



**HAL**  
open science

## Large permutation invariant random matrices are asymptotically free over the diagonal

Benson Au, Guillaume Cébron, Antoine Dahlqvist, Franck Gabriel, Camille Male

► **To cite this version:**

Benson Au, Guillaume Cébron, Antoine Dahlqvist, Franck Gabriel, Camille Male. Large permutation invariant random matrices are asymptotically free over the diagonal. *Annals of Probability*, inPress. hal-01824543

**HAL Id: hal-01824543**

**<https://hal.science/hal-01824543>**

Submitted on 27 Jun 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Large permutation invariant random matrices are asymptotically free over the diagonal

Benson Au

Department of Mathematics  
University of California  
Berkeley, CA 94720-3840

Guillaume Cébron

IMT; UMR 5219  
Université de Toulouse; CNRS  
UPS, F-31400 Toulouse; France

Antoine Dahlqvist

University College Dublin  
School of Mathematics and Statistics  
Belfield Dublin 4; Ireland

Franck Gabriel

Imperial College London  
South Kensington Campus  
London SW7 2AZ; UK

Camille Male

IMB; UMR 5251  
Université de Bordeaux; CNRS  
F-33405 Talence; France

May 23, 2018

## Abstract

We prove that independent families of permutation invariant random matrices are asymptotically free over the diagonal, both in probability and in expectation, under a uniform boundedness assumption on the operator norm. We can relax the operator norm assumption to an estimate on sums associated to graphs of matrices, further extending the range of applications (for example, to Wigner matrices with exploding moments and so the sparse regime of the Erdős-Rényi model). The result still holds even if the matrices are multiplied entrywise by bounded random variables (for example, as in the case of matrices with a variance profile and percolation models).

## 1 Introduction

Non-commutative probability is a generalization of classical probability that extends the probabilistic perspective to non-commuting random variables. Following the seminal work of Voiculescu [26], this setting provides a robust framework for studying the spectral theory of random multi-matrix models in the large  $N$  limit. We outline the basic approach of this program as follows.

1. In the non-commutative framework, Voiculescu's notion of *free independence* plays the role of classical independence. This simple parallel yields a surprisingly rich theory, with free analogues of many classical concepts (e.g., the free CLT, free cumulants, free entropy, and conditional expectations). The scope of free probability further benefits from a robust analytic framework, allowing for the *free harmonic analysis* [24].

- Free independence describes the large  $N$  limit behavior of independent random matrices in many generic situations: for example, unitarily invariant random matrices [26] and Wigner matrices [11]. The free probability machinery allows for tractable computations. In particular, we can compute the limiting spectral distributions of rational functions in these random matrices.

Yet, many random matrix models of interest lie beyond the scope of free probability. One can accommodate such models by defining a new framework. This is the perspective of traffic probability [2, 10, 12, 13, 14, 16, 17, 18].

- The traffic framework enlarges the non-commutative probability framework. The notion of traffic distribution is richer and allows to consider a notion of independence which encodes the non-commutative notions of independences.
- Permutation invariant random matrices provide a canonical model of traffic independence in the large  $N$  limit.

The purpose of this article is to demonstrate that Voiculescu's notion of conditional expectation provides an analytic tool for traffic independence, in the context of large random matrices.

## 1.1 Background

We first recall the notion of freeness with amalgamation [25].

**Definition 1.1.** • An operator-valued probability space is a triple  $(\mathcal{A}, \mathcal{B}, E)$  consisting of a unital algebra  $\mathcal{A}$ , a unital subalgebra  $\mathcal{B} \subset \mathcal{A}$  and a conditional expectation  $E : \mathcal{A} \rightarrow \mathcal{B}$ , i.e. a unit-preserving linear map such that  $E(b_1 a b_2) = b_1 E(a) b_2$  for any  $a \in \mathcal{A}$  and  $b_1, b_2 \in \mathcal{B}$ .

- Let  $K$  be an arbitrary index set. We write  $\mathcal{B}\langle X_k : k \in K \rangle$  for the algebra generated by freely noncommuting indeterminates  $(X_k)_{k \in K}$  and  $\mathcal{B}$ . For a monomial  $P = b_0 X_{k_1} b_1 X_{k_2} \cdots X_{k_n} b_n \in \mathcal{B}\langle X_k : k \in K \rangle$ , the degree of  $P$  is  $n$  and its coefficients are  $b_0, \dots, b_n$ . The operator-valued distribution, or  $E$ -distribution, of a family  $\mathbf{A} = (A^{(k)})_{k \in K}$  of elements in  $\mathcal{A}$  is the map of operator-valued moments

$$E_{\mathbf{A}} : P \in \mathcal{B}\langle X_k : k \in K \rangle \mapsto E[P(\mathbf{A})] \in \mathcal{B}.$$

- Sub-families  $\mathbf{A}_1 = (A_1^{(k)})_{k \in K}, \dots, \mathbf{A}_L = (A_L^{(k)})_{k \in K}$  of  $\mathcal{A}$  are called free with amalgamation over  $\mathcal{B}$  (in short, free over  $\mathcal{B}$ ) whenever for any  $n \geq 2$ , any  $\ell_1 \neq \ell_2 \neq \cdots \neq \ell_n$  and any monomials  $P_1, \dots, P_n \in \mathcal{B}\langle X_k : k \in K \rangle$ , one has

$$E \left[ \left( P_1(\mathbf{A}_{\ell_1}) - E[P_1(\mathbf{A}_{\ell_1})] \right) \cdots \left( P_n(\mathbf{A}_{\ell_n}) - E[P_n(\mathbf{A}_{\ell_n})] \right) \right] = 0.$$

Ordinary freeness is the special case of operator-valued freeness when the algebra  $\mathcal{B}$  equals  $\mathbb{C}$ . Operator-valued freeness works mostly like the ordinary one, and the freeness with amalgamation of  $\mathbf{A}_1, \dots, \mathbf{A}_L$  allows to compute the operator-valued distribution of  $\mathbf{A}_1 \cup \dots \cup \mathbf{A}_L$  from the separate operator-valued distributions.

For  $N \in \mathbb{N}$ , the set of  $N \times N$  matrices is denoted by  $\mathcal{M}_N$ . We define the diagonal map  $\Delta$  on  $\mathcal{M}_N$  by  $\Delta A = (\delta_{i,j} A(i, j))_{i,j=1}^N$  for any  $A = (A(i, j))_{i,j=1}^N$ . The set of  $N \times N$  diagonal matrices is denoted by  $\mathcal{D}_N$ . Note that  $(\mathcal{M}_N, \mathcal{D}_N, \Delta)$  is an operator-valued non-commutative probability space.

For random matrices, freeness with amalgamation over the diagonal first appeared in the work of Shlyakhtenko [22] on Gaussian Wigner matrices with variance profile.

Our motivation to consider this question for permutation invariant matrices is a recent result by Boedihardjo and Dykema [7]. They proved that certain random Vandermonde matrices based on i.i.d. random variables are asymptotically  $\mathcal{R}$ -diagonal over the diagonal matrices. These random matrices are invariant by right multiplication by permutation matrices. If this symmetry was with respect to multiplication by any unitary matrices, then we would have obtained convergence to ordinary  $\mathcal{R}$ -diagonal elements [21, Theorem 15.10]. This then suggests a link between permutation invariance and freeness with amalgamation over the diagonal.

## 1.2 Asymptotic freeness with amalgamation over the diagonal

In the sequel, when we speak about a family  $\mathbf{A}_N$  of  $N \times N$  random matrices, we implicitly refer to a whole sequence  $(\mathbf{A}_N)_{N \in \mathbb{N}}$  of families of  $N \times N$  random matrices  $\mathbf{A}_N = (A_N^{(k)})_{k \in K}$ , where  $K$  is independent of  $N$ . The family  $\mathbf{A}_N$  is said to be *permutation invariant* if for any permutation  $\sigma$  of the set  $[N] := \{1, \dots, N\}$ , the following two families of random variables have the same distribution:

$$\left( A_N^{(k)}(i, j) \right)_{k \in K, 1 \leq i, j \leq N} \stackrel{\mathcal{L}aw}{=} \left( A_N^{(k)}(\sigma(i), \sigma(j)) \right)_{k \in K, 1 \leq i, j \leq N}.$$

Equivalently, for any permutation matrix  $S$  of size  $N$ ,

$$\left( A_N^{(k)} \right)_{k \in K} \stackrel{\mathcal{L}aw}{=} \left( S A_N^{(k)} S^{-1} \right)_{k \in K}.$$

The following result is a simplified version of our main result (Theorem 2.2).

**Theorem 1.2.** *Let  $\mathbf{A}_{N,1} = (A_{N,1}^{(k)})_{k \in K}, \dots, \mathbf{A}_{N,L} = (A_{N,L}^{(k)})_{k \in K}$  be independent families of random matrices which are uniformly bounded in operator norm. Assume moreover that each family is permutation invariant.*

*Then  $\mathbf{A}_{N,1}, \dots, \mathbf{A}_{N,L}$  are asymptotically free with amalgamation over  $\mathcal{D}_N$  in the following sense. For any  $n \geq 2$ , any  $\ell_1 \neq \ell_2 \neq \dots \neq \ell_n$  and any monomials  $P_{N,1}, \dots, P_{N,n} \in \mathcal{D}_N \langle X_k : k \in K \rangle$  such that their degrees and the norms of their coefficients are bounded uniformly in  $N$ , the matrix*

$$\varepsilon_N := \Delta \left[ (P_{N,1}(\mathbf{A}_{N,\ell_1}) - \Delta P_{N,1}(\mathbf{A}_{N,\ell_1})) \cdots (P_{N,n}(\mathbf{A}_{N,\ell_n}) - \Delta P_{N,n}(\mathbf{A}_{N,\ell_n})) \right]$$

*converges to zero in Schatten  $p$ -norm for all  $p \geq 1$ . Namely, for any  $p \geq 1$ , we have*

$$\mathbb{E} \left[ \frac{1}{N} \text{Tr}[(\varepsilon_N \varepsilon_N^*)^p] \right] \xrightarrow{N \rightarrow \infty} 0.$$

This convergence implies the convergence to zero in probability of the Schatten  $p$ -norms of  $\varepsilon_N$  as  $N$  tends to infinity. Thus, independent large permutation invariant matrices are asymptotically free over the diagonal in probability in  $(\mathcal{M}_N, \mathcal{D}_N, \Delta)$ : for a typical realization of  $\mathbf{A}_{N,1}, \dots, \mathbf{A}_{N,L}$ , the operator-valued distribution  $\mathbf{A}_{N,1} \cup \dots \cup \mathbf{A}_{N,L}$  is close to the distribution of copies of  $\mathbf{A}_{N,1}, \dots, \mathbf{A}_{N,L}$  which are free with amalgamation over  $\mathcal{D}_N$ .

In Theorem 2.2 we first strengthen this result for the operator-valued space of random matrices generated by the  $\mathbf{A}_{N,\ell}$ . In Section 3 we also weaken the assumption of invariance by permutations: our results also apply if each matrix of  $\mathbf{A}_\ell^{(N)}$  is multiplied entrywise by bounded random variables.

### 1.3 Numerical validation

Free harmonic analysis in operator-valued spaces and Theorem 2.2 introduce new perspectives for the analysis of large random matrices. In particular one can use the numerical method of Belinschi, Mai and Speicher [4] to calculate the limiting empirical spectral distribution of rational functions random matrices.

We illustrate our result by computing the limiting distribution of the sum of two independent Hermitian matrices of various models: GUE matrices with variance profile, adjacency matrices of Erdős-Rényi graphs, adjacency matrices of percolation on the cycle, diagonal matrices conjugated by a unitary Brownian motion or by the Fast Fourier Transform matrix.

The rest of the article is organized as follows. In Section 2, we state our main results: Theorem 2.2 and its generalizations Proposition 2.3 and Proposition 2.4. In Section 3, we present an algorithm to compute the eigenvalue distribution of the sum of independent permutation invariant random matrices, and apply it to various models. Finally, Section 4 contains the proofs of the different results of Section 2.

## 2 Statements of results

### 2.1 Freeness with amalgamation of large random matrices

We now consider the operator-valued probability space of random matrices over a probability space. We denote it  $(\mathcal{M}_N(L^\infty), \mathcal{D}_N(L^\infty), \Delta)$ , where  $\mathcal{M}_N(L^\infty)$  is the space of  $N \times N$  matrices of bounded random variables, and  $\mathcal{D}_N(L^\infty)$  is the space of  $N \times N$  diagonal matrices of bounded random variables. In this context,  $\mathcal{D}_N(L^\infty)\langle X_k : k \in K \rangle$  is the space of monomials of *random* diagonal matrices.

The minimal setting in order to formalize Theorem 1.2 in  $(\mathcal{M}_N(L^\infty), \mathcal{D}_N(L^\infty), \Delta)$  requires the following definition.

**Definition 2.1.** *Let  $\mathbf{A}_N$  be a family of random matrices. We define  $\mathcal{A}_N$  to be the smallest unital subalgebra containing the matrices of  $\mathbf{A}_N$  that is closed under  $\Delta$ , and we denote  $\mathcal{B}_N = \Delta(\mathcal{A}_N)$ .*

Remark that  $\mathcal{A}_N = \mathcal{B}_N\langle \mathbf{A}_N \rangle$ , and  $\mathcal{B}_N$  is the smallest subalgebra of  $\mathcal{D}_N(L^\infty)$  such that  $\Delta(\mathcal{B}_N\langle \mathbf{A}_N \rangle) \subset \mathcal{B}_N$ . The triplet  $(\mathcal{A}_N, \mathcal{B}_N, \Delta)$  is an operator-valued probability space. In order to formulate a notion of asymptotic freeness with amalgamation over  $\mathcal{B}_N$ , the coefficients of the polynomials should be consistent as the dimension  $N$  grows. The way we shall encode the coefficients independently of the dimension is by the following notion of graph monomial.

In this article, a *graph monomial*  $g$  is the data of a finite, bi-rooted, connected and directed graph  $G = (V, E, v_{in}, v_{out})$  and edge labels  $\ell : E \rightarrow \{1, \dots, L\}$  and  $k : E \rightarrow K$ . We allow multiplicity of edges and loops in our graphs. The roots  $v_{in}, v_{out}$ , are elements of the vertex set  $V$ , which we term the input and the output respectively. The roots are allowed to be equal. The labels  $\ell$  and  $k$  allows to assign a matrix  $A_{N, \ell(e)}^{(k(e))}$  to each edge  $e \in E$ .

The evaluation of the graph monomial  $g$  in the family of matrices  $\mathbf{A}_N = \mathbf{A}_{N,1} \cup \dots \cup \mathbf{A}_{N,L}$  is the  $N \times N$  matrix  $g(\mathbf{A}_N) \in \mathcal{M}_N(L^\infty)$  given by

$$g(\mathbf{A}_N)(i, j) := \sum_{\substack{\phi: V \rightarrow [N] \\ \phi(v_{in})=j, \phi(v_{out})=i}} \prod_{e=(v,w) \in E} A_{N, \ell(e)}^{(k(e))}(\phi(w), \phi(v)).$$

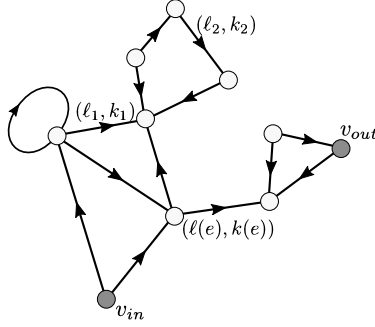


Figure 1: A graph monomial

In Lemma 4.3, we prove that  $\mathcal{B}_N$  is spanned by the matrices  $g(\mathbf{A}_N)$  where the graph of  $g$  is a so-called cactus-type monomial, i.e. a graph monomial whose underlying graph is a *cactus* and whose input and output are equal.

**Theorem 2.2.** *Let  $\mathbf{A}_{N,1} = (A_{N,1}^{(k)})_{k \in K}, \dots, \mathbf{A}_{N,L} = (A_{N,L}^{(k)})_{k \in K}$  be independent families of random matrices that are uniformly bounded in operator norm. Assume moreover that each family, except possibly one, is permutation invariant. We write  $\mathbf{A}_N = \mathbf{A}_{N,1} \cup \dots \cup \mathbf{A}_{N,L}$  for the union of the families of matrices.*

*Then  $\mathbf{A}_{N,1}, \dots, \mathbf{A}_{N,L}$  are asymptotically free with amalgamation in the operator-valued probability space  $(\mathcal{A}_N, \mathcal{B}_N, \Delta)$  in the following sense. Let  $n \geq 2$ ,  $\ell_1 \neq \ell_2 \neq \dots \neq \ell_n$  and  $P_{N,1}, \dots, P_{N,n} \in \mathcal{B}_N \langle X_k : k \in K \rangle$  which are explicitly given by*

$$P_{N,i} = g_{0,i}(\mathbf{A}_N) X_{k_i(1)} g_{1,i}(\mathbf{A}_N) \cdots X_{k_i(d_i)} g_{d_i,i}(\mathbf{A}_N),$$

where each  $g_{j,i}$  is a cactus-type monomial (which does not depend on  $N$ ). Then, the matrix

$$\varepsilon_N := \Delta \left[ (P_{N,1}(\mathbf{A}_{N,\ell_1}) - \Delta P_{N,1}(\mathbf{A}_{N,\ell_1})) \cdots (P_{N,n}(\mathbf{A}_{N,\ell_n}) - \Delta P_{N,n}(\mathbf{A}_{N,\ell_n})) \right]$$

converges to zero in Schatten  $p$ -norm for all  $p \geq 1$ . Namely, for any  $p \geq 1$ , we have

$$\mathbb{E} \left[ \frac{1}{N} \text{Tr} [(\varepsilon_N \varepsilon_N^*)^p] \right] \xrightarrow{N \rightarrow \infty} 0.$$

Note that there is no assumption of convergence of the  $\Delta$ -distribution of the matrices, as it is commonly assumed in the context of *deterministic equivalents* [20, Chapter 10].

## 2.2 Generalizations

In this section, we present some generalizations of Theorem 2.2. First, we relax the bounded operator norm assumption.

**Proposition 2.3.** *The conclusion of Theorem 2.2 remains valid if the operator norm bound assumption is replaced by the following weaker assumption: for any  $\ell \in \{1, \dots, L\}$ , any graph monomials  $g_1, \dots, g_n$  with equal input and output,*

$$\mathbb{E} \left[ \prod_{i=1}^n \text{Tr} (g_i(\mathbf{A}_{N,\ell})) \right] = O(N^{\sum_{i=1}^n \mathfrak{f}(g_i)/2}), \quad (2.1)$$

where  $\mathfrak{f}(g_i) \geq 2$  is determined from the forest  $F$  of two-edge connected components of the graph underlying  $g_i$  (Definition 4.8).

In particular, the conclusion of Proposition 2.3 holds if

$$\mathbb{E} \left[ \prod_{i=1}^n \text{Tr}(g_i(\mathbf{A}_{N,\ell})) \right] = O(N^n),$$

which is the boundedness of the so-called traffic distribution [17]. For example, this holds for Wigner matrices with exploding moments as the matrix of the Erdős-Rényi graph (see for instance [18, Proposition 4.1]), which do not satisfy the bounded operator norm property [27, Proposition 12].

The integer  $\mathfrak{f}(g)$  appearing in the above theorem has been considered by Mingo and Speicher in [19], where they proved that (2.1) is true if  $\mathbf{A}_{N,1}, \dots, \mathbf{A}_{N,L}$  are uniformly bounded in operator norms (see Section 4.4).

**Proposition 2.4.** *Let  $\mathbf{A}_{N,1}, \dots, \mathbf{A}_{N,L}$  be independent families of random matrices that are uniformly bounded in operator norms, or which satisfies the bound (2.1). Assume moreover that each family is permutation invariant. Let  $(\Gamma_\ell^{(k)})_{\ell,k}$  be a family of random matrices with uniformly bounded entries, independent of  $(\mathbf{A}_{N,1}, \dots, \mathbf{A}_{N,L})$ . Then the conclusion of Theorem 1.2 and the conclusion of Theorem 2.2 remains true for the family  $\tilde{\mathbf{A}}_{N,1}, \dots, \tilde{\mathbf{A}}_{N,L}$ , where*

$$\tilde{\mathbf{A}}_{N,\ell} = \left( A_{N,\ell}^{(k)} \circ \Gamma_\ell^{(k)} \right)_{k \in K},$$

and  $\circ$  denotes the entrywise product.

Let us emphasize that the families of matrices  $(\Gamma_1^{(k)})_k, \dots, (\Gamma_L^{(k)})_k$  are not assumed to be neither independent nor permutation invariant, and so too for the  $\tilde{\mathbf{A}}_{N,1}, \dots, \tilde{\mathbf{A}}_{N,L}$ . In particular, Proposition 2.4 can be applied to Wigner random matrices with variance profile [22].

### 3 Examples and numerical validation

In this section, we consider various models of independent permutation invariant random matrices  $X_N$  and  $Y_N$ . On the one hand, we construct the empirical spectral distribution of  $X_N + Y_N$ . On the other hand, we use the fixed point algorithm of Belinschi, Mai and Speicher [4] to compute the spectral distribution of the free convolution with amalgamation over the diagonal of  $X_N$  and  $Y_N$ . This depends only on the separated  $\Delta$ -distributions of  $X_N$  and  $Y_N$ . Our main theorem guarantees that the difference between this two pictures becomes negligible in the limit in expectation and in probability. We actually obtain the numerical evidence for realizations of large matrices that asymptotic freeness over the diagonal holds.

#### 3.1 Amalgamated subordination property

We need the following notion to explain the fixed point algorithm. Let  $(\mathcal{A}, \mathcal{D}, \Delta)$  be an operator-valued probability space such that  $\mathcal{A}$  is a  $C^*$ -algebra. The operator-valued Stieltjes transform of a self-adjoint element  $X$  of  $\mathcal{A}$  is the map

$$G_X : \Lambda \in \mathcal{D}^+ \mapsto \Delta[(\Lambda - X)^{-1}],$$

where  $\mathcal{D}^+$  is the set of elements  $\Lambda \in \mathcal{D}$  such that  $\Im(\Lambda) := \frac{\Lambda - \Lambda^*}{2i} > 0$ . We also write  $H_X$  for the map  $H : \Lambda \mapsto G_X(\Lambda)^{-1} - \Lambda$ .

For random matrices, this map is then the diagonal of the resolvent of the matrices. Let us remark that this object was known to be an important tool in the analysis of large random matrices, for instance for random matrices with heavy tailed entries [9, 5, 8, 3], for adjacency matrices of random graphs [15], or more recently for the local analysis of Wigner matrices with variance profile [1].

**Theorem 3.1** (Theorem 2.2 of [4]). *Two self-adjoint operator-valued random variables  $X$  and  $Y$  free with amalgamation over  $\mathcal{D}$  satisfy the following subordination property. For all  $\Lambda \in \mathcal{D}^+$ , we have  $G_{X+Y}(\Lambda) = G_X(\Omega(\Lambda))$ , where the subordination function  $\Omega : \mathcal{D}^+ \rightarrow \mathcal{D}^+$  is the unique solution of the fixed point equation  $\Omega(\Lambda) = F_\Lambda(\Omega(\Lambda))$ , where*

$$F_\Lambda(\Omega) = H_Y(H_X(\Omega) + \Lambda) + \Lambda.$$

Moreover, it is the limit of the sequence  $\Omega_n$  given by  $\Omega_{n+1} = F_\Lambda(\Omega_n)$  for any  $\Omega_0 \in \mathcal{D}^+$ .

Let  $X$  and  $Y$  be two self-adjoint operator valued random variables, free with amalgamation. Assume that the  $\Delta$ -distributions of  $X$  and  $Y$  are those of our matrices  $X_N$  and  $Y_N$  respectively. The following algorithm generates an approximation of the value of the spectral density  $g_{X+Y}(x)$  of  $X + Y$  at  $x \in \mathbb{R}$  when it exists.

**Step 1** Simulate a realization of  $X_N$  and a realization of  $Y_N$ .

**Step 2** Set  $\Lambda = (x + iy)\mathbb{1}_N$  where  $y > 0$  is small. Compute the terms of the sequence  $\Omega_1(\Lambda) := \Lambda$ ,  $\Omega_{n+1}(\Lambda) := H_{Y_N}(H_{X_N}(\Omega_n(\Lambda)) + \Lambda) + \Lambda$ ,  $\forall n \geq 1$ , until the difference between  $\Omega_n$  and  $\Omega_{n+1}$  is under a threshold.

**Step 3** The value of the density  $g_{X+Y}(x)$  is then close to

$$\frac{1}{\pi} \Im \left( \frac{1}{N} \text{Tr} [(\Omega_n(\Lambda) - X_N)^{-1}] \right)$$

provided  $y$  is small enough and the distribution admits a density at  $x$ .

*Remark 3.2.* Let us mention two strengthening of this method.

1. The amalgamated  $\mathcal{R}$ -transform of  $Y$  is the map  $\mathcal{D}^+ \rightarrow \mathcal{D}^+$  characterized by  $G_Y(\Lambda) = (\Lambda - R_Y(G_Y(\Lambda)))^{-1}$ . The function  $\Omega(\Lambda)$  of Theorem 3.1 actually equals  $\Lambda - R_Y(G_{X+Y}(\Lambda))$ . The knowledge of the amalgamated  $\mathcal{R}$ -function of  $X$  and/or  $Y$  provides faster algorithms [20, Theorem 11 of Chapter 9] which do not require the simulation of the matrices  $X$  and/or  $Y$ .
2. Thanks to a linearization trick, the fixed point algorithm described by Belinschi, Mai and Speicher can be extended in order to compute the distribution of any non-commuting rational function in  $X$  and  $Y$ , see [4, Theorem 2.2].

### 3.2 Matrix models

In the rest of the section, we will present examples of different models to illustrate of our results.

1. We write  $GUEvp(\eta)$  for the matrix with variance profile decomposed in blocks

$$GUEvp(\eta) = \sqrt{\frac{8}{5\eta + 3}} \begin{pmatrix} \sqrt{\eta}X_{11} & X_{12} \\ X_{21} & \sqrt{\eta}X_{22} \end{pmatrix},$$



where  $\eta > 0$  is a parameter,  $X_{11}$  has size  $N/4 \times N/4$  and  $\begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix}$  is a GUE matrix.

2. We write  $ER(d)$  for the standardized adjacency matrix of a sparse Erdős-Rényi graph, namely

$$ER(d) = \frac{Y_N - d\mathbb{J}_N}{\sqrt{d(1 - d/(N-1))}},$$

where

- $Y_N$  is a real symmetric random matrix, with null diagonal, and such that the strict subdiagonal entries are i.i.d. Bernoulli random variables with parameter  $d/(N-1)$ ,
  - $\mathbb{J}_N$  is the matrix whose non diagonal entries are all  $1/(N-1)$  and diagonal entries are zero,
  - $d > 0$  is a parameter.
3. We write  $Perm$  for the matrix  $\frac{1}{\sqrt{2}}(V + V' - 2\mathbb{J}_N)$ , where  $V$  is a uniform permutation matrix, and  $Perm(p)$  for the matrix  $\frac{1}{\sqrt{p}}Perm \circ \Gamma(p)$ , where  $\Gamma(p)$  is a symmetric matrix, with null diagonal, and whose strict subdiagonal entries are i.i.d. Bernoulli random variables with parameter  $p$ .
  4. For a given diagonal matrix  $D$ , we write  $FFT_D$  for the matrix  $VUDU^*V^*$  where  $V$  is a uniform permutation matrix and  $U$  is the *Fast Fourier Transform* unitary matrix  $(\frac{1}{\sqrt{N}}\omega e^{-i2\pi\frac{(n-1)(m-1)}{N}})_{nm}$ . We also denote  $FFT_D(p)$  for the matrix  $\frac{1}{\sqrt{p}}FFT_D \circ \Gamma(p)$ , where  $\Gamma(p)$  is as in the definition of  $Perm(p)$ .
  5. For a given diagonal matrix  $D$ , we write  $UMB_D(t)$  for the matrix  $U_{t,N}DU_{t,N}^*$ , where  $U_{t,N}$  is a unitary Brownian motion starting from a uniform permutation matrix at time  $t$ . We refer to [6] for the definition of the unitary Brownian motion and the related notion of  $t$ -freeness.
  6. We also define the three following diagonal matrix:

$$D1 := \text{diag}(1, -1, 1, -1, \dots, 1, -1),$$

$$D2 := \text{diag}(\underbrace{-1, \dots, -1}_{N/2}, \underbrace{1, \dots, 1}_{N/2}),$$

$$D3 := \sqrt{\frac{3}{14}} \text{diag}\left(-2, -2 + \frac{2}{N}, -2 + \frac{4}{N}, \dots, -2 + \frac{N-1}{N}, 1 + \frac{2}{N}, 1 + \frac{4}{N}, \dots, 2\right).$$

### 3.3 Experiment framework and comments

Figure 2 reports the result of 8 numerical simulations, for different models for  $X_N$  and  $Y_N$  indicated in the legend of the pictures.

In each case, we represent

- the histogram of the eigenvalues of one realization of  $\frac{1}{\sqrt{2}}(X_N + Y_N)$  for matrices of size  $N = 1000$ ,
- in blue, the density of the spectral distribution of  $\frac{1}{\sqrt{2}}(X + Y)$ , where  $X$  and  $Y$  are free over the diagonal and such that their  $\Delta$ -distribution is the one of  $X_N$  and  $Y_N$

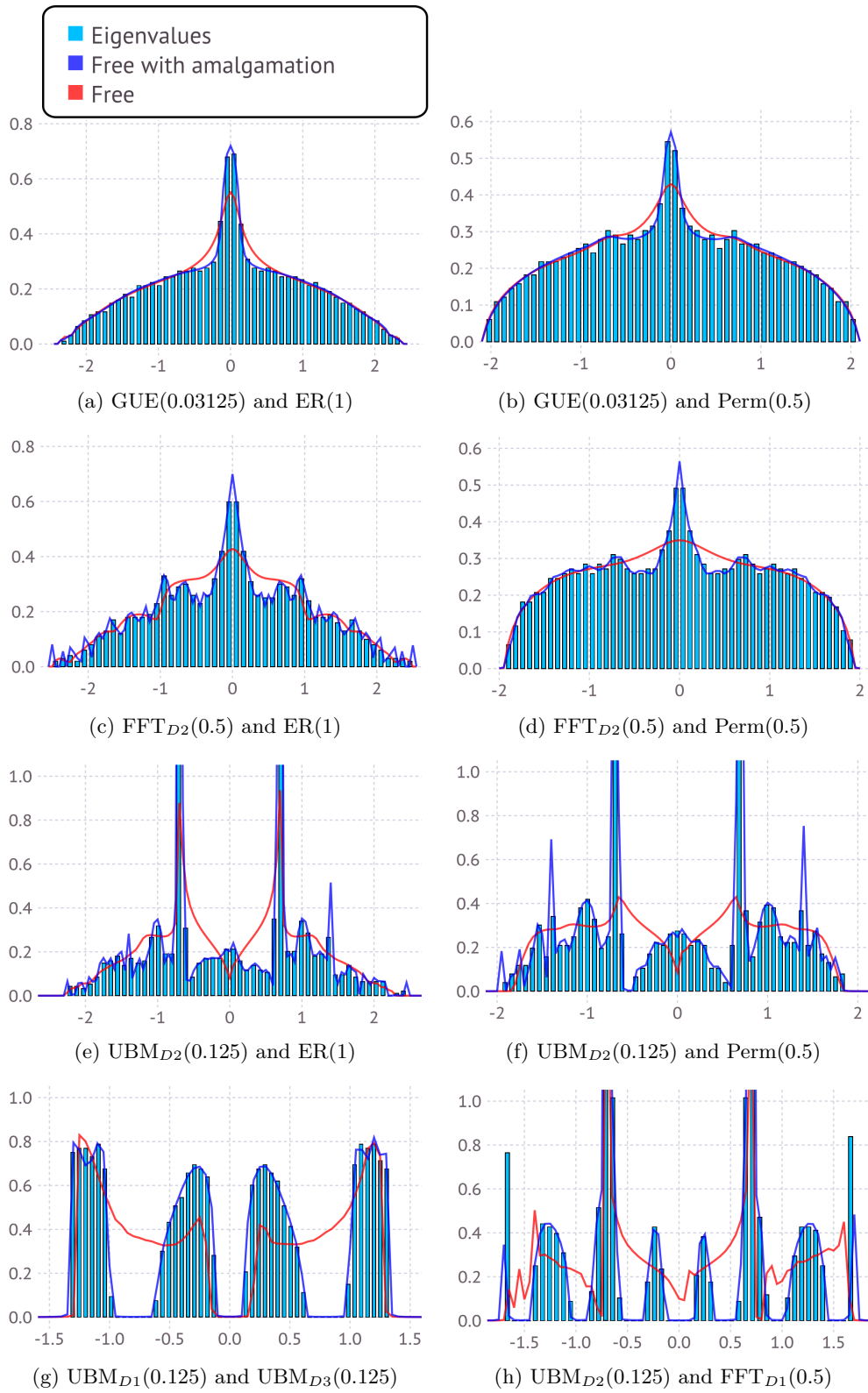


Figure 2: Results of the numerical simulations.

- in red, the density of the free convolution of the empirical eigenvalues distributions of  $X_N$  and  $Y_N$ .

In the algorithm, the small parameter  $y$  equals 0.001. We stop the fixed point process at the threshold  $\|\Omega_n - \Omega_{n+1}\|_\infty \leq 0.001$ .

The prediction by asymptotic freeness over the diagonal fits accurately the histogram in all situations we have investigated numerically. In Picture (a) to (d) we see that the deviation between the free case and the free with amalgamation case is mainly at the center of the spectrum. In Picture (e) to (h), the time of the Brownian motion is quite small and so the difference between the red and the blue line is more important.

## 4 Proof of the theorems

### 4.1 Algebras of graph polynomials

We first make precise the notion of cacti mentioned previously.

**Definition 4.1.** *A connected graph is a cactus if every edge belongs to exactly one simple cycle. An oriented cactus is a cactus such that each simple cycle is directed.*

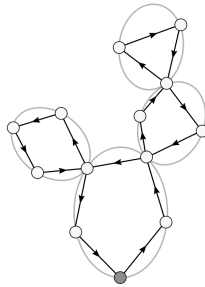


Figure 3: Cactus with four simple cycles

We define  $\mathcal{C}$  to be the set of *cactus-type monomials*, that is such that the underlying graph is an oriented cactus and the input and output are equal. We denote by  $\mathcal{C}_N$  the vector space generated by  $(g(\mathbf{A}_N))_{g \in \mathcal{C}}$ . More generally we define  $\mathcal{F}$  to be the set of graph monomials obtained by considering a directed simple path oriented from the right to the left and by attaching oriented cacti on the vertices of the path. The input is the far right vertex of the path, the output is the far left. We then denote by  $\mathcal{F}_N$  the vector space generated by  $(g(\mathbf{A}_N))_{g \in \mathcal{F}}$ . It is clear that  $\mathcal{C} \subset \mathcal{F}$  and so  $\mathcal{C}_N \subset \mathcal{F}_N$ . Moreover, note that  $\mathcal{F}_N = \mathcal{C}_N \langle \mathbf{A}_N \rangle$ , namely it is the vector space generated by elements of the form

$$P = g_0(\mathbf{A}_N) A_{N, \ell_1}^{(k_1)} g_1(\mathbf{A}_N) \cdots A_{N, \ell_d}^{(k_d)} g_d(\mathbf{A}_N), \quad (4.1)$$

where, for  $i \in \{0, \dots, d\}$ ,  $g_i$  is in  $\mathcal{C}$ .

**Lemma 4.2.** *The spaces  $\mathcal{C}_N$  and  $\mathcal{F}_N$  are unital algebras and  $\Delta(\mathcal{F}_N) = \mathcal{C}_N$ .*

*Proof.* For any graph monomials  $g_1$  and  $g_2$ , the product of matrices  $g_1(\mathbf{A}_N)g_2(\mathbf{A}_N)$  is equal to  $g(\mathbf{A}_N)$ , where  $g$  is obtained by identifying the input of  $g_1$  with the output of  $g_2$ . It follows that  $\mathcal{C}_N$  is an algebra, and so is  $\mathcal{F}_N$ . Both are unital by considering the trivial graph with a single isolated vertex.

For any graph monomial  $g$ , the matrix  $\Delta(g(\mathbf{A}_N))$  is equal to  $\Delta(g)(\mathbf{A}_N)$ , where  $\Delta(g)$  is obtained by identifying the input with the output of  $g$ . It follows that  $\Delta(\mathcal{F}_N) = \mathcal{C}_N$ .  $\square$

**Lemma 4.3.** *One has  $\mathcal{F}_N = \mathcal{A}_N$  and consequently  $\mathcal{C}_N = \mathcal{B}_N$ .*

*Proof.* Recall that  $\mathcal{A}_N$  is the smallest unital algebra containing  $\mathbf{A}_N$  stable under  $\Delta$ . It is clear that  $\mathcal{A}_N \subset \mathcal{F}_N$  since  $\mathbf{A}_N \subset \mathcal{F}_N$  and  $\Delta(\mathcal{F}_N) \subset \mathcal{F}_N$  by Lemma 4.2.

We now prove the reverse inclusion by induction on the number of edges. If  $g$  has a single edge then  $g(\mathbf{A}_N) \in \mathbf{A}_N \cup \Delta(\mathbf{A}_N) \subset \mathcal{A}_N$ .

Assume that for some  $n \geq 2$ , every  $g(\mathbf{A}_N) \in \mathcal{F}_N$  whose graph has fewer than  $n$  edges belongs to  $\mathcal{A}_N$ . Let  $g(\mathbf{A}_N) \in \mathcal{F}_N$  for some graph with  $n$  edges. If the input and the output of  $g$  are not equal, or if they are equal but belong to more than one cycle, then we can write  $g(\mathbf{A}_N) = g_1(\mathbf{A}_N)g_2(\mathbf{A}_N)$ , where  $g_1$  and  $g_2$  have fewer than  $n$  edges. Since  $\mathcal{A}_N$  is an algebra, then  $g(\mathbf{A}_N) \in \mathcal{A}_N$ .

Assume that the input and the output are equal to a vertex  $v$  and belong to exactly one cycle. Then one can split the vertex  $v$  into two vertices  $v_{in}$  and  $v_{out}$ , allowing one to exhibit a product  $P$  of the form (4.1) such that  $\Delta(P) = g(\mathbf{A}_N)$ . Now we can factorize  $P = g_1(\mathbf{A}_N)g_2(\mathbf{A}_N)$  as in the previous case, where  $g_1(\mathbf{A}_N)$  and  $g_2(\mathbf{A}_N)$  belong to  $\mathcal{A}_N$  by the induction hypothesis. Hence  $g(\mathbf{A}_N) = \Delta(g_1(\mathbf{A}_N)g_2(\mathbf{A}_N))$  belongs to  $\mathcal{A}_N$ .  $\square$

## 4.2 Preliminary lemmas

We start by considering arbitrary families  $\mathbf{A}_{N,1}, \dots, \mathbf{A}_{N,L}$  of random matrices. By enlarging them if necessary, we can assume that the families are closed under  $*$ .

**Lemma 4.4.** *If*

$$\mathbb{E} \left[ \frac{1}{N} \text{Tr} \varepsilon_N \right] \xrightarrow{N \rightarrow \infty} 0$$

for every  $\varepsilon_N$  as in Theorem 2.2, i.e.

$$\varepsilon_N := (P_{N,1}(\mathbf{A}_{N,\ell_1}) - \Delta P_{N,1}(\mathbf{A}_{N,\ell_1})) \cdots (P_{N,n}(\mathbf{A}_{N,\ell_n}) - \Delta P_{N,n}(\mathbf{A}_{N,\ell_n}))$$

where  $n \geq 2$ ,  $\ell_1 \neq \ell_2 \neq \cdots \neq \ell_n$  and  $P_{N,1}, \dots, P_{N,n} \in \mathcal{B}_N \langle X_k : k \in K \rangle$  are explicitly given by  $P_{N,i} = g_{0,i}(\mathbf{A}_N)X_{k_i(1)}g_{1,i}(\mathbf{A}_N) \cdots X_{k_i(d_i)}g_{d_i,i}(\mathbf{A}_N)$ , with  $g_{j,i}$  cactus-type monomials (which does not depend on  $N$ ), then, we also have

$$\mathbb{E} \left[ \frac{1}{N} \text{Tr}(\varepsilon_N \varepsilon_N^*)^p \right] \xrightarrow{N \rightarrow \infty} 0$$

for any such  $\varepsilon_N$ .

*Proof.* Let us remark that  $(\varepsilon_N \varepsilon_N^*)^p = \varepsilon_N \varepsilon_N^* (\varepsilon_N \varepsilon_N^*)^{p-1}$  can be written as:

$$(P_{N,1}(\mathbf{A}_{N,\ell_1}) - \Delta P_{N,1}(\mathbf{A}_{N,\ell_1})) \cdots (P_{N,n}(\mathbf{A}_{N,\ell_n}) - \Delta P_{N,n}(\mathbf{A}_{N,\ell_n})) \varepsilon_N^* (\varepsilon_N \varepsilon_N^*)^{p-1}.$$

Moreover,  $\varepsilon_N^* (\varepsilon_N \varepsilon_N^*)^{p-1}$  is an element of  $\mathcal{B}_N$ , that we denote  $g(\mathbf{A}_N)$ . Since  $\Delta$  is a conditional expectation, one has  $\Delta(P_{N,n}(\mathbf{A}_{N,\ell_n}))g(\mathbf{A}_N) = \Delta(P_{N,n}(\mathbf{A}_{N,\ell_n})g(\mathbf{A}_N))$ . Hence one can update the last coefficient of the polynomial  $P_{N,n}$  in order to include the factor  $\varepsilon_N^* (\varepsilon_N \varepsilon_N^*)^{p-1}$ . More precisely, writing

$$P_{N,n} = g_{0,n}(\mathbf{A}_N)X_{k_n(1)}g_{1,n}(\mathbf{A}_N) \cdots X_{k_n(d_n)}g_{d_n,n}(\mathbf{A}_N),$$

we can replace this polynomial by

$$\tilde{P}_{N,n} = g_{0,n}(\mathbf{A}_N) X_{k_n(1)} g_{1,n}(\mathbf{A}_N) \cdots X_{k_n(d_n)} \tilde{g}_{d_n,n}(\mathbf{A}_N),$$

where  $\tilde{g}_{d_n,n}(\mathbf{A}_N) = g_{d_n,n}(\mathbf{A}_N)g(\mathbf{A}_N)$ . Hence we are left to prove that  $\mathbb{E}\left[\frac{1}{N}\text{Tr}\varepsilon_N\right]$  converges to zero.  $\square$

For any  $i \in \{1, \dots, n\}$ , let  $P_{N,i} \in \mathcal{B}_N\langle X_k : k \in K \rangle$ . We define  $M_i := P_{N,i}(\mathbf{A}_{N,\ell_i})$  and  $\overset{\circ}{M}_i := M_i - \Delta(M_i)$ . In order to compute the quantity  $\mathbb{E}\left[\frac{1}{N}\text{Tr}(\overset{\circ}{M}_1 \dots \overset{\circ}{M}_n)\right]$ , we use the moment method, involving sums over unrooted labeled graphs that we introduce now.

In this article, a *test graph*  $T$  is the data of a finite, directed and labeled graph  $T = (V, E, K)$ . Note that the graphs are possibly disconnected. The labelling map  $K : E \rightarrow \mathbf{A}_N$  corresponds to an assignment of matrices  $e \mapsto K_e$ . We define the quantities

$$\tau_N[T] = \mathbb{E}\left[\frac{1}{N^{c(T)}} \sum_{\phi: V \rightarrow [N]} \prod_{e=(v,w) \in E} K_e(\phi(w), \phi(v))\right] \quad (4.2)$$

$$\tau_N^0[T] = \mathbb{E}\left[\frac{1}{N^{c(T)}} \sum_{\substack{\phi: V \rightarrow [N] \\ \text{injective}}} \prod_{e=(v,w) \in E} K_e(\phi(w), \phi(v))\right], \quad (4.3)$$

where  $c(T)$  is the number of connected components of  $T$ .

For any partition  $\pi$  of  $V$ , we define  $T^\pi = (V^\pi, E^\pi, K^\pi)$  as the graph obtained from  $T$  by identifying vertices in the same block of  $\pi$ . More precisely, the vertices of  $T^\pi$  are the blocks of  $\pi$ , each edge  $e = (v, w)$  induces an edge  $e^\pi = (B_v, B_w)$  where  $B_v$  and  $B_w$  are the blocks of  $\pi$  containing  $v$  and  $w$  respectively. The label  $K_{e^\pi}^\pi$  is the same as the label  $K_e$ . We say that  $T^\pi$  is a quotient of  $T$ . We then have the two relations [17, Lemma 2.6]

$$\tau_N[T] = \sum_{\pi \in \mathcal{P}(V)} N^{c(T^\pi) - c(T)} \tau_N^0[T^\pi] \quad (4.4)$$

$$\tau_N^0[T] = \sum_{\pi \in \mathcal{P}(V)} N^{c(T^\pi) - c(T)} \text{Mob}(0, \pi) \tau_N[T^\pi], \quad (4.5)$$

where  $\text{Mob}(0, \pi) = \left(\prod_{B \in \pi} (-1)^{|B|} (|B| - 1)!\right)$  is the Möbius function on the poset of partitions [23, Example 3.10.4].

We write  $M_i = g_{0,i}(\mathbf{A}_N) A_{k_i(1)} g_{1,i}(\mathbf{A}_N) \cdots A_{k_i(d_i)} g_{d_i,i}(\mathbf{A}_N)$ , and consider the graph monomial  $t_i \in \mathcal{F}$  such that  $t_i(\mathbf{A}_N) = M_i$ . It is explicitly given by

$$t_i := \left( \begin{array}{c} \overset{g_{0,i}}{\underset{\vee}{\cdot}} \quad (\ell_i, k_i(1)) \quad \overset{g_{1,i}}{\underset{\vee}{\cdot}} \quad (\ell_i, k_i(2)) \quad \overset{g_{2,i}}{\underset{\vee}{\cdot}} \quad \dots \quad \overset{g_{d_i-1,i}}{\underset{\vee}{\cdot}} \quad (\ell_i, k_i(d_i)) \quad \overset{g_{d_i,i}}{\underset{\vee}{\cdot}} \end{array} \right),$$

obtained by considering a directed simple path oriented from the right to the left whose edges are labelled by  $(\ell_i, k_i(d_i)), \dots, (\ell_i, k_i(1))$  and attaching, following the decreasing order on  $j = d_i, \dots, 0$ , the root of the graph monomial  $g_{j,i}$  on the vertices of the path.

Let  $T = (V, E, K)$  be the test graph obtained by identifying the output vertex of  $t_i$  with the input vertex of  $t_{i-1}$  for  $i \in \{1, \dots, n\}$  (with notation modulo  $n$  for the index  $i$ ). It inherits the edge labels from the  $t_i$ . Note that as an unrooted graph,  $T$  equals  $\Delta(t_1 \cdots t_n)$ . One can easily verify that  $\mathbb{E}\left[\frac{1}{N}\text{Tr}[M_1 \dots M_n]\right] = \tau_N[T]$ .

For each  $i = 1, \dots, n$ , the graph  $t_i$  can be seen as a subgraph of  $T$ . We denote by  $v_i$  the vertex of  $T$  corresponding to the input of  $t_i$  with the convention  $v_0 = v_n$ . The following lemma describes the cancellations in Formula (4.4) when one replaces  $M_i$  by  $\overset{\circ}{M}_i = M_i - \Delta(M_i)$ .

**Lemma 4.5.** *With above notations for the test graph  $T$  we have*

$$\mathbb{E} \left[ \frac{1}{N} \text{Tr}[\overset{\circ}{M}_1 \dots \overset{\circ}{M}_n] \right] = \sum_{\substack{\pi \in \mathcal{P}(V) \text{ s.t.} \\ v_i \not\sim_{\pi} v_{i-1}, \forall i}} \tau_N^0[T^\pi].$$

*Proof.* Let us consider  $I \subset \{1, \dots, n\}$ . Denote by  $T_I$  the test graph obtained from  $T$  by identifying, for each  $i \in I$ , the input and output of  $t_i$ . Then we have

$$\mathbb{E} \left[ \frac{1}{N} \text{Tr}[\overset{\circ}{M}_1 \dots \overset{\circ}{M}_n] \right] = \sum_{I \subset \{1, \dots, n\}} (-1)^{|I|} \tau_N[T_I].$$

For each  $i = 1, \dots, n$ , we denote by  $V_i$  the vertex set of  $t_i$ , seen in the graph  $T$ .

$$\mathbb{E} \left[ \frac{1}{N} \text{Tr}[\overset{\circ}{M}_1 \dots \overset{\circ}{M}_n] \right] = \sum_{I \subset \{1, \dots, n\}} (-1)^{|I|} \sum_{\substack{\pi \in \mathcal{P}(V) \\ \text{s.t. } v_{i-1} \sim_{\pi} v_i, \forall i \in I}} \tau_N^0[T^\pi]$$

For any  $\pi \in \mathcal{P}(V)$ , denote by  $J_\pi$  the set of indices  $i \in \{1, \dots, n\}$  such that  $v_{i-1} \sim_{\pi} v_i$ . Then, exchanging the order of the sums, we get

$$\mathbb{E} \left[ \frac{1}{N} \text{Tr}[\overset{\circ}{M}_1 \dots \overset{\circ}{M}_n] \right] = \sum_{\pi \in \mathcal{P}(V)} \left( \sum_{I \subset J_\pi} (-1)^{|I|} \right) \tau_N^0[T^\pi].$$

Since the sum in the parentheses vanishes as soon as the index set  $J_\pi$  is nonempty, we get the expected result.  $\square$

Our analysis of each term  $\tau_N^0[T^\pi]$  relies on the geometry of a graph we introduce now. In this article, we call colored component of  $T^\pi$  a connected maximal subgraph of  $T^\pi$  whose edges are labelled in one of the sets  $\{1\} \times K, \dots, \{L\} \times K$ . We denote by  $\mathcal{CC}(T^\pi)$  the set of colored components. Hence a colored component has labels corresponding to only one of the families  $\mathbf{A}_{N,1}, \dots, \mathbf{A}_{N,L}$ . We call graph of colored components of  $T^\pi$  the undirected bipartite graph  $\mathcal{GCC}(T^\pi) = (\mathcal{V}_\pi, \mathcal{E}_\pi)$ , where  $\mathcal{V}_\pi = \mathcal{CC}(T^\pi) \sqcup V_\pi$  and there is an edge between each vertex and the colored components it belongs to. The definition is slightly different from the one in [17] where  $\mathcal{V}_\pi$  is the union of the colored components and the vertices of  $T^\pi$  that belong to more than one colored component.

**Lemma 4.6.** *There is no partition  $\pi$  in the sum of Lemma 4.5 such that  $\mathcal{GCC}(T^\pi)$  is a tree.*

*Proof.* Consider the simple cycle  $c$  of  $T$  that visits the edges of  $t_1, \dots, t_n$  labeled with  $b$  in  $\bigsqcup_{\ell} \mathcal{A}_\ell$ . For any partition  $\pi$  of  $\mathcal{P}(V)$ , the cycle induces a cycle  $c^\pi$  (non simple in general) on the graph of colored components of  $T^\pi$ . If this graph is a tree, then eventually  $c^\pi$  backtracks and  $\pi$  must identify  $v_i$  with  $v_{i-1}$  for some  $i \in \{1, \dots, n\}$ .  $\square$

The conclusion so far is that to prove the converge to zero of any  $\varepsilon_N$  as in Lemma 4.4, it suffices to prove the convergence to zero of  $\tau_N^0[T^\pi]$ , as in Lemma 4.5, for any  $T^\pi$  such that  $\mathcal{GCC}(T^\pi)$  is not a tree.

### 4.3 Proof of Proposition 2.3

Henceforth, we assume that the families of matrices  $\mathbf{A}_{N,2}, \dots, \mathbf{A}_{N,L}$  are permutation invariant and we are operating under the assumptions of Proposition 2.3. Notations are those of the previous section.

We first prove that, without loss of generality, we can assume that  $\mathbf{A}_{N,1}$  is also permutation invariant. By Lemma 4.2, the quantity  $\mathbb{E}[\frac{1}{N}\text{Tr}\varepsilon_N]$  is a linear combination of terms of the form  $\mathbb{E}[\frac{1}{N}\text{Tr}g(\mathbf{A}_N)]$ , for some graph monomials  $g$ . Let  $U$  be a uniform permutation matrix, independent of  $\mathbf{A}_N$ . By [17, Lemma 1.4], for any graph monomial  $g$  one has  $\mathbb{E}[\frac{1}{N}\text{Tr}g(\mathbf{A}_N)] = \mathbb{E}[\frac{1}{N}\text{Tr}g(U\mathbf{A}_N U^t)]$ . Since the families of matrices are independent, we have equality in distribution

$$(U\mathbf{A}_{N,1}U^t, \dots, U\mathbf{A}_{N,L}U^t) \stackrel{Law}{=} (U\mathbf{A}_{N,1}U^t, \mathbf{A}_{N,2}, \dots, \mathbf{A}_{N,L}).$$

This proves that we can replace  $\mathbf{A}_{N,1}$  by the permutation invariant family  $U\mathbf{A}_{N,1}U^t$ .

**Lemma 4.7.** *For any partition  $\pi$  of  $V$  such that  $\mathcal{GCC}(T^\pi)$  is not a tree, we have  $\tau_N^0[T^\pi] = O(N^{-1})$ .*

*Proof.* Let  $\phi$  be an arbitrary injective function from  $V^\pi$  to  $[N]$ . By permutation invariance, we have [17, Lemma 4.1]

$$\tau_N^0[T^\pi] = \frac{1}{N} \frac{N!}{(N - |V^\pi|)!} \mathbb{E} \left[ \prod_{e=(v,w) \in E^\pi} K_e^\pi(\phi(w), \phi(v)) \right]. \quad (4.6)$$

For any  $1 \leq \ell \leq L$ , we denote by  $T_\ell^\pi = (V_\ell, E_\ell)$  the test graph which is composed of the colored components of  $T^\pi$  labelled by the matrices  $\mathbf{A}_{N,\ell}$ . Then, by independence of the families of matrices:

$$\mathbb{E} \left[ \prod_{e=(v,w) \in E^\pi} K_e^\pi(\phi(w), \phi(v)) \right] = \prod_{\ell=1}^L \mathbb{E} \left[ \prod_{e=(v,w) \in E_\ell} K_e^\pi(\phi(w), \phi(v)) \right].$$

Hence, using again (4.6) for each graph  $T_\ell^\pi$  we get

$$\begin{aligned} \tau_N^0[T^\pi] &= \frac{1}{N} \frac{N!}{(N - |V^\pi|)!} \left( \prod_{\ell=1}^L \frac{(N - |V_\ell|)!}{N!} \right) N^{|\mathcal{CC}(T^\pi)|} \times \prod_{\ell=1}^L \tau_N^0[T_\ell^\pi] \\ &= N^{-1+|V^\pi|-\sum_\ell |V_\ell|+|\mathcal{CC}(T^\pi)|} \left( 1 + O\left(\frac{1}{N}\right) \right) \times \prod_{\ell=1}^L \tau_N^0[T_\ell^\pi], \end{aligned}$$

where  $|\mathcal{CC}(T^\pi)|$  is the number of colored components of  $T^\pi$ . Note that the cardinals of the vertex and edge sets of  $\mathcal{GCC}(T^\pi)$  are  $|\mathcal{V}_\pi| = |\mathcal{CC}(T^\pi)| + |V_\pi|$  and  $|\mathcal{E}_\pi| = \sum_\ell |V_\ell|$ , so that

$$\tau_N^0[T^\pi] = N^{|\mathcal{V}_\pi|-1-|\mathcal{E}_\pi|} \left( 1 + O\left(\frac{1}{N}\right) \right) \times \prod_{\ell=1}^L \tau_N^0[T_\ell^\pi].$$

A bound for  $\prod_{\ell=1}^L \tau_N^0[T_\ell^\pi]$  is obtained from the growth condition (2.1), for which we need the following definition.

**Definition 4.8.** *1. Recall that a cut edge of a finite graph is an edge whose deletion increases the number of connected components.*

2. A two-edge connected graph is a connected graph that does not contain a cut edge. A two-edge connected component of a graph is a maximal two-edge connected subgraph.
3. For a finite graph  $G$ , we define  $F(G)$  to be the graph whose vertices are the two-edge connected components of  $G$  and whose edges are the cut edges linking the two-edge components that contain its adjacent vertices. Note that  $F(G)$  is always a forest: we call it the forest of two-edge connected components of  $G$ .
4. We define  $\mathfrak{f}(G)$  to be the number of leaves of  $F(G)$ , with the convention that the trees of  $F(G)$  that consist only of one vertex have two leaves.

**Lemma 4.9.** *We have the following estimate*

$$\prod_{\ell=1}^L \tau_N^0[T_\ell^\pi] = O\left(N^{\sum_\ell \mathfrak{f}(T_\ell^\pi)/2 - |\mathcal{CC}(T^\pi)|}\right).$$

*Proof.* Let  $\hat{T} = (\hat{V}, \hat{E})$  be a test graph. By (4.5), we have

$$\tau_N^0[\hat{T}] = \sum_{\sigma \in \mathcal{P}(\hat{V})} N^{c(\hat{T}^\sigma) - c(\hat{T})} \text{Mob}(0, \sigma) \tau_N[\hat{T}^\sigma].$$

Under the assumption of Proposition 2.3, we have  $\tau_N[\hat{T}^\sigma] = O(N^{\mathfrak{f}(\hat{T}^\sigma)/2 - c(\hat{T}^\sigma)})$ . But  $\mathfrak{f}(\hat{T}^\sigma) \leq \mathfrak{f}(\hat{T})$ , so we get  $\tau_N^0[\hat{T}] = O(N^{\mathfrak{f}(\hat{T})/2 - c(\hat{T})})$ . Applying this fact for each  $T_\ell^\pi$  provides the expected result.  $\square$

Thus, we have the estimate

$$\tau_N^0[T^\pi] = O(N^{|\mathcal{V}_\pi| - 1 - |\mathcal{E}_\pi| + \sum_\ell \mathfrak{f}(T_\ell^\pi)/2 - |\mathcal{CC}(T^\pi)|}). \quad (4.7)$$

Let us recall that, for a finite and undirected graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , denoting by  $\text{deg}_{\mathcal{G}}(v)$  the degree of a vertex  $v \in \mathcal{V}$ , we have

$$|\mathcal{E}| - |\mathcal{V}| = \sum_{v \in \mathcal{V}} \left( \frac{\text{deg}_{\mathcal{G}}(v)}{2} - 1 \right).$$

Assume  $\mathcal{GCC}(T^\pi)$  is not a tree. Let  $\tilde{\mathcal{G}}_\pi = (\tilde{\mathcal{V}}_\pi, \tilde{\mathcal{E}}_\pi)$  be the graph obtained *pruning*  $\mathcal{GCC}(T^\pi)$ , by removing the vertices of  $\mathcal{GCC}(T^\pi)$  that are of degree one (as well as the edges attached to these vertices), and iterating this procedure until it does not remain vertices of degree one. We denote by  $\tilde{V}_1$  and  $\tilde{V}_2$  the vertices of  $\mathcal{CC}(T^\pi)$  and  $V_\pi$  respectively that remain in  $\tilde{\mathcal{G}}_\pi$  after this process. Then we have

$$\begin{aligned} & |\mathcal{V}_\pi| - |\mathcal{E}_\pi| - 1 + \sum_{\ell=1}^L \frac{\mathfrak{f}(T_\ell^\pi)}{2} - |\mathcal{CC}(T^\pi)| \\ &= |\tilde{\mathcal{V}}_\pi| - |\tilde{\mathcal{E}}_\pi| - 1 + \sum_{S \in \mathcal{CC}(T^\pi)} \left( \frac{\mathfrak{f}(S)}{2} - 1 \right) \\ &= -1 - \sum_{S \in \tilde{V}_1} \left( \frac{\text{deg}_{\tilde{\mathcal{G}}_\pi}(S)}{2} - 1 \right) - \sum_{v \in \tilde{V}_2} \left( \frac{\text{deg}_{\tilde{\mathcal{G}}_\pi}(v)}{2} - 1 \right) + \sum_{S \in \mathcal{CC}(T^\pi)} \left( \frac{\mathfrak{f}(S)}{2} - 1 \right). \end{aligned}$$

Note that since  $T^\pi$  is two-edge connected, the colored components  $S$  such that  $\mathfrak{f}(S) \geq 3$  remain in  $\tilde{\mathcal{G}}_\pi$ . Indeed, since  $T^\pi$  is two-edge connected, each leaf of the tree of two-edge connected components of  $S$  corresponds to at least one vertex  $v \in \tilde{V}_\pi$ . So the



components  $S$  that have been removed have necessarily  $\mathfrak{f}(S) = 2$ . Moreover, for the same reason, we have  $\mathfrak{f}(S) \leq \deg_{\tilde{g}_\pi}(S)$ . Hence, since  $\deg_{\tilde{g}_\pi}(v) \geq 2$  for each  $v \in \tilde{V}_2$ , the last equality yields

$$|\mathcal{V}_\pi| - |\mathcal{E}_\pi| - 1 + \sum_{\ell} \frac{\mathfrak{f}(T_\ell^\pi)}{2} - |\mathcal{CC}(T^\pi)| \leq -1 - \sum_{v \in \tilde{V}_2} \left( \frac{\deg_{\tilde{g}_\pi}(v)}{2} - 1 \right) \leq -1.$$

By Formula (4.7), we get as expected  $\tau_N^0[T^\pi] \xrightarrow{N \rightarrow \infty} 0$ .  $\square$

#### 4.4 Proof of Theorem 2.2

Theorem 6 of [19] ensures that

$$\mathrm{Tr}(g(\mathbf{A}_{N,\ell})) \leq N^{\mathfrak{f}(g)/2} \prod_{e \in E} \|A_{N,\ell(e)}^{k(e)}\|$$

for any graph monomial  $g$  with set of edges  $E$  and equal input and output. Because the norm of our matrix are uniformly bounded, the matrices satisfies the growth assumption (2.1) of Proposition 2.3, and all the results of the previous section are valid.

#### 4.5 Proof of Theorem 1.2

Let  $\mathbf{A}_{N,1}, \dots, \mathbf{A}_{N,L}$  and  $\varepsilon_N$  be as in Theorem 1.2. By enlarging the index set  $K$  if necessary, let us denote by  $\mathbf{A}_{N,L+1}$  the family of the coefficients  $D_{i,\ell}$  of the polynomial  $P_{N,\ell}$  defining  $\varepsilon_N$ . We apply Theorem 2.2 to the family  $\mathbf{A}_{N,1}, \dots, \mathbf{A}_{N,L+1}$ , specified for the polynomials

$$\tilde{P}_{N,\ell} = D_{0,\ell} X_{k_i(1)} D_{1,\ell} \cdots X_{k_i(d_\ell)} D_{d_\ell,\ell}.$$

Note that each  $D_{i,\ell}$  can indeed be written as  $g_{i,\ell}(\mathbf{A}_N)$  for  $g_{i,\ell}$  a graph monomial consisting in a single loop labeled by the matrix  $D_{i,\ell}$  of  $\mathbf{A}_{N,L+1}$ . This proves the assertion of Theorem 1.2 for any  $P_{N,\ell}$  with fixed degree and bounded coefficients, which is sufficient to ensure the validity of the theorem in full generality.

#### 4.6 Proof of Proposition 2.4

Let us consider the families of random matrices  $\tilde{\mathbf{A}}_{N,1}, \dots, \tilde{\mathbf{A}}_{N,L}$ , where we recall that  $\tilde{\mathbf{A}}_{N,\ell} := (A_{N,\ell}^{(k)} \circ \Gamma_\ell^{(k)})_{k \in K}$ , and  $\Gamma = (\Gamma_\ell^{(k)})_{\ell,k}$  is a family of random matrices, with uniformly bounded entries, independent of the matrices  $(\mathbf{A}_{N,1}, \dots, \mathbf{A}_{N,L})$ .

According to Section 4.2, in order to have the conclusion of Theorem 2.2, it suffices to prove that  $\tau_N^0[T^\pi(\tilde{\mathbf{A}}_N)]$  converges to 0 for all test graph  $T$  and partition  $\pi$  such that  $\mathcal{GCC}(T^\pi)$  is not a tree. Using (4.6), we have

$$\begin{aligned} & \tau_N^0[T^\pi(\tilde{\mathbf{A}}_N)] \\ &= \frac{1}{N} \sum_{\substack{\phi: V \rightarrow [N] \\ \text{injective}}} \mathbb{E} \left[ \prod_{e=(v,w) \in E} A_{\ell(e)}^{k(e)}(\phi(w), \phi(v)) \right] \times \mathbb{E} \left[ \prod_{e=(v,w) \in E} \Gamma_{\ell(e)}^{k(e)}(\phi(w), \phi(v)) \right] \\ &= \tau_N^0[T^\pi(\mathbf{A}_N)] \cdot \frac{(N - |V^\pi|)!}{N!} \sum_{\substack{\phi: V \rightarrow [N] \\ \text{injective}}} \mathbb{E} \left[ \prod_{e=(v,w) \in E} \Gamma_{\ell(e)}^{k(e)}(\phi(w), \phi(v)) \right]. \end{aligned}$$

The term  $\tau_N^0[T^\pi(\mathbf{A}_N)]$  tends to 0 thanks to Section 4.3 and Section 4.4, and the rightmost term is bounded, since the entries of the matrices  $\Gamma_\ell^{(k)}$  are bounded. Thus the conclusion of Theorem 2.2 are true for  $\tilde{\mathbf{A}}_{N,1}, \dots, \tilde{\mathbf{A}}_{N,L}$ .

Reasoning as in Section 4.5, in order to have the conclusion of Theorem 1.2 true for  $\tilde{\mathbf{A}}_{N,1}, \dots, \tilde{\mathbf{A}}_{N,L}$ , it suffices to have the conclusion of Theorem 2.2 true for  $\tilde{\mathbf{A}}_{N,1}, \dots, \tilde{\mathbf{A}}_{N,L}, \tilde{\mathbf{A}}_{N,L+1}$ , where  $\tilde{\mathbf{A}}_{N,L+1}$  is a family of bounded deterministic diagonal matrices. We can write  $\tilde{\mathbf{A}}_{N,L+1}$  as  $(A_{N,\ell}^{(k)} \circ \Gamma_\ell^{(k)})_{k \in K}$  with  $A_{N,L+1}^{(k)} = I_N$  and  $\Gamma_{L+1}^{(k)} = \tilde{A}_{N,L+1}^{(k)}$ , in such a way that  $\mathbf{A}_{N,L+1} = (A_{N,L+1}^{(k)})_{k \in K}$  is permutation invariant. Then we have proved that the conclusion of Theorem 2.2 is true for  $\tilde{\mathbf{A}}_{N,1}, \dots, \tilde{\mathbf{A}}_{N,L+1}$ .

## References

- [1] Oskari H. Ajanki, László Erdős, and Torben Krüger. Universality for general Wigner-type matrices. *Probability Theory and Related Fields*, 169(3-4):667–727, 2017.
- [2] Benson Au. Traffic distributions of random band matrices. *arXiv e-prints*, (arXiv:1601.02188), January 2016.
- [3] S. Belinschi, A. Dembo, and A. Guionnet. Spectral measure of heavy tailed band and covariance random matrices. *Comm. Math. Phys.*, 289(3):1023–1055, 2009.
- [4] Serban T Belinschi, Tobias Mai, and Roland Speicher. Analytic subordination theory of operator-valued free additive convolution and the solution of a general random matrix problem. *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 2017(732):21–53, 2017.
- [5] Gérard Ben Arous and Alice Guionnet. The spectrum of heavy tailed random matrices. *Communications in Mathematical Physics*, 278(3):715–751, 2008.
- [6] Florent Benaych-Georges and Thierry Lévy. A continuous semigroup of notions of independence between the classical and the free one. *The Annals of Probability*, pages 904–938, 2011.
- [7] March Boedihardjo and Ken Dykema. Asymptotic \*-moments of some random Vandermonde matrices. *Adv. Math.*, 318:1–45, 2017.
- [8] Charles Bordenave and Alice Guionnet. Localization and delocalization of eigenvectors for heavy-tailed random matrices. *Probability Theory and Related Fields*, 157(3-4):885–953, 2013.
- [9] J. P. Bouchaud and P. Cizeau. Theory of Lévy matrices. *Phys. Rev. E*, 50(3):1810–1822, 1994.
- [10] Guillaume Cébron, Antoine Dahlqvist, and Camille Male. Universal constructions for spaces of traffics. *arXiv preprint arXiv:1601.00168*, 2016.
- [11] Ken Dykema. On certain free product factors via an extended matrix model. *J. Funct. Anal.*, 112(1):31–60, 1993.
- [12] Franck Gabriel. A combinatorial theory of random matrices III: random walks on  $\mathfrak{S}(N)$ , ramified coverings and the  $\mathfrak{S}(\infty)$  Yang-Mills measure. *arXiv preprint*, (arXiv:1510.01046), oct 2015.

- [13] Franck Gabriel. Combinatorial theory of permutation-invariant random matrices I: Partitions, geometry and renormalization. *arXiv preprint*, (arXiv:1503.02792), mar 2015.
- [14] Franck Gabriel. Combinatorial theory of permutation-invariant random matrices II: cumulants, freeness and Lévy processes. *arXiv preprint*, (arXiv:1507.02465), jul 2015.
- [15] O. Khorunzhy, M. Shcherbina, and V. Vengerovsky. Eigenvalue distribution of large weighted random graphs. *Journal of Mathematical Physics*, 45(4):1648–1672, 2004.
- [16] C. Male and S. Péché. Uniform regular weighted graphs with large degree: Wigner’s law, asymptotic freeness and graphons limit. *ArXiv e-prints*, October 2014.
- [17] Camille Male. Traffic distributions and independence: permutation invariant random matrices and the three notions of independence. *arXiv preprint, accepted for publication at Memoirs of the AMS*, (arXiv:1111.4662), nov 2011.
- [18] Camille Male. The limiting distributions of large heavy Wigner and arbitrary random matrices. *Journal of Functional Analysis*, 272(1):1 – 46, 2017.
- [19] James A. Mingo and Roland Speicher. Sharp bounds for sums associated to graphs of matrices. *Journal of Functional Analysis*, 262(5):2272 – 2288, 2012.
- [20] James A Mingo and Roland Speicher. *Free probability and random matrices*, volume 35. Springer, 2017.
- [21] Alexandru Nica and Roland Speicher. *Lectures on the combinatorics of free probability*, volume 335 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 2006.
- [22] Dimitri Shlyakhtenko. Random Gaussian band matrices and freeness with amalgamation. *International Mathematics Research Notices*, 1996(20):1013–1025, 1996.
- [23] Richard P. Stanley. *Enumerative combinatorics. Volume 1*, volume 49 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2012.
- [24] D. V. Voiculescu, K. J. Dykema, and A. Nica. *Free random variables*, volume 1 of *CRM Monograph Series*. American Mathematical Society, Providence, RI, 1992. A noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups.
- [25] Dan-Virgil Voiculescu. *Operations on certain non-commutative operator-valued random variables*. Preprint series in mathematics. Inst. de Mat., 1986.
- [26] Dan-Virgil Voiculescu. Limit laws for random matrices and free products. *Invent. Math.*, 104(1):201–220, 1991.
- [27] Inna Zakharevich. A generalization of Wigner’s law. *Communications in mathematical physics*, 268(2):403–414, 2006.