{i,\tau})$, where (i,τ), $\tau \ge \hat{t}$, is the next step outside $T_{i,t}$ (it is the next first or important time after $(i,\hat{t})$, by convention $\tau = \ell + 1$ if the path remain on the tree). Important times (i,t') with $1 \le t' < t$ or $\tau \le t' \le \ell$ are called long cycling times. The other important times are called superfluous. As argued in the proof of Lemma 3, for each $i \in [2m]$ and $\varepsilon \in \{', ''\}$, the number of long cycling times in γ_i^ε is bounded by g' (recall that there are at most $g' + 1$ excess edges in the connected component of γ_i^ε).

We now have our second encoding. We can reconstruct the sequence $(\bar{\gamma}_{i,t})_{i\in[2m],t\in[\ell]}$ from the positions of the merging times, the long cycling and the short cycling times and their respective marks. For each i and $\varepsilon \in \{', ''\}$, there are at most 1 short cycling time, 1 merging times and g' long cycling times. There are at most $\ell^{4m(g'+2)}$ ways to position them. Note that $|Y_\gamma| \le a + 2m = a'$, the term $2m$ coming from the elements $y_{i,k}$, $i \in [2m]$. Hence, as argued in the proof of Lemma 3, there are at most 2^{4m} possibilities for a connected component mark, at most $a's^2$ different possible marks for a long cycling time, $a's^2\ell^2$ marks for a short cycling time, at most $a's^2$ marks for the first merging time mark and a'^2s^2 for the second merging time. We deduce that

$$\begin{aligned}\big|\mathcal{W}'_{\ell,m}(s,a,p)\big| &\le \ell^{4m(g'+2)}\big(2^{4m}\big)^p\big(a's^2\big)^m\big(a'^2s^2\big)^{2m}\big(a's^2\big)^{4mg'}\big(a's^2\ell^2\big)^{4m}. \\ &\le \ell^{4m(g'+4)}2^{4mp}\big(a'^2s^2\big)^{4m(g'+1)}.\end{aligned}$$

It concludes the proof. □

Lemma 9. *For any $\gamma \in W'_{\ell,m}(s,a,p)$, we have*

$$\sum_{\substack{\gamma' \in W'_{\ell,m}(s,a,p): \\ \gamma' \sim \gamma}} q\big(\gamma'\big) \le d^p n^{s-p}.$$

Proof. The proof follows easily from the proof of Lemma 5. Let $\Gamma' = (V_\gamma, E'_\gamma)$ be the graph defined in Proposition 7. Arguing as in (26), we have an upper bound of the form

$$\sum_{\gamma':\gamma'\sim\gamma} q\big(\gamma'\big) \le \sum_{\star} \prod_{e\in E_\gamma} \bigg(\sum_y Q^{k_1}_{yx'_1} \cdots Q^{k_b}_{yx'_b}\bigg),$$

where the first sum $\sum_\star$ is over all possible choices for the distinct elements in $X_{\gamma'}$, and the positive integers k_j and the elements $x'_j \in X_{\gamma'}$ are determined by the edge e. Since $k_j \ge 1$ and $\sum_y Q_{yx} = 1$, we have

$$\sum_y Q^{k_1}_{yx'_1} \cdots Q^{k_d}_{yx'_d} \le 1.$$

It follows that $\sum_{\gamma':\gamma'\sim\gamma} q(\gamma')$ is upper bounded by number of possible choices for $X_{\gamma'}$. The latter is bounded by $d^p n^{s-p}$ as explained in the proof of Lemma 5. □

We finally estimate $\mu'(\gamma)$.

Lemma 10. *There is a constant $c > 0$ such that, if $\gamma \in W_{\ell,m}(s,a,p)$, $g = a - s + p$ and a_1 is the number of arcs in A_γ which are visited exactly once in γ, then we have*

$$|\mu'(\gamma)| \le c^{m+g'} n^{-a}.$$

Proof. Let A_* be the set of inconsistent arcs of A'_γ (as defined above Proposition 1). Using the terminology of the proof of Proposition 7 and as argued in Lemma 10, $|A_*|$ is upper bounded by four times the number of excess edges plus twice the number of merging times. There are at most $g' + 2m$ excess edges and $3m$ merging times, hence,

$$|A_*| \le 4(g' + 2m) + 6m.$$

It remains to apply Proposition 1. □

We are ready to prove Proposition 3.

Proof of Proposition 3. We define

$$m = \lceil \sqrt{\log n} \rceil. \tag{41}$$

For this choice of m, $n^{1/m} \le \exp(\sqrt{\log n})$. Hence, from (23) and Markov inequality, it suffices to prove that for some $c > 0$,

$$S = \sum_{s,a,p} |\mathcal{W}'(s,a,p)| \max_{\gamma \in W'(s,a,p)} \Bigg(|\mu'(\gamma)| \sum_{\substack{\gamma' \in W_{\ell,m}(s,a,p): \\ \gamma' \sim \gamma}} q(\gamma') \Bigg) \le e^{cm^2}. \tag{42}$$

Let $\gamma \in W'_{\ell,m}(s,a,p)$. Set $g' = g'(s,a,p) = a - s + p$, by Lemma 9 and Lemma 10,

$$|\mu'(\gamma)| \sum_{\substack{\gamma' \in W'_{\ell,m}(s,a,p): \\ \gamma' \sim \gamma}} q(\gamma') \le d^p c^{m+g'} n^{-g'}.$$

Now, by Lemma 7, since $a \le 2\ell m$, $s \le 2\ell m + 1 \le 3\ell m$, for some new constant $c > 1$ changing from line to line,

$$\begin{aligned} S &\le \sum_{s,a,p: g'(s,a,p) \ge 0, p \le 2g'(s,a,p) + 10m} 2^{4mp} \big((a+2m)^2 s^2 \ell\big)^{4m(g'+4)} d^p n^{-g'} c^{m+g'} \\ &\le c^m (c\ell m)^{80m} \sum_{s,g',p: g' \ge 0, p \le 2g' + 10m} 2^{4mp} (c\ell m)^{20mg'} d^p n^{-g'}, \end{aligned}$$

where at the last line, we have performed the change of variable $a \to g' = a + p - s$. Then, we may sum over p, using $(\log n)^c = e^{o(m)}$ and $d \le e^m$, we get for some new constant $c > 0$,

$$S \le e^{cm^2} \sum_{s,g' \ge 0} \left(\frac{L}{n}\right)^{g'},$$

where we have set $L = (c\ell m)^{20m}$. Since $s \le 3\ell m = e^{o(m)}$ and $L/n = o(1)$, we deduce that (42) holds. □

5. Proof of Theorem 4

All ingredients are finally gathered to prove Theorem 4. We start by reducing the range of ℓ and d where there is something to be proven. Up to adjusting the final constant c_1, we may assume without loss of generality that $d \le \exp(\sqrt{\log n})$ and $\ell \le \log n / \log d$ (otherwise the probabilistic bound is larger than 1). We fix any $0 < c_0 < c'_0 < \delta$. Then by Lemma 2 and Lemma 1, if Ω is the event that G is ℓ-tangle-free, for any $c > 0$,

$$\begin{aligned} \mathbb{P}\big(\|P^\ell_{|\mathbf{1}^\perp}\| \ge e^{c\sqrt{\log n}} \rho^\ell\big) &= \mathbb{P}\big(\|P^\ell_{|\mathbf{1}^\perp}\| \ge e^{c\sqrt{\log n}} \rho^\ell; \Omega\big) + O\big(d^{\ell+2h} n^{-c'_0}\big) \\ &\le \mathbb{P}\big(J \ge e^{c\sqrt{\log n}} \rho^\ell\big) + O\big(d^{\ell+2h} n^{-c'_0}\big), \end{aligned}$$

where

$$J = \|\underline{P}^{(\ell)}\| + \frac{1}{n}\sum_{k=1}^{\ell}\|R_k^{(\ell)}\|.$$

On the other end, by Propositions 2-3, for some $c_1' > 0$, with probability at least $1 - 2n^{-c_0'}$,

$$\begin{aligned} J &\le e^{c_1'\sqrt{\log n}}\rho^\ell + \frac{1}{n}\sum_{k=1}^{\ell} e^{c_1'\sqrt{\log n}} \\ &\le \left(e^{c_1'\sqrt{\log n}} + \ell e^{\frac{\ell}{2}\log d - \log n}\right)\rho^\ell, \end{aligned}$$

where we have used $\rho \ge 1/\sqrt{d}$ by (32). Since $\ell \le \log n/\log d$, we find that the event

$$J \le \left(e^{c_1'\sqrt{\log n}} + \frac{\ell}{\sqrt{n}}\right)\rho^\ell$$

has probability at least $1 - 2n^{-c_0'}$. We take any $c > c_1'$ and it remains to adjust the final constant $c_1 > c$ to deal with bounded values of n. It concludes the proof of Theorem 4.

Remark 2. Lemma 2 and Proposition 1 are the only properties of the uniform measures on $\mathbb{S}_n$ which have been used in the proof. Proposition 1 is used in Lemma 6 and Lemma 10 where we use that the number of inconsistent arcs is at most $c(g + m)$. The proof may thus be extended to other probability measures on $\mathbb{S}_n$ with other notions of inconsistency. For example, if n is even, the set of matching $\mathbb{M}_n$ is the subset of permutations $\sigma \in \mathbb{S}_n$ such that $\sigma(x) \ne x$ and $\sigma^2(x) = x$ for all $x \in [n]$. Following [2], analogs of Lemma 2 and Proposition 1 hold for the uniform measure on $\mathbb{M}_n$ (the definition of a consistent arc is slightly more constrained for matchings, but in Lemma 6 and Lemma 10, we may still upper bound the number of inconsistent arcs by $c(m + g)$).

Remark 3. Proposition 2 and Proposition 3 are true beyond bistochastic matrices. An inspection of the proof reveals that they hold for any matrix Q provided that $\max_x \sum_y |Q_{xy}| \le c$ for some constant $c > 0$ (which will have an influence on all other constants).

6. Proof of corollaries

6.1. *Proof of Theorem* 2

By construction, we have $Q_{xy} = \mathbf{1}((x, y) \in E)/r$. It follows that

$$\|Q\|_{1\to\infty} = \frac{1}{r} \quad \text{and} \quad \|Q\|_{\mathrm{HS}} = \frac{1}{\sqrt{r}}. \tag{43}$$

It remains to apply Theorem 1 with $\delta = 1$.

6.2. *Proof of Corollary* 1

Let $\mathcal{P}$ be the set of bi-stochastic matrices of size n with entries in $\{0, 1/r\}$. From the proof of Theorem 2, for any $Q \in \mathcal{P}$, (43) holds. Note that $A = MB$ for some permutation matrix M is equivalent to $M^*A = B$. It follows that for any permutation matrix M, if P is uniformly sampled over $\mathcal{P}$, P and MP have the same distribution. In particular, P and MP have the same distribution for M uniformly distributed and independent of P. We may thus apply Theorem 2 to MP by conditioning on the value of P.

6.3. *Proof of Theorem* 3

Up to increasing the constant c_1, we may assume that $r \le \exp(\sqrt{\log n})$. Obviously, if $x \notin S$,

$$\max_y Q_{xy} = \max_i p_i \le \sqrt{\sum_i p_i^2}.$$

From our assumption on S, it follows that $\|Q\|^{(\delta)}_{1\to\infty} \le \sqrt{\sum_i p_i^2}$.

Moreover, we have

$$Q^{\mathsf{T}} Q = \sum_{i,j} p_i p_j M_i^* M_j = \sum_i p_i^2 I + \sum_{j\neq i} p_i p_j M_i^* M_j.$$

From the triangle inequality, we deduce that

$$\|Q\|_{\mathrm{HS}} \le \left\| \sum_i p_i^2 I \right\|_{\mathrm{HS}} + \sum_{j\neq i} p_i p_j \|M_i^* M_j\|_{\mathrm{HS}} = \sqrt{\sum_i p_i^2} + \sum_{i\neq j} p_i p_j \sqrt{\frac{1}{n}\sum_{x=1}^n \mathbf{1}\big(\sigma_i(x) = \sigma_j(x)\big)}.$$

It follows that $\|Q\|_{\mathrm{HS}} \le \rho + \sqrt{|S|/n} \le (1 + r^{1/2} n^{-\delta/2})\rho$ (where we have used $\sum_i p_i = 1$ and $\sum_i p_i^2 \ge 1/r$). It remains to apply Theorem 1.

6.4. *Proof of Corollary* 2

Let $0 < c_0 < 1$ and fix some $c_0 < \delta < 1$. Up to increasing the constant c_1, we may assume that $r \le \exp(\sqrt{\log n})$. For any permutation matrix M, P has the same distribution as MP. In particular, P and MP have the same distribution for M uniformly distributed and independent of $M_1, \dots, M_r$. Now, let $S = \{x \in [n] : \exists i \neq j, \sigma_i(x) = \sigma_j(x)\}$. From the union bound, we have

$$\mathbb{E}|S| \le r(r-1)\mathbb{P}\big(\sigma_1(x) = \sigma_2(x)\big) = \frac{r(r-1)}{n}.$$

Hence, from Markov inequality,

$$\mathbb{P}\big(|S| \ge n^{1-\delta}\big) \le r^2 n^{\delta-2}.$$

Finally, on the event $\{S < n^{1-\delta}\}$, we apply Theorem 3 for MP by conditioning on the value of P.

References

[1] A. Basak, N. Cook and O. Zeitouni. Circular law for the sum of random permutation matrices. *Electron. J. Probab.* **23** (33) (2018) 51. MR3798243 https://doi.org/10.1214/18-EJP162

[2] C. Bordenave. A new proof of Friedman's second eigenvalue theorem and its extension to random lifts. *Ann. Sci. Éc. Norm. Supér.* To appear.

[3] C. Bordenave, M. Lelarge and L. Massoulié. Nonbacktracking spectrum of random graphs: Community detection and nonregular Ramanujan graphs. *Ann. Probab.* **46** (1) (2018) 1–71. MR3758726 https://doi.org/10.1214/16-AOP1142

[4] G. Brito, I. Dumitriu and K. D. Harris. Spectral gap in random bipartite biregular graphs and its applications. Available at arXiv:1804.07808.

[5] N. Cook. The circular law for random regular digraphs. Available at arXiv:1703.05839. MR4029149 https://doi.org/10.1214/18-AIHP943

[6] S. Coste. The spectral gap of sparse random digraphs. Available at arXiv:1708.00530.

[7] A. Figà-Talamanca and T. Steger. Harmonic analysis for anisotropic random walks on homogeneous trees. *Mem. Amer. Math. Soc.* **110** (531) (1994) xii+68. MR1219707 https://doi.org/10.1090/memo/0531

[8] Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica* **1** (3) (1981) 233–241. MR0637828 https://doi.org/10.1007/BF02579329

[9] A. Guionnet, M. Krishnapur and O. Zeitouni. The single ring theorem. *Ann. of Math. (2)* **174** (2) (2011) 1189–1217. MR2831116 https://doi.org/10.4007/annals.2011.174.2.10

[10] A. Guionnet and O. Zeitouni. Support convergence in the single ring theorem. *Probab. Theory Related Fields* **154** (3–4) (2012) 661–675. MR3000558 https://doi.org/10.1007/s00440-011-0380-5

[11] U. Haagerup and F. Larsen. Brown's spectral distribution measure for R-diagonal elements in finite von Neumann algebras. *J. Funct. Anal.* **176** (2) (2000) 331–367. MR1784419 https://doi.org/10.1006/jfan.2000.3610

[12] D. A. Levin, Y. Peres and E. L. Wilmer. *Markov Chains and Mixing Times*. J. G. Propp and D. B. Wilson (Eds), 2nd edition. American Mathematical Society, Providence, RI, 2017. MR3726904

[13] A. Litvak, A. Lytova, K. Tikhomirov, N. Tomczak-Jaegermann and P. Youssef. Circular law for sparse random regular digraphs. Available at arXiv:1801.05576.

[14] L. Massoulié. Community detection thresholds and the weak Ramanujan property. In *STOC'14—Proceedings of the 2014 ACM Symposium on Theory of Computing* 694–703. ACM, New York, 2014. MR3238997

[15] J. A. Mingo and R. Speicher. *Free Probability and Random Matrices*. *Fields Institute Monographs* **35**. Springer, New York, 2017. MR3585560 https://doi.org/10.1007/978-1-4939-6942-5

[16] M. Rudelson and R. Vershynin. Invertibility of random matrices: Unitary and orthogonal perturbations. *J. Amer. Math. Soc.* **27** (2) (2014) 293–338. MR3164983 https://doi.org/10.1090/S0894-0347-2013-00771-7