The Annals of Probability
2014, Vol. 42, No. 6, 2454–2496
DOI: 10.1214/13-AOP866

A LARGE DEVIATION PRINCIPLE FOR WIGNER MATRICES WITHOUT GAUSSIAN TAILS[1]

BY CHARLES BORDENAVE AND PIETRO CAPUTO

IMT UMR 5219 CNRS and Université Paul-Sabatier Toulouse III, and Università Roma Tre

We consider $n \times n$ Hermitian matrices with i.i.d. entries X_{ij} whose tail probabilities $\mathbb{P}(|X_{ij}| \geq t)$ behave like $e^{-at^{\alpha}}$ for some $a > 0$ and $\alpha \in (0, 2)$. We establish a large deviation principle for the empirical spectral measure of $X/\sqrt{n}$ with speed $n^{1+\alpha/2}$ with a good rate function $J(\mu)$ that is finite only if μ is of the form $\mu = \mu_{\mathrm{sc}} \boxplus \nu$ for some probability measure ν on $\mathbb{R}$, where $\boxplus$ denotes the free convolution and μ_{sc} is Wigner's semicircle law. We obtain explicit expressions for $J(\mu_{\mathrm{sc}} \boxplus \nu)$ in terms of the αth moment of ν. The proof is based on the analysis of large deviations for the empirical distribution of very sparse random rooted networks.

1. Introduction. Let $\mathcal{H}_n(\mathbb{C})$ denote the set of $n \times n$ Hermitian matrices. The *empirical spectral measure* of a matrix $A \in \mathcal{H}_n(\mathbb{C})$ is the probability measure on $\mathbb{R}$ defined by

$$\mu_A = \frac{1}{n} \sum_{k=1}^{n} \delta_{\lambda_k(A)},$$

where $\lambda_1(A) \geq \cdots \geq \lambda_n(A)$ denote the eigenvalues of A counting multiplicity. Below, we consider the empirical spectral measure of a *Wigner random matrix* X described as follows. Let $(X_{ij})_{1 \leq i < j}$ be i.i.d. complex random variables with variance $\mathbb{E}|X_{12} - \mathbb{E}X_{12}|^2 = 1$, and let $(X_{ii})_{i \geq 1}$ be an independent family of i.i.d. real random variables. Extend this array by setting $X_{ij} = \overline{X}_{ji}$ for $1 \leq j < i$, and consider the sequence of $n \times n$ Hermitian random matrices

$$X(n) = (X_{ij})_{1 \leq i, j \leq n}. \tag{1}$$

For ease of notation, we often drop the argument n and simply write X for $X(n)$.

The space $\mathcal{P}(\mathbb{R})$ of probability measures on $\mathbb{R}$ is endowed with the topology of weak convergence: a sequence of probability measures $(\mu_n)_{n \geq 1}$ *converges weakly* to μ if for any bounded continuous function $f : \mathbb{R} \mapsto \mathbb{R}$, $\int f \, d\mu_n \to \int f \, d\mu$ as

Received August 2012; revised March 2013.

[1]Supported by the GDRE GREFI-MEFI CNRS-INdAM and supported in part by the European Research Council through the "Advanced Grant" PTRELSS 228032.

MSC2010 subject classifications. 60B20, 47A10, 15A18.

Key words and phrases. Random matrices, spectral measure, large deviations, free convolution, random networks, local weak convergence.

n goes to infinity. We denote this convergence by $\mu_n \rightsquigarrow \mu$. Wigner's celebrated theorem asserts that almost surely,

$$\mu_{X/\sqrt{n}} \rightsquigarrow \mu_{\mathrm{sc}}, \tag{2}$$

where μ_{sc} is the semicircle law, that is, the probability measure with density $\frac{1}{2\pi}\sqrt{4-x^2}$ on $[-2, 2]$; see, for example, [3, 4, 19].

We consider large deviations, that is, events of the form $\mu_{X/\sqrt{n}} \in B$ where B is a measurable set in $\mathcal{P}(\mathbb{R})$ whose closure does not contain the limiting law μ_{sc}. Clearly, (2) implies that $\mathbb{P}(\mu_{X/\sqrt{n}} \in B) \to 0$, $n \to \infty$. It follows from known concentration estimates that if the entries X_{ij} are bounded, or if they satisfy a logarithmic Sobolev inequality, then $\mathbb{P}(\mu_{X/\sqrt{n}} \in B)$ decays to 0 as fast as e^{-cn^2} for some constant $c > 0$; see Guionnet and Zeitouni [17] or [3]. Further, if the X_{ij} have a Gaussian law such that X belongs to the Gaussian unitary ensemble GUE or the Gaussian orthogonal ensemble GOE, then a full large deviation principle for $\mu_{X/\sqrt{n}}$ with speed n^2 has been established by Ben Arous and Guionnet in [7]. However, apart from the GUE and GOE cases, we are not aware of any case for which the large deviation principle for $\mu_{X/\sqrt{n}}$ has been obtained. We refer to the recent work of Chatterjee and Varadhan [13] for the large deviations of the largest eigenvalues of X/n. For other models of random matrices where the joint law of the eigenvalues has a tractable form, large deviation principles have been proved; see, for example, [3], Section 2.6, or Eichelsbacher, Sommerauer and Stolz [16].

In this paper, we prove a large deviation principle under the assumption that X_{ij} has tail probabilities $\mathbb{P}(|X_{ij}| \geq t)$ of order e^{-at^α} for some $a > 0$, and $\alpha \in (0, 2)$. Before stating our assumptions and results in detail, let us make some preliminary remarks.

It is not hard to see why $n^{1+\alpha/2}$ is the natural speed for large deviations in our setting. For instance, for a fixed $x \in \mathbb{R}$, consider the event $|X_{ii}| \sim x\sqrt{n}$, for all $i = 1, \dots, n$, which has probability $e^{-cn^{1+\alpha/2}}$, for some $c > 0$. This event forces all eigenvalues of $X/\sqrt{n}$ to shift by x and, therefore, produces a shift by x of the limiting spectral measure μ_{sc}. Similarly, by considering deviations on the scale $\sqrt{n}$ of few elements X_{ij} in each row of the matrix X, one expects to be able to produce more general deformations of μ_{sc} at a cost of order $n^{1+\alpha/2}$ on the exponential scale. It turns out that this picture is correct, provided the deformations of μ_{sc} are of the form $\mu = \mu_{\mathrm{sc}} \boxplus \nu$ for some $\nu \in \mathcal{P}(\mathbb{R})$, where $\boxplus$ denotes the *free convolution*. Roughly speaking, the idea is that the entries of X that are visible on a scale $\sqrt{n}$ form a very sparse weighted random graph or random network G_n that is asymptotically independent from the rest of the matrix, and a large deviation principle for $\mu_{X/\sqrt{n}}$ can be deduced from a large deviation principle for the law of the random network G_n. This approach also allows us to obtain explicit expressions for the rate function. The strategy of proof developed in the present work for Wigner matrices could certainly be generalized to other models such as random covariance matrices or random band matrices with the same type of tail assumptions on the entries. Large deviations with speed n^α of the largest eigenvalue may also be handled with similar techniques.

Main result. We recall that a sequence of random variables $(Z_n)_{n\geq 1}$ with values in a topological space $\mathcal{X}$ with Borel σ-field $\mathcal{B}$, satisfies the *large deviation principle* (LDP) with rate function J and speed v, if $J:\mathcal{X}\mapsto[0,\infty]$ is a lower semicontinuous function, $v:\mathbb{N}\mapsto[0,\infty)$ is a function which increases to infinity, and for every $B\in\mathcal{B}$:

$$\begin{aligned}(3)\qquad -\inf_{x\in B^\circ} J(x) &\leq \liminf_{n\to\infty}\frac{1}{v(n)}\log\mathbb{P}(Z_n\in B)\leq \limsup_{n\to\infty}\frac{1}{v(n)}\log\mathbb{P}(Z_n\in B)\\ &\leq -\inf_{x\in\overline{B}} J(x),\end{aligned}$$

where B° denotes the interior of B and $\overline{B}$ denotes the closure of B. We recall that the lower semicontinuity of J means that the level sets $\{x\in\mathcal{X}:J(x)\leq t\}$, $t\in[0,\infty)$, are closed subsets of $\mathcal{X}$. When the level sets are compact, the rate function J is said to be *good*.

We now introduce our statistical assumption. Let $a,\alpha\in(0,\infty)$. We say that a complex random variable Y belongs to the class $\mathcal{S}_\alpha(a)$, and write $Y\in\mathcal{S}_\alpha(a)$, if

$$(4)\qquad \lim_{t\to\infty} -t^{-\alpha}\log\mathbb{P}\big(|Y|\geq t\big)=a,$$

and if $Y/|Y|$ and $|Y|$ are independent for large values of $|Y|$, that is, there exists $t_0>0$ and a probability $\vartheta\in\mathcal{P}(\mathbb{S}^1)$ on the unit circle $\mathbb{S}^1$ such that for all $t\geq t_0$, all measurable sets $U\subset\mathbb{S}^1$, one has

$$(5)\qquad \mathbb{P}\big(Y/|Y|\in U \text{ and } |Y|\geq t\big)=\vartheta(U)\mathbb{P}\big(|Y|\geq t\big).$$

For instance, if Y is Weibull, that is, Y is a nonnegative random variable with distribution function $F(t)=1-e^{-at^\alpha}$, with $\alpha>0$, and $a>0$, then $Y\in\mathcal{S}_\alpha(a)$, with $\vartheta=\delta_1$, the unit mass at the point 1. Clearly, if $Y\in\mathcal{S}_\alpha(a)$ is real valued, then the associated measure ϑ must have support in $\{-1,1\}$. It will be convenient to allow the value $a=\infty$ in (4). Namely, for $\alpha>0$ we write $Y\in\mathcal{S}_\alpha(\infty)$ whenever (4) holds with $a=\infty$. We do not require (5) in this case. For instance, if Y is a bounded random variable, then $Y\in\mathcal{S}_\alpha(\infty)$, for all $\alpha>0$, and if Y has a Gaussian tail, then $Y\in\mathcal{S}_\alpha(\infty)$, for all $\alpha\in(0,2)$. Moreover, if $Y\in\mathcal{S}_\alpha(a)$ for some $\alpha,a>0$, then $Y\in\mathcal{S}_\beta(\infty)$ for all $\beta\in(0,\alpha)$. We remark that (5) is a mild technical condition that we do not expect to be crucial. However, it will turn out to be convenient for the analysis of random networks in Section 3 below.

Throughout the paper, we assume that the array $\{X_{ij}\}$ is given as above, that is, we have two independent families of random variables: the off-diagonal entries X_{ij}, $i<j$, which are i.i.d. copies of a complex random variable X_{12} with unit variance, and the on-diagonal entries X_{ii}, which are i.i.d. copies of a real random variable X_{11}. The matrix $X=X(n)$ is defined as in (1). Moreover, the following main assumption will always be understood without explicit mention.

ASSUMPTION 1. There exist $\alpha \in (0,2)$ and $a, b \in (0,\infty]$ such that $X_{12} \in \mathcal{S}_\alpha(a)$ and $X_{11} \in \mathcal{S}_\alpha(b)$.

The main result can be formulated as follows.

THEOREM 1.1. *Fix $\alpha \in (0,2)$ as in Assumption 1. The measures $\mu_{X/\sqrt{n}}$ satisfy the LDP with speed $n^{1+\alpha/2}$ and good rate function*

$$J(\mu) = \begin{cases} \Phi(\nu), & \text{if } \mu = \mu_{\text{sc}} \boxplus \nu \text{ for some } \nu \in \mathcal{P}(\mathbb{R}), \\ \infty, & \text{otherwise}, \end{cases} \tag{6}$$

where $\Phi : \mathcal{P}(\mathbb{R}) \mapsto [0,\infty]$ is a good rate function.

More details on the rate function Φ will be given in Theorems 1.2 and 1.3 below. We anticipate that $\Phi(\nu) = 0$ if and only if $\nu = \delta_0$, where δ_0 is the Dirac mass at 0. Moreover, as one should expect, in the case $a = b = \infty$, one has $\Phi(\nu) = \infty$ for all $\nu \neq \delta_0$.

The proof of Theorem 1.1 consists of two main parts. The first part, the "random matrix theory part" of the work, is discussed in Section 2. Here, we show that at speed $n^{1+\alpha/2}$ the large deviations are governed by the sparse $n \times n$ random matrix $C = C(n)$ defined by

$$C_{ij} = \begin{cases} \dfrac{X_{ij}}{\sqrt{n}}, & \text{if } \varepsilon(n) \le \dfrac{X_{ij}}{\sqrt{n}} \le \varepsilon(n)^{-1}, \\ 0, & \text{otherwise}, \end{cases}$$

where $\varepsilon(n)$ is a cutoff sequence that for convenience will be set equal to $1/\log n$. In particular, we show that as far as the LDP with speed $n^{1+\alpha/2}$ is concerned, $\mu_{X/\sqrt{n}}$ behaves as $\mu_{\text{sc}} \boxplus \mu_C$, where μ_C is the spectral measure of the matrix C; see Proposition 2.1 below. As a consequence, the LDP for $\mu_{X/\sqrt{n}}$ will be obtained by contraction if one has the LDP for μ_C with speed $n^{1+\alpha/2}$ and rate function Φ.

The second part, the "random graph theory part" of the work, is presented in Section 3. Here, we prove the above mentioned LDP for the spectral measures μ_C. By viewing the matrix C as the adjacency matrix of a weighted graph, one runs naturally into the analysis of large deviations for sparse random networks. This is best formulated within the theory of local convergence for networks that was recently developed by Benjamini and Schramm [5], Aldous and Steele [2] and Aldous and Lyons [1]. Let us briefly sketch the main ideas—all details will be given in Section 3. Let G_n be the sparse random network naturally associated to the $n \times n$ matrix C, that is, G_n is the weighted graph with n vertices whose adjacency matrix is given by C. Notice that the weights can have a sign, and there are loops corresponding to nonzero diagonal entries of C. Take a vertex at random, call it the root, and consider the connected component of G_n at that vertex. This gives rise to a random connected rooted network, we call ρ_n its law. By identifying two networks which differ only by a permutation of the vertex labels, the law ρ_n is regarded as an element of the space $\mathcal{P}(\mathcal{G}_*)$ of probability measures on $\mathcal{G}_*$, where $\mathcal{G}_*$ is the space of equivalence classes (under rooted isomorphisms) of connected rooted networks. The essential point is that the eigenvalue distribution μ_C can be identified with a suitable "spectral measure" μ_{ρ_n} associated to the law ρ_n; see also [9–11] for recent works based on the same idea.

Since the network G_n is very sparse, one has that almost surely ρ_n converges (under the weak local convergence [1]) to the Dirac mass on the trivial element of $\mathcal{G}_*$, namely the network consisting of a single isolated vertex (the root). We introduce a suitable weak topology on $\mathcal{P}(\mathcal{G}_*)$, and prove that the measures ρ_n satisfy a LDP with speed $n^{1+\alpha/2}$ and a good rate function $I(\rho)$. The latter is finite only if ρ belongs to the so called sofic measures, that is, if ρ is the weak local limit of finite networks, and if the support of ρ satisfies some natural constraints. Call $\mathcal{P}_s(\mathcal{G}_*)$ the set of such probability measures. We find that for $\rho \in \mathcal{P}_s(\mathcal{G}_*)$, one has

$$I(\rho) = b\mathbb{E}_\rho\left|\omega_G(o)\right|^\alpha + \frac{a}{2}\mathbb{E}_\rho \sum_{v \in V_G \setminus o} \left|\omega_G(o, v)\right|^\alpha, \tag{7}$$

where $\mathbb{E}_\rho$ denotes expectation w.r.t. ρ, the law of the equivalence class of a connected rooted network (G, o), o denoting the root; $\omega_G(o)$ denotes the weight of the loop at the root, and $\omega_G(o, v)$ denotes the weight of the edge (o, v) if v is an element of the vertex set V_G of the network. We refer to Proposition 3.9 for the precise result.

It turns out that the choice of a "myopic" topology on $\mathcal{P}(\mathcal{G}_*)$ is crucial to have the desired result. On the other hand, we want this topology to be fine enough to have that the map $\rho \mapsto \mu_\rho$ defining the spectral measure associated to ρ is continuous. If all this is satisfied, then a LDP for the spectral measure $\mu_C = \mu_{\rho_n}$ can be obtained by contraction from the LDP for ρ_n; see Proposition 3.14. In particular, we find that the function Φ in Theorem 1.1 is given by

$$\Phi(\nu) = \inf\{I(\rho), \rho \in \mathcal{P}_s(\mathcal{G}_*) : \mu_\rho = \nu\}. \tag{8}$$

We now turn to more explicit characterizations of the rate function in Theorem 1.1. From the approach discussed above, we will see that the rate function Φ depends on the laws of X_{11} and X_{12} only through α, a, b and the supports of the associated measures on $\mathbb{S}^1$. While the variational principle (8) is not always explicitly solvable, there is a large class of $\nu \in \mathcal{P}(\mathbb{R})$ for which $\Phi(\nu)$ can be computed. This allows us to give explicit expressions for the rate function $J(\mu)$ in Theorem 1.1. Recall that the free convolution with μ_{sc} is injective: for any $\mu \in \mathcal{P}(\mathbb{R})$ there is at most one $\nu \in \mathcal{P}(\mathbb{R})$ such that $\mu = \mu_{\mathrm{sc}} \boxplus \nu$. Let $\mathcal{P}_{\mathrm{sym}}(\mathbb{R})$ denote the set of symmetric probability measures on $\mathbb{R}$. If $\mu = \mu_{\mathrm{sc}} \boxplus \nu$, then $\mu \in \mathcal{P}_{\mathrm{sym}}(\mathbb{R})$ is equivalent to $\nu \in \mathcal{P}_{\mathrm{sym}}(\mathbb{R})$. For more details on free convolution with the semicircular distribution, we refer to Biane [8]. For $\nu \in \mathcal{P}(\mathbb{R})$, we use the notation

$$m_\alpha(\nu) = \int |x|^\alpha \, d\nu(x) \tag{9}$$

for the αth moment of ν. If $X_{11} \in \mathcal{S}_\alpha(b)$ for some $b < \infty$, then we write ϑ_b for the associated measure given in (5). Recall that since X_{11} is real, ϑ_b is a measure on $\{-1, 1\}$. The following theorem summarizes the main facts we can establish about the rate function.

THEOREM 1.2. (a) *For any* $\nu \in \mathcal{P}(\mathbb{R})$,

$$\Phi(\nu) \geq \left(\frac{a}{2} \wedge b\right) m_\alpha(\nu).$$

(b) *If* $b < \infty$ *and* $\operatorname{supp}(\vartheta_b) = \{-1, 1\}$, *then for any* $\nu \in \mathcal{P}(\mathbb{R})$:

$$\Phi(\nu) \leq b m_\alpha(\nu).$$

(c) *If* $b < \infty$ *and* $\operatorname{supp}(\vartheta_b) = \{-1, 1\}$, *and* $\nu \in \mathcal{P}_{\mathrm{sym}}(\mathbb{R})$, *then*

$$\Phi(\nu) = \left(\frac{a}{2} \wedge b\right) m_\alpha(\nu).$$

Some remarks about Theorem 1.2. Part (a) shows clearly that $\Phi(\nu) = 0$ is equivalent to $\nu = \delta_0$, that is, $J(\mu) = 0$ is equivalent to $\mu = \mu_{\mathrm{sc}}$. It also shows that J is a good rate function since the level sets $\{m_\alpha(\cdot) \leq t\}$, $t \in [0, \infty)$ are compact in $\mathcal{P}(\mathbb{R})$. Concerning the remaining statements, the fact that the moments $m_\alpha(\nu)$ appear naturally in the rate function and the special role played by symmetric measures ν can be understood as follows.

As one could expect, there is a natural way to achieve a large deviation $\mu_{X/\sqrt{n}} \sim \mu_{\mathrm{sc}} \boxplus \nu$ by tilting only the diagonal entries of X, namely by considering events of the form $\mu_{\mathcal{D}/\sqrt{n}} \sim \nu$, where $\mathcal{D}$ denotes the diagonal matrix with entries $X_{11}, \dots, X_{nn}$, and

$$\mu_{\mathcal{D}/\sqrt{n}} = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_{i,i}/\sqrt{n}}.$$

In view of (5), one can consider an arbitrary $\nu \in \mathcal{P}(\mathbb{R})$ here if $b < \infty$ and $\operatorname{supp}(\vartheta_b) = \{-1, 1\}$. If $b < \infty$ and $\operatorname{supp}(\vartheta_b) = \{+1\}$ (or $\{-1\}$) then only ν whose support is $\mathbb{R}_+$ (or $\mathbb{R}_-$) can be considered. If $b = \infty$, then no measure $\nu \neq \delta_0$ will have a finite cost on the scale $n^{1+\alpha/2}$.

Similarly, one can try to reach a large deviation $\mu_{X/\sqrt{n}} \sim \mu_{\mathrm{sc}} \boxplus \nu$ by tilting only the off-diagonal entries of X. For instance, for n even, let $\mathcal{A}$ denote the block diagonal matrix made up of the 2×2 blocks

$$\begin{pmatrix} 0 & X_{i,i+1} \\ \bar{X}_{i,i+1} & 0 \end{pmatrix}, \qquad i = 1, \dots, n/2.$$

That is, $\mathcal{A}$ is defined by $\mathcal{A}_{2i-1,2i} = X_{i,i+1}$, $\mathcal{A}_{2i,2i-1} = \bar{X}_{i,i+1}$, $i = 1, \dots, n/2$, and $\mathcal{A}_{i,j} = 0$ for all other entries. It is straightforward to see that the empirical spectral measures of $\mathcal{A}/\sqrt{n}$ is given by

$$\mu_{\mathcal{A}/\sqrt{n}} = \frac{1}{n}\sum_{i=1}^{n/2}(\delta_{|X_{i,i+1}|/\sqrt{n}} + \delta_{-|X_{i,i+1}|/\sqrt{n}}).$$

Notice that $\mu_{\mathcal{A}/\sqrt{n}}$ is a symmetric distribution. Thus, if we try to obtain $\mu_{X/\sqrt{n}} \sim \mu_{\mathrm{sc}} \boxplus \nu$ by requiring $\mu_{\mathcal{A}/\sqrt{n}} \sim \nu$ we are forced to restrict to $\nu \in \mathcal{P}_{\mathrm{sym}}(\mathbb{R})$.

In view of this discussion, it is natural to look for upper bounds on the rate function Φ in terms of the rate function associated to large deviations of $\mu_{\mathcal{D}/\sqrt{n}}$ and $\mu_{\mathcal{A}/\sqrt{n}}$. Our results will show in particular that if the variables X_{ij} are as in Assumption 1, with $b < \infty$ and $\mathrm{supp}(\vartheta_b) = \{-1, 1\}$, then:

(1) $\mu_{\mathcal{D}/\sqrt{n}}$ satisfies a LDP on $\mathcal{P}(\mathbb{R})$ with speed $n^{1+\alpha/2}$ and rate function $I_b(\nu) = bm_\alpha(\nu)$, for all $\nu \in \mathcal{P}(\mathbb{R})$;

(2) $\mu_{\mathcal{A}/\sqrt{n}}$ satisfies a LDP on $\mathcal{P}(\mathbb{R})$ with speed $n^{1+\alpha/2}$ and rate function equal to $I_a(\nu) = \frac{a}{2}m_\alpha(\nu)$, for all $\nu \in \mathcal{P}_{\mathrm{sym}}(\mathbb{R})$, and $I_a(\nu) = +\infty$ if $\nu \notin \mathcal{P}_{\mathrm{sym}}(\mathbb{R})$.

Since $\mu_{\mathcal{D}/\sqrt{n}}$ and $\mu_{\mathcal{A}/\sqrt{n}}$ are the empirical measures induced by i.i.d. random variables rescaled by $\sqrt{n}$, the statements above can be seen as extremal instances of Sanov's theorem, in the case of variables with exponential tails of the form (4). Thus, roughly speaking, part (b) in Theorem 1.2 can be interpreted as the bound obtained by adopting the strategy $\mu_{\mathcal{D}/\sqrt{n}} \sim \nu$ to reach the deviation $\mu_{X/\sqrt{n}} \sim \mu_{\mathrm{sc}} \boxplus \nu$. When $b \le a/2$, parts (a) and (b) above yield the expression

$$\Phi(\nu) = bm_\alpha(\nu),$$

showing that this strategy is optimal. Similarly, to illustrate part (c), observe that if $\nu \in \mathcal{P}_{\mathrm{sym}}(\mathbb{R})$, then for the deviation $\mu_{X/\sqrt{n}} \sim \mu_{\mathrm{sc}} \boxplus \nu$ one can also use the strategy $\mu_{\mathcal{A}/\sqrt{n}} \sim \nu$. This reasoning will produce the bound $\Phi(\nu) \le (a/2 \wedge b)m_\alpha(\nu)$. The general bound in part (a) then shows that this is actually an optimal strategy if $a/2 \le b$.

If the support of ϑ_b is only $\{+1\}$ (or $\{-1\}$) then the above scenario changes in that one can use the diagonal matrix $\mathcal{D}$ only to reach deviations ν whose support is $\mathbb{R}_+$ (or $\mathbb{R}_-$). In this case, we have the following estimates. Without loss of generality, we restrict to $\mathrm{supp}(\vartheta_b) = \{+1\}$.

THEOREM 1.3. *Suppose* $b < \infty$, *and* $\mathrm{supp}(\vartheta_b) = \{+1\}$.

(a) *If* $\mathrm{supp}(\nu) \subset \mathbb{R}_+$, *then*

$$\Phi(\nu) \le bm_\alpha(\nu).$$

(b) *Suppose* $\alpha \in (1, 2)$. *If* $\nu \in \mathcal{P}_{\mathrm{sym}}(\mathbb{R})$, *then*

$$\Phi(\nu) = \frac{a}{2}m_\alpha(\nu).$$

(c) *Suppose* $\alpha \in (1, 2)$. *If* $\int x\,d\nu(x) < 0$ *then* $\Phi(\nu) = +\infty$.

The above result can be interpreted as before by appealing to the large deviations of $\mu_{\mathcal{D}/\sqrt{n}}$ and $\mu_{\mathcal{A}/\sqrt{n}}$. In particular, part (b) shows that since one cannot realize a symmetric deviation $\nu \in \mathcal{P}_{\mathrm{sym}}(\mathbb{R})$ using the matrix $\mathcal{D}$ only, it is less costly to realize it using the matrix $\mathcal{A}$ only. Similarly, in part (c), one has that neither $\mathcal{D}$ nor $\mathcal{A}$, nor any other matrix with vanishing trace, can be used to produce a measure ν with $\int x\,d\nu(x) < 0$ and, therefore, the rate function must be $+\infty$. We believe that results in parts (b) and (c) above should hold without the additional condition $\alpha \in (1, 2)$.

The proofs of Theorems 1.2 and 1.3 are given in Section 3.10.

2. Exponential equivalences. Throughout the rest of the paper, we fix the cutoff sequence $\varepsilon(n)$ as

$$\varepsilon(n) = \frac{1}{\log n}. \tag{10}$$

We decompose the matrix X as

$$\frac{X}{\sqrt{n}} = A + B + C + D, \tag{11}$$

where the matrices A, B, C, D are defined by

$$A_{ij} = \mathbf{1}_{|X_{ij}|<(\log n)^{2/\alpha}} \frac{X_{ij}}{\sqrt{n}}, \qquad B_{ij} = \mathbf{1}_{(\log n)^{2/\alpha} \le |X_{ij}| \le \varepsilon(n) n^{1/2}} \frac{X_{ij}}{\sqrt{n}},$$

$$C_{ij} = \mathbf{1}_{\varepsilon(n) n^{1/2} < |X_{ij}| < \varepsilon(n)^{-1} n^{1/2}} \frac{X_{ij}}{\sqrt{n}}, \qquad D_{ij} = \mathbf{1}_{\varepsilon(n)^{-1} n^{1/2} < |X_{ij}|} \frac{X_{ij}}{\sqrt{n}}.$$

The matrix A represents the bulk of the original matrix, while the matrix C yields the elements that are visible on the scale $\sqrt{n}$. The starting point of our analysis (see Lemmas 2.2 and 2.3 below) is to show that the contribution of both B and D is negligible for large deviations with speed $n^{1+\alpha/2}$.

We define the distance on $\mathcal{P}(\mathbb{R})$ as

$$d(\mu, \nu) = \sup\{|g_\mu(z) - g_\nu(z)| : \mathfrak{Im}(z) \ge 2\}, \tag{12}$$

where g_μ is the *Cauchy–Stieltjes transform* of μ, that is, for $z \in \mathbb{C}_+ = \{z \in \mathbb{C} : \mathfrak{Im}(z) > 0\}$,

$$g_\mu(z) = \int \frac{\mu(dx)}{x - z}. \tag{13}$$

Recall that this distance is a metric for the weak convergence; see, for example, [3], Theorem 2.4.4. Let also d_{KS} denote the Kolmogorov–Smirnov distance and let W_p denote the L^p-Wasserstein distance; see Appendix B below for the relevant definitions. The introduction of the distance d_{KS} is mainly due to the use of the rank inequality of Lemma B.1. The Wasserstein distance on the other hand can be controlled in terms of the matrix elements thanks to the Hoffman–Wielandt inequality in Lemma B.2. We shall relate these distances to the distance (12) via the following estimate, which is a consequence of (75) and (77):

$$d(\mu,\nu) \le d_{\mathrm{KS}}(\mu,\nu) \wedge W_1(\mu,\nu). \tag{14}$$

The following proposition is the first major step on the way to prove Theorem 1.1.

PROPOSITION 2.1. *The random probability measures $\mu_{\mathrm{sc}} \boxplus \mu_C$ and $\mu_{X/\sqrt{n}}$ are exponentially equivalent: for any $\delta > 0$,*

$$\limsup_{n\to\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}\big(d(\mu_{X/\sqrt{n}}, \mu_{\mathrm{sc}} \boxplus \mu_C) \ge \delta\big) = -\infty.$$

The rest of this section is devoted to the proof of Proposition 2.1. The strategy is as follows: we start by showing that the contribution of D in (11) can be neglected (Lemma 2.2), then we show that B can also be neglected (Lemma 2.3). The last step will then consist in proving that μ_{A+C} and $\mu_{\mathrm{sc}} \boxplus \mu_C$ are exponentially equivalent. We note that the assumption (5) is not needed for the proof of Proposition 2.1. Actually, a careful look at the proof shows that it is sufficient to replace condition (4) by the weaker assumption $\limsup_{t\to\infty} t^{-\alpha} \log \mathbb{P}(|Y| \ge t) < 0$; see Remark 2.7 below.

2.1. *Preliminary estimates.*

LEMMA 2.2 (Very large entries). *The random probability measures μ_{A+B+C} and $\mu_{X/\sqrt{n}}$ are exponentially equivalent: for any $\delta > 0$,*

$$\limsup_{n\to\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}\big(d(\mu_{X/\sqrt{n}}, \mu_{A+B+C}) \ge \delta\big) = -\infty.$$

PROOF. From (14), it is sufficient to prove that for any $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}\big(d_{\mathrm{KS}}(\mu_{X/\sqrt{n}}, \mu_{A+B+C}) \ge \delta\big) = -\infty.$$

Then, using the rank inequality Lemma B.1, it is sufficient to prove that for any $\delta > 0$

$$\limsup_{n\to\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}\big(\mathrm{rank}(D) \ge \delta n\big) = -\infty.$$

However, since the rank is bounded by the number of nonzeros entries of a matrix, one has

$$\mathbb{P}\big(\mathrm{rank}(D) \ge 2\delta n\big) \le \mathbb{P}\Bigg(\sum_{1\le i\le j\le n} \mathbf{1}\big(|X_{ij}| \ge \varepsilon(n)^{-1} n^{1/2}\big) \ge \delta n\Bigg).$$

The Bernoulli variables $\mathbf{1}(|X_{ij}| \geq \varepsilon(n)^{-1}n^{1/2})$, $1 \leq i \leq j \leq n$, are independent. Also, by assumption (4), their mean value $p_{ij} = \mathbb{P}(|X_{ij}| \geq \varepsilon(n)^{-1}n^{1/2})$ satisfies

$$p_{ij} \leq p(n) := e^{-c\varepsilon(n)^{-\alpha}n^{\alpha/2}}$$

for some $c > 0$. For our choice of $\varepsilon(n)$ in (10), one has $p(n) = o(1/n^2)$. Hence, it is sufficient to prove that for any $\delta > 0$:

$$\limsup_{n\to\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}\left(\sum_{1\leq i\leq j\leq n} (\mathbf{1}(|X_{ij}| \geq \varepsilon(n)^{-1}n^{1/2}) - p_{ij}) \geq \delta n\right) = -\infty.$$

Recall Bennett's inequality [6]: if W_i, $i = 1, \dots, m$ are independent Bernoulli(p_i) variables, and $h(x) = (x+1)\log(x+1) - x$, then one has

$$\mathbb{P}\left(\sum_{i=1}^{m} (W_i - p_i) \geq t\right) \leq \exp\left(-\sigma^2 h\left(\frac{t}{\sigma^2}\right)\right) \tag{15}$$

with $\sigma^2 = \sum_{i=1}^m p_i(1-p_i)$. In our case, for all n large enough,

$$\sigma^2 = \sum_{1\leq i\leq j\leq n} p_{ij}(1-p_{ij}) \leq \frac{n(n+1)p(n)}{2}.$$

Therefore, using $h(x) \sim x \log x$ as $x \to \infty$,

$$\begin{aligned}\mathbb{P}\left(\sum_{1\leq i\leq j\leq n} (\mathbf{1}(|X_{ij}| \geq \varepsilon(n)^{-1}n^{1/2}) - p_{ij}) \geq \delta n\right) &\leq \exp\left(-\sigma^2 h\left(\frac{n\delta}{\sigma^2}\right)\right) \\ &\leq \exp(c_0 n \log(np(n)))\end{aligned}$$

for some constant $c_0 > 0$ depending on δ. Now, since $n \leq p(n)^{-1/2}$ for n large, we find that for some $c_1 > 0$, for all n large enough the last expression is upper bounded by

$$\exp\left(\frac{1}{2}c_0 n \log p(n)\right) \leq \exp(-c_1 n^{1+\alpha/2}\varepsilon(n)^{-\alpha}).$$

This proves the claim. $\square$

We now show that the contribution of B in (11) is also negligible. While Lemma 2.2 would work for any $\alpha > 0$, the next results use the fact that $\alpha \in (0, 2)$.

LEMMA 2.3 (Moderately large entries). *The random probability measures μ_{A+C} and $\mu_{X/\sqrt{n}}$ are exponentially equivalent: for any $\delta > 0$,*

$$\limsup_{n\to\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}(d(\mu_{X/\sqrt{n}}, \mu_{A+C}) \geq \delta) = -\infty.$$

PROOF. From (14), Lemma 2.2 and the triangle inequality, it is sufficient to check that for any $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}\big(W_2(\mu_{A+B+C}, \mu_{A+C}) \geq \delta\big) = -\infty,$$

where $W_2 \geq W_1$ is the L^2-Wasserstein distance defined by (76). From the Hoffman–Wielandt inequality Lemma B.2, it is sufficient to prove that for any $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}\left(\frac{1}{n}\operatorname{tr}(B^2) \geq \delta\right) = -\infty.$$

We write

$$\frac{1}{n}\operatorname{tr}(B^2) \leq \frac{2}{n^2} \sum_{1\leq i\leq j\leq n} |X_{ij}|^2 \mathbf{1}\big((\log n)^{2/\alpha} \leq |X_{ij}| \leq \varepsilon(n) n^{1/2}\big).$$

Thus, from Markov's inequality, for any $\lambda > 0$,

$$\mathbb{P}\left(\frac{1}{n}\operatorname{tr}(B^2) \geq 2\delta\right) \leq e^{-\lambda\delta} \prod_{1\leq i,j\leq n} \mathbb{E}\big[e^{n^{-2}\lambda|X_{ij}|^2 \mathbf{1}((\log n)^{2/\alpha} \leq |X_{ij}| \leq \varepsilon(n)n^{1/2})}\big].$$

To estimate the last expectation, we use the integration by part formula, for $\mu \in \mathcal{P}(\mathbb{R})$ and $g \in C^1$,

$$\begin{aligned}\int_a^b g(x)\,d\mu(x) &= g(a)\mu\big([a,\infty)\big) - g(b)\mu\big((b,\infty)\big) \\ &\quad + \int_a^b g'(x)\mu\big([x,\infty)\big)\,dx.\end{aligned} \tag{16}$$

Define the function

$$f(x) = n^{-2}\lambda x^2 - cx^\alpha. \tag{17}$$

Let μ denote the law of $|X_{ij}|$, and $g(x) = e^{n^{-2}\lambda x^2}$. By Assumption 1, there exists a constant $c > 0$ such that

$$\mu\big([t,\infty)\big) = \mathbb{P}\big(|X_{ij}| \geq t\big) \leq \exp\big(-ct^\alpha\big) \tag{18}$$

for all t large enough. In particular, $g(t)\mu([t,\infty)) \leq e^{f(t)}$. From (16), it follows that

$$\begin{aligned}&\mathbb{E}\big[e^{n^{-2}\lambda|X_{ij}|^2 \mathbf{1}((\log n)^{2/\alpha} \leq |X_{ij}| \leq \varepsilon(n)n^{1/2})}\big] \\ &\quad \leq 1 + \int_{(\log n)^{2/\alpha}}^{\varepsilon(n)n^{1/2}} g(x)\,d\mu(x) \\ &\quad \leq 1 + e^{f((\log n)^{2/\alpha})} + \int_{(\log n)^{2/\alpha}}^{\varepsilon(n)n^{1/2}} \frac{2\lambda x}{n^2} e^{f(x)}\,dx \\ &\quad \leq 1 + e^{f((\log n)^{2/\alpha})} + \frac{\lambda\varepsilon(n)^2}{n} \max_{x\in[(\log n)^{2/\alpha}, \varepsilon(n)n^{1/2}]} e^{f(x)}.\end{aligned} \tag{19}$$

We choose $\lambda = \frac{1}{2}c\varepsilon(n)^{\alpha-2}n^{1+\alpha/2}$, with the constant $c > 0$ given in (18). Simple computations show that $f(x)$ reaches its maximum for $x \in [(\log n)^{2/\alpha}, \varepsilon(n)n^{1/2}]$ at $x = (\log n)^{2/\alpha}$, where it is equal to

$$\frac{1}{2}c\varepsilon(n)^{\alpha-2}n^{\alpha/2-1}(\log n)^{4/\alpha} - c(\log n)^2.$$

Using (10), for $n \geq n_0$ this is smaller than $-\frac{c}{2}(\log n)^2$. Therefore, using $1+x \leq e^x$, $x \geq 0$, one has that (19) is bounded by $\exp[e^{-(c/4)(\log n)^2}]$ for n large enough. It follows that

$$\frac{1}{n^{1+\alpha/2}} \log \mathbb{P}\left(\frac{1}{n}\operatorname{tr}(B^2) \geq 2\delta\right) \leq -\frac{1}{2}c\delta\varepsilon(n)^{\alpha-2} + n^{1-\alpha/2}e^{-(c/4)(\log n)^2}.$$

The desired conclusion follows. $\square$

For $s > 0$, we define the compact set for the weak topology

$$K_s = \left\{\mu \in \mathcal{P}(\mathbb{R}) : \int x^2\, d\mu \leq s\right\}.$$

For a suitable choice of s, we now check that μ_C is in K_s with large probability.

LEMMA 2.4 (Exponential tightness estimates).

$$\limsup_{n\to\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}(\mu_C \notin K_{(\log n)^2}) = -\infty.$$

Moreover, if $I = \{(i,j) : |X_{ij}| > (\log n)^{2/\alpha}\}$, *for any* $\delta > 0$,

$$\lim_{n\to\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}\big(|I| \geq \delta n^{1+\alpha/2}\big) = -\infty.$$

PROOF. Notice that

$$\int x^2\, d\mu_C = \frac{1}{n}\operatorname{tr}(C^2) \leq \frac{2}{n^2} \sum_{1\leq i\leq j\leq n} |X_{ij}|^2 \mathbf{1}\big(\varepsilon(n)n^{1/2} < |X_{ij}| \leq \varepsilon(n)^{-1}n^{1/2}\big).$$

We may repeat the argument in the proof of Lemma 2.3. This time we take $\lambda = \frac{1}{2}c\varepsilon(n)^{2-\alpha}n^{1+\alpha/2}$, where c is as in (18), and then define f as in (17). For any $s > 0$, one has

$$\mathbb{P}(\mu_C \notin K_{2s}) \leq e^{-\lambda s}\left(1 + e^{f(\varepsilon(n)\sqrt{n})} + \frac{1}{2}cn^{\alpha/2}\varepsilon(n)^{-\alpha} \max_{x\in[\varepsilon(n)n^{1/2}, \varepsilon(n)^{-1}n^{1/2}]} e^{f(x)}\right)^{n^2}.$$

Simple considerations show that $f(x)$, for $x \in [\varepsilon(n)n^{1/2}, \varepsilon(n)^{-1}n^{1/2}]$ is maximized at $x = \varepsilon(n)n^{1/2}$, where it satisfies $f(\varepsilon(n)n^{1/2}) \leq -\frac{1}{2}c\varepsilon(n)^{\alpha}n^{\alpha/2}$. This gives, for n large enough,

$$\frac{1}{n^{1+\alpha/2}} \log \mathbb{P}(\mu_C \notin K_{2s}) \leq -\frac{1}{2}cs\varepsilon(n)^{2-\alpha} + n^{1-\alpha/2}e^{-(1/4)c\varepsilon(n)^{\alpha}n^{\alpha/2}}.$$

We choose finally $s = 1/(2\varepsilon(n)^2)$. For our choice of $\varepsilon(n)$ in (10), this implies the first claim.

For the second claim, we have

$$\mathbb{P}\big(|I| \geq 2\delta n^{1+\alpha/2}\big) \leq \mathbb{P}\bigg(\sum_{1\leq i\leq j\leq n} \mathbf{1}\big(|X_{ij}| \geq (\log n)^{2/\alpha}\big) \geq \delta n^{1+\alpha/2}\bigg).$$

The Bernoulli variables $\mathbf{1}(|X_{ij}| \geq (\log n)^{2/\alpha})$, $1 \leq i \leq j \leq n$, are independent. Also, by Assumption 1, their average $p_{ij} = \mathbb{P}(|X_{ij}| \geq (\log n)^{2/\alpha})$ satisfies

$$p_{ij} \leq p'(n) := e^{-c(\log n)^2}$$

for some $c > 0$. We argue as in the proof of Lemma 2.2. From Bennett's inequality (15),

$$\mathbb{P}\bigg(\sum_{1\leq i\leq j\leq n} \big(\mathbf{1}\big(|X_{ij}| \geq (\log n)^{2/\alpha}\big) - p_{ij}\big) \geq \delta n^{1+\alpha/2}\bigg)$$
$$\leq \exp\bigg(-c_0 n^{1+\alpha/2} \log\bigg(\frac{n^{\alpha/2-1}}{p'(n)}\bigg)\bigg)$$

for some constant $c_0 = c_0(\delta) > 0$. Since $p'(n) = o(n^{\alpha/2-1})$, this gives the claim. □

2.2. *Auxiliary estimates.* To complete the proof of Proposition 2.1, we shall need two extra results. The first is due to Guionnet and Zeitouni [17], Corollary 1.4.

THEOREM 2.5 (Concentration for matrices with bounded entries). *Let $\kappa \geq 1$, let $Y \in \mathcal{H}_n(\mathbb{C})$ be a random matrix with independent entries $(Y_{ij})_{1\leq i\leq j\leq n}$ bounded by κ, and let $M \in \mathcal{H}_n(\mathbb{C})$ be a deterministic matrix such that $\int x^2\, d\mu_M \leq \kappa^2$. There exists a universal constant $c > 0$ such that for all $(c\kappa^2/n)^{2/5} \leq t \leq 1$,*

$$\mathbb{P}\big(W_1(\mu_{Y/\sqrt{n}+M}, \mathbb{E}\mu_{Y/\sqrt{n}+M}) \geq t\big) \leq \frac{c\kappa}{t^{3/2}} \exp\bigg(-\frac{n^2 t^5}{c\kappa^4}\bigg).$$

In [17], Corollary 1.4, the result is stated for matrices Y in $\mathcal{H}_n(\mathbb{C})$ such that the entries have independent real and imaginary parts. The extension to our setting follows by using a version of Talagrand's concentration inequality for independent bounded variables in $\mathbb{C}$. Also, the matrix M is not present in [17]. It is, however, not hard to check that its presence does not change the argument in [17], page 132, since one can use the bound

$$\int x^2\, d\mu_{Y/\sqrt{n}+M} \leq 2\int x^2\, d\mu_{Y/\sqrt{n}} + 2\int x^2\, d\mu_M \leq 4\kappa^2.$$

The latter is an easy consequence of, for example, Lemma B.2.

The second result we need is a uniform bound on the rate of the convergence of the empirical spectral measure of sums of random matrices.

THEOREM 2.6 (Uniform asymptotic freeness). *Let $Y = (Y_{ij})_{1\le i,j\le n} \in \mathcal{H}_n(\mathbb{C})$ be a Wigner random matrix with* $\mathrm{Var}(Y_{12}) = 1$, $\mathbb{E}|Y_{12}|^3 < \infty$ *and* $\mathbb{E}|Y_{11}|^2 < \infty$. *There exists a universal constant $c > 0$ such that for any integer $n \ge 1$ and any $M \in \mathcal{H}_n(\mathbb{C})$,*

$$d(\mathbb{E}\mu_{Y/\sqrt{n}+M}, \mu_{\mathrm{sc}} \boxplus \mu_M) \le c\frac{\sqrt{\mathbb{E}|Y_{11}|^2} + \mathbb{E}|Y_{12}|^3}{\sqrt{n}}.$$

A striking point of the above theorem is that the constant c does not depend on M. The result is a variation around Pastur and Shcherbina [19], Theorem 18.3.1. The detailed proof of Theorem 2.6 is given in Appendix A below. We are now ready to finish the proof of Proposition 2.1.

2.3. *Proof of Proposition* 2.1. By Lemmas 2.2 and 2.3, it is sufficient to prove that μ_{A+C} and $\mu_{\mathrm{sc}} \boxplus \mu_C$ are exponentially equivalent: for any $\delta > 0$,

$$\lim_{n\to\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}\big(d(\mu_{\mathrm{sc}} \boxplus \mu_C, \mu_{A+C}) \ge \delta\big) = -\infty. \tag{20}$$

Let $\mathcal{F}$ be the σ-algebra generated by the random variables

$$\{X_{ij}\mathbf{1}_{|X_{ij}|\ge(\log n)^{2/\alpha}}\}.$$

Then the random matrix C is $\mathcal{F}$-measurable. Define the event

$$E = \left\{\int x^2\, d\mu_C \le (\log n)^2\right\}.$$

Then $E \in \mathcal{F}$. Lemma 2.4 implies that for some sequence $s_1(n) \to \infty$ and all n large enough,

$$\mathbb{P}(E^c) \le e^{-s_1(n)n^{1+\alpha/2}}. \tag{21}$$

Conditional on $\mathcal{F}$, $\sqrt{n}A$ is a random matrix with independent entries $(\sqrt{n}A_{ij})_{1\le i\le j\le n}$ bounded by $(\log n)^{2/\alpha}$. Thus, we may apply Theorem 2.5 with $Y/\sqrt{n}$ replaced by A, and M replaced by C. Using (14) to replace $W_1(\cdot,\cdot)$ by $d(\cdot,\cdot)$, taking $t = \delta$, and $\kappa = (\log n)^{2/\alpha}$ in Theorem 2.5, one has that for all $\delta > 0$, there is a sequence $s_2(n) \to \infty$, $n \to \infty$, such that

$$\mathbf{1}_E \mathbb{P}_{\mathcal{F}}\big(d(\mathbb{E}_{\mathcal{F}}\mu_{A+C}, \mu_{A+C}) \ge \delta\big) \le e^{-s_2(n)n^{1+\alpha/2}}, \tag{22}$$

where $\mathbb{P}_{\mathcal{F}}$ and $\mathbb{E}_{\mathcal{F}}$ are the conditional probability and expectation given $\mathcal{F}$. Notice that Theorem 2.5 can be applied here since on the event E one has $\int x^2\, d\mu_C \le (\log n)^2 \le \kappa^2$. Moreover, (22) holds uniformly within E, since the bound of Theorem 2.5 is uniform with respect to M satisfying $\int x^2\, d\mu_M \le \kappa^2$.

From (21) and (22), using the triangle inequality one has that (20) follows once we prove that for any $\delta > 0$:

$$\lim_{n\to\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}\big(d(\mu_{\mathrm{sc}} \boxplus \mu_C, \mathbb{E}_{\mathcal{F}}\mu_{A+C}) \ge \delta\big) = -\infty. \tag{23}$$

Next, we use a coupling argument to remove the dependency between A and C. Let P_n be the law of X_{12} conditioned on $\{|X_{12}| < (\log n)^{2/\alpha}\}$, and Q_n be the law of X_{11} conditioned on $\{|X_{11}| < (\log n)^{2/\alpha}\}$. We also define $I = \{(i, j) : |X_{ij}| \geq (\log n)^{2/\alpha}\}$. Given $\mathcal{F}$, if $(i, j) \in I$, then $A_{ij} = 0$ while, if $(i, j) \notin I$ and $1 \leq i \leq j \leq n$, then $\sqrt{n} A_{ij}$ has conditional law P_n or Q_n depending on whether $i < j$ or $i = j$.

On our probability space, we now consider Y an independent Hermitian random matrix such that $(Y_{ij})_{1 \leq i \leq j \leq n}$ are independent, and for $1 \leq i \leq n$, Y_{ii} has law Q_n, while for $1 \leq i < j \leq n$, Y_{ij} has law P_n. We form the matrix

$$A'_{ij} = \mathbf{1}\big((i, j) \notin I\big) A_{ij} + \mathbf{1}\big((i, j) \in I\big) \frac{Y_{ij}}{\sqrt{n}}.$$

By construction, $\sqrt{n} A'$ and Y have the same distribution and are independent of $\mathcal{F}$. Also, by Lemma B.2 and Jensen's inequality,

$$\begin{aligned}\mathbb{E}_{\mathcal{F}} d(\mu_{A+C}, \mu_{A'+C}) &\leq \sqrt{\mathbb{E}_{\mathcal{F}} \frac{\operatorname{tr}(A - A')^2}{n}} \\ &\leq \sqrt{\frac{1}{n^2} \sum_{1 \leq i, j \leq n} \mathbb{E}_{\mathcal{F}} \mathbf{1}\big((i, j) \in I\big) |Y_{ij}|^2} \leq c_0 \sqrt{\frac{|I|}{n^2}},\end{aligned}$$

where we have used the fact that, for some constant $c_0 > 0$,

$$\max\big(\mathbb{E}|Y_{11}|^2, \mathbb{E}|Y_{12}|^2\big) \leq c_0^2.$$

Define the event

$$F = \big\{|I| \leq \delta^2 n^2 / c_0^2\big\}.$$

Then $F \in \mathcal{F}$ and

$$\mathbf{1}_F \mathbb{E}_{\mathcal{F}} d(\mu_{A+C}, \mu_{A'+C}) \leq \delta. \tag{24}$$

From Lemma 2.4, for some sequence $s_3(n) \to \infty$, for all n large enough,

$$\mathbb{P}\big(F^c\big) \leq e^{-s_3(n) n^{1+\alpha/2}}. \tag{25}$$

Observe that by definition of the distance (12),

$$d(\mathbb{E}_{\mathcal{F}} \mu_{A'+C}, \mathbb{E}_{\mathcal{F}} \mu_{A+C}) \leq \mathbb{E}_{\mathcal{F}} d(\mu_{A'+C}, \mu_{A+C}).$$

Since A' and $Y/\sqrt{n}$ have the same distribution, we deduce from (24), (25) and the triangle inequality that the proof of (23) can be reduced to the proof of

$$\lim_{n \to \infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}\big(d(\mu_{\mathrm{sc}} \boxplus \mu_C, \mathbb{E}_{\mathcal{F}} \mu_{Y/\sqrt{n}+C}) \geq \delta\big) = -\infty. \tag{26}$$

Clearly, $\mathbb{E}|Y_{12}|^3 \le c_0(\log n)^{6/\alpha}$ and $\sigma^2 = \mathrm{Var}(Y_{12}) \to 1$. We may apply the uniform estimate of Theorem 2.6, applied to $Y/(\sigma\sqrt{n})$ and $M = C$, which is $\mathcal{F}$-measurable. We find for any $\delta > 0$,

$$\mathbb{P}\big(d(\mu_{\mathrm{sc}} \boxplus \mu_C, \mathbb{E}_{\mathcal{F}}\mu_{Y/(\sigma\sqrt{n})+C}) \ge \delta\big) = 0$$

for all $n \ge n_0(\delta)$ where $n_0(\delta)$ is a constant depending only on δ.

On the other hand, arguing as above, from Hoffman–Wielandt's inequality (Lemma B.2) and Jensen's inequality, for any $\delta > 0$,

$$\begin{aligned} d(\mathbb{E}_{\mathcal{F}}\mu_{Y/\sqrt{n}+C}, \mathbb{E}_{\mathcal{F}}\mu_{Y/(\sigma\sqrt{n})+C}) &\le \mathbb{E}_{\mathcal{F}} d(\mu_{Y/\sqrt{n}+C}, \mu_{Y/(\sigma\sqrt{n})+C}) \\ &\le \mathbb{E}_{\mathcal{F}}\sqrt{\frac{(1-1/\sigma)^2}{n^2}\operatorname{tr}(Y^2)} \\ &\le \frac{|1-1/\sigma|}{n}\sqrt{\mathbb{E}\operatorname{tr}(Y^2)} \le \delta \end{aligned}$$

for all $n \ge n_1(\delta)$ where $n_1(\delta)$ is a constant depending only on δ.

This concludes the proof of (26) and of Proposition 2.1.

REMARK 2.7. In the proof of Proposition 2.1, we have only used the following assumptions on the Wigner matrix X: (i) $\mathrm{Var}(X_{12}) = 1$ and (ii) there exists $c > 0$ such that for all $i \le j$,

$$\limsup_{t\to\infty} \frac{1}{t^\alpha} \log \mathbb{P}\big(|X_{ij}| \ge t\big) \le -c.$$

3. Large deviations of very sparse rooted networks. In this section, we start by adapting to our setting the notion of local weak convergence of rooted networks, introduced in [2, 5] and [1]. Next, we introduce a suitable projective limit topology on the space of networks. Then we prove the LDP for the network G_n induced by the very sparse matrix C. Finally, we introduce the spectral measure associated to a network and project the LDP for networks onto a LDP for spectral measures.

3.1. *Locally finite Hermitian networks.* Let V be a countable set, the *vertex* set. A pair $(u, v) \in V^2$ is an *oriented edge*. A *network* or *weighted graph* $G = (V, \omega)$ is a vertex set V together with a map ω from V^2 to $\mathbb{C}$. We say that a network is *Hermitian*, if for all $(u, v) \in V^2$,

$$\omega(u, v) = \overline{\omega(v, u)}.$$

For ease of notation, we sometimes set $\omega(v) = \omega(v, v)$ for the weight of the loop at v. The degree of v in G is defined by

$$\deg(v) = \sum_{u\in V} \big|\omega(v, u)\big|^2.$$

The network G is *locally finite* if for any vertex v, $\deg(v) < \infty$.

A path π from u to v in V is a sequence $\pi = (u_0, \dots, u_k)$ with $u_0 = u$, $u_k = v$ and, for $1 \le i \le k$, $|\omega(u_{i-1}, u_i)| > 0$. If such $\pi : u \to v$ exists, then one defines the ℓ_2 distance

$$D_\pi(u, v) = \left(\sum_{i=1}^{k} \bigl|\omega(u_{i-1}, u_i)\bigr|^{-2}\right)^{1/2}.$$

The distance between u and v is defined as

$$D(u, v) = \inf_{\pi : u \to v} D_\pi(u, v).$$

Notice that weights are thought of as inverse of distances. If there is no path $\pi : u \to v$, then the distance $D(u, v)$ is set to be infinite. A network is *connected* if $D(u, v) < \infty$ for any $u \ne v \in V$.

All networks we consider below will be Hermitian and locally finite, but not necessarily connected. We call $\mathcal{G}$ the set of all such networks. For a network $G \in \mathcal{G}$, to avoid possible confusion, we will often denote by $V_G, \omega_G, \deg_G$ the corresponding vertex set, weight and degree functions.

Clearly, any $n \times n$ Hermitian matrix $H_n \in \mathcal{H}_n(\mathbb{C})$ defines a finite network $G = G(H_n)$ in a natural way, by taking

$$V_G = \{1, \dots, n\}, \qquad \omega_G(i, j) = H_n(i, j). \tag{27}$$

For simplicity, we often write simply H_n instead of $G(H_n)$.

3.2. *Rooted networks*. Below, a *rooted network* $(G, o) = (V, \omega, o)$ is a Hermitian, locally finite and connected network (V, ω) with a distinguished vertex $o \in V$, the root. For $t > 0$, we denote by $(G, o)_t$ the rooted network with vertex set $\{u \in V : D(o, u) \le t\}$, and with the weights induced by ω. Two rooted networks $(G_i, o_i) = (V_i, \omega_i, o_i)$, $i \in \{1, 2\}$, are *isomorphic* if there exists a bijection $\sigma : V_1 \to V_2$ such that $\sigma(o_1) = o_2$ and $\sigma(G_1) = G_2$, where σ acts on G_1 through $\sigma(u, v) = (\sigma(u), \sigma(v))$ and $\sigma(\omega) = \omega \circ \sigma$.

We define the semidistance d_{loc} between two rooted networks (G_1, o_1) and (G_2, o_2) to be

$$d_{\mathrm{loc}}\bigl((G_1, o_1), (G_2, o_2)\bigr) = \frac{1}{1 + T},$$

where T is the supremum of those $t > 0$ such that there is a bijection $\sigma : V_{(G_1, o_1)_t} \to V_{(G_2, o_2)_t}$ with $\sigma(o_1) = o_2$ and such that the function $\omega_{G_2} \circ \sigma - \omega_{G_1}$ is bounded by $1/t$ on $V^2_{(G_1, o_1)_t}$.

The rooted network isomorphism defines a space $\mathcal{G}_*$ of equivalence classes of rooted networks (G, o). On the space $\mathcal{G}_*$, d_{loc} becomes a distance. The associated topology will be referred to as the *local topology*. We write $\mathbf{g}$ for an element of $\mathcal{G}_*$. We shall denote the convergence on $(\mathcal{G}_*, d_{\mathrm{loc}})$ by $d_{\mathrm{loc}}(\mathbf{g}_n, \mathbf{g}) \to 0$ or $\mathbf{g}_n \overset{\mathrm{loc}}{\to} \mathbf{g}$.

The space $(\mathcal{G}_*, d_{\mathrm{loc}})$ is separable and complete [1]. Let $\mathcal{P}(\mathcal{G}_*)$ denote the space of probability measures on $\mathcal{G}_*$. For $\mu, \mu_n \in \mathcal{P}(\mathcal{G}_*)$, we write $\mu_n \overset{\mathrm{loc}}{\rightsquigarrow} \mu$ when μ_n converges weakly, that is, when $\int f\, d\mu_n \to \int f\, d\mu$ for every bounded continuous function f on $(\mathcal{G}_*, d_{\mathrm{loc}})$. This notion of weak convergence is often referred to as *local weak convergence*. See [1] for more details and examples.

For a network $G \in \mathcal{G}$, and $v \in V_G$, one writes $G(v)$ for the connected component of G at v, that is, the largest connected network $G' \subset G$ with $v \in V_{G'}$. If $G \in \mathcal{G}$ is finite, that is, V_G is finite, one defines the probability measure $U(G) \in \mathcal{P}(\mathcal{G}_*)$ as the law of the equivalence class of the rooted network $(G(o), o)$ where the root o is sampled uniformly at random from V_G:

$$U(G) = \frac{1}{V_G} \sum_{v \in V_G} \delta_{\mathbf{g}(v)},$$

where $\mathbf{g}(v)$ stands for the equivalence class of $(G(v), v)$. If $G_n, n \geq 1$, is a sequence of finite networks from $\mathcal{G}$, we shall say that G_n has *local weak limit* $\rho \in \mathcal{P}(\mathcal{G}_*)$ if $U(G_n) \overset{\mathrm{loc}}{\rightsquigarrow} \rho$.

3.3. *Sofic measures*. Following [1], a measure $\rho \in \mathcal{P}(\mathcal{G}_*)$ is called *sofic* if there exists a sequence of finite networks $G_n, n \geq 1$, whose local weak limit is ρ. We shall need to identify a subset of the sofic measures. Let ϑ_a, ϑ_b denote the laws of $X_{12}/|X_{12}|$ and $X_{11}/|X_{11}|$, respectively, for $X_{12} \in \mathcal{S}_\alpha(a)$ and $X_{11} \in \mathcal{S}_\alpha(b)$; see Assumption 1, and let $S_a, S_b \subset \mathbb{S}^1$ denote their supports. Let $\mathcal{A}_n \subset \mathcal{H}_n(\mathbb{C})$ be the set of $n \times n$ Hermitian matrices H such that either $H_{ij} = 0$ or $H_{ij}/|H_{ij}| \in S_a$ for all $i < j$, and such that either $H_{ii} = 0$ or $H_{ii}/|H_{ii}| \in S_b$ for all i. We say that $\rho \in \mathcal{P}(\mathcal{G}_*)$ is *admissible sofic* if there exists a sequence of matrices $H_n \in \mathcal{A}_n$ such that $U(H_n) \overset{\mathrm{loc}}{\rightsquigarrow} \rho$, where H_n is identified with the associated network $G(H_n)$ as in (27). We denote by $\mathcal{P}_s(\mathcal{G}_*)$ the set of admissible sofic probability measures. Measures in $\mathcal{P}_s(\mathcal{G}_*)$ will often be called simply sofic if no confusion can arise.

Let $\mathbf{g}_\varnothing$ stand for the trivial network consisting of a single isolated vertex (the root) with zero weights. We refer to $\mathbf{g}_\varnothing$ as the empty network. Clearly, the Dirac mass at the empty network $\rho = \delta_{\mathbf{g}_\varnothing}$ is sofic (it suffices to consider matrices with zero entries). Let us consider some more examples.

EXAMPLE 3.1. Suppose that $S_b = \{-1, +1\}$. Let $Y_1, Y_2, \ldots$ be i.i.d. random variables with distribution $\nu \in \mathcal{P}(\mathbb{R})$. Consider the random diagonal matrix H_n with $H_n(i,i) = Y_i$. Then, by the law of large numbers, almost surely $U(H_n) \overset{\mathrm{loc}}{\rightsquigarrow} \rho$, where ρ is given by

$$\rho = \int_{\mathbb{R}} \delta_{\mathbf{g}_x}\, d\nu(x),$$

if $\mathbf{g}_x$ is the network consisting of a single vertex (the root) with loop weight equal to x.

EXAMPLE 3.2. Suppose that $Z_1, Z_3, Z_5, \dots$ are i.i.d. complex random variables with law $\mu \in \mathcal{P}(\mathbb{C})$ such that μ-a.s. one has either $Z_1 = 0$, or $Z_1/|Z_1| \in S_a$. Consider the $n \times n$ matrix H such that $H_n(j, j+1) = Z_j$, $H_n(j+1, j) = \bar{Z}_j$, for all odd $1 \le j \le n-1$, and all other entries of H_n are zero. By construction, $H_n \in \mathcal{A}_n$ almost surely. From the law of large numbers, almost surely $U(H_n) \overset{\text{loc}}{\rightsquigarrow} \rho$, where ρ is given by

$$\rho = \frac{1}{2} \int_{\mathbb{C}} (\delta_{\hat{\mathbf{g}}_z} + \delta_{\hat{\mathbf{g}}_{\bar{z}}})\, d\mu(z),$$

if $\hat{\mathbf{g}}_z$ denotes the equivalence class of the two vertex network (V, ω, o), with $V = \{o, 1\}$, $\omega(o, 1) = z$, $\omega(1, o) = \bar{z}$ and $\omega(o, o) = \omega(1, 1) = 0$.

EXAMPLE 3.3. For any fixed $n \in \mathbb{N}$, if $H_n \in \mathcal{A}_n$, then $U(H_n) \in \mathcal{P}_s(\mathcal{G}_*)$. Indeed, take a sequence of $m \times m$ matrices $A_m \in \mathcal{A}_m$ defined as follows. Let $k, r \ge 0$, with $r < n$, be integers such that $m = kn + r$, and take A_m as the block diagonal matrix with the first k blocks all equal to H_n and the last block of size r equal to zero. Then $U(A_m) = \frac{n}{n+(r/k)} U(H_n) + \frac{1}{1+(kn/r)} \delta_{\mathbf{g}_\varnothing}$. As $m \to \infty$, $r/k \to 0$, $kn/r \to \infty$ and, therefore, $U(A_m)$ converges to $U(H_n)$.

3.4. *Truncated networks.* It will be important to work with suitable truncations of the weights. To this end we consider, for $0 < \theta < 1$, networks $G \in \mathcal{G}$ such that for any $(u, v) \in V_G^2$,

$$\deg_G(v) \le \theta^{-2} \quad \text{and} \quad |\omega_G(u, v)| \ge \theta \mathbf{1}(\omega_G(u, v) \ne 0). \tag{28}$$

We call $\mathcal{G}^\theta$ the set of all such networks. Clearly, any $G \in \mathcal{G}^\theta$ is locally finite and has at most θ^{-4} outgoing nonzero edges from any vertex. As before, one defines the space $\mathcal{G}_*^\theta$ by taking equivalence classes of connected rooted networks from $\mathcal{G}^\theta$. We define $\mathcal{P}(\mathcal{G}_*^\theta)$ as the sets of $\rho \in \mathcal{P}(\mathcal{G}_*)$ with support in $\mathcal{G}_*^\theta$, and set $\mathcal{P}_s(\mathcal{G}_*^\theta) = \mathcal{P}(\mathcal{G}_*^\theta) \cap \mathcal{P}_s(\mathcal{G}_*)$.

LEMMA 3.4. (i) *$\mathcal{P}_s(\mathcal{G}_*)$ is closed for the local weak topology.*

(ii) *For any $\theta > 0$, $\mathcal{G}_*^\theta$ is a compact set for the local topology.*

PROOF. For (i): by definition, $\mathcal{P}_s(\mathcal{G}^*)$ is the closure of the set of $U(G)$ such that G is an admissible finite network [i.e., for some integer $n \ge 1$, $H \in \mathcal{A}_n$ and $G = G(H)$ as in (27)].

For (ii): let $\mathbf{g} \in \mathcal{G}_*^\theta$ and (G, o) be a rooted network in the equivalence class $\mathbf{g}$. Observe that each edge of G has a weight bounded above by θ^{-1}. This implies that in G each path whose total length is bounded by $t > 0$, contains at most t^2/θ^2 edges. Moreover, G has at most θ^{-4} outgoing edges from any vertex. Hence, G has at most $n(t) = \theta^{-4t^2/\theta^2}$ vertices at distance less than t from any given vertex.

Now, we denote by $\mathcal{G}_*^{\theta,t}$ the set of equivalence classes of $(G,o)_t$ such that the equivalence class of (G,o) is in $\mathcal{G}_*^{\theta}$. There is a finite number, say $m(t)$, of equivalence classes of rooted connected graphs with less than $n(t)$ vertices (without weights). Since all weights of $\mathbf{g} \in \mathcal{G}_*^{\theta}$ are in $[\theta, \theta^{-1}]$, there is a covering of $\mathcal{G}_*^{\theta,t}$ with balls of radius $1/(1+t)$ of cardinal at most $k(t) = m(t)(t\theta^{-1})^{n(t)^2}$.

Notice that for any rooted network $d_{\mathrm{loc}}((G,o),(G,o)_t) \le 1/(1+t)$. Hence, from the definition of d_{loc}, we have proved that, for any $t>0$, there exists a finite covering of $\mathcal{G}_*^{\theta}$ with balls of radius $1/(1+t)$. This proves that $\mathcal{G}_*^{\theta}$ is precompact. The fact that $\mathcal{G}_*^{\theta}$ is closed follows directly from (28). □

Next, we describe a canonical way to obtain a network in $\mathcal{G}^{\theta}$ by truncating a network from $\mathcal{G}$. This will allow us to introduce a topology on $\mathcal{P}(\mathcal{G}_*)$ that is weaker than the local weak topology. In particular, a topology for which $\mathcal{P}_s(\mathcal{G}_*)$ is compact; compare Lemmas 3.4 and 3.8. For $0<\theta<1$, define the two continuous functions

$$\chi_\theta(x) = \begin{cases} 0, & \text{if } x \in [0,\theta), \\ (x-\theta)/\theta, & \text{if } x \in [\theta, 2\theta), \\ 1, & \text{if } x \in [2\theta, \infty), \end{cases}$$

$$\widetilde{\chi}_\theta(x) = \begin{cases} 1, & \text{if } x \in [0,\theta^{-2}-1), \\ \theta^{-2}-x, & \text{if } x \in [\theta^{-2}-1, \theta^{-2}), \\ 0, & \text{if } x \in [\theta^{-2}, \infty) \end{cases}$$

that will serve as approximations for the indicator functions $\mathbf{1}(x \ge \theta)$ and $\mathbf{1}(x \le \theta^{-2})$.

If $G=(V,\omega)$, we define $\widetilde{G}_\theta = (V, \widetilde{\omega}_\theta)$ as the network with vertex set V and, for all $u,v \in V$,

$$\widetilde{\omega}_\theta(u,v) = \omega(u,v)\widetilde{\chi}_\theta\big(\deg_G(u) \vee \deg_G(v)\big). \tag{29}$$

Next, we define $G_\theta = (V, \omega_\theta)$ as the network with vertex set V and, for all $u,v \in V$,

$$\omega_\theta(u,v) = \widetilde{\omega}_\theta(u,v)\chi_\theta\big(|\widetilde{\omega}_\theta(u,v)|\big). \tag{30}$$

Clearly, G_θ satisfies (28), and for any $u,v \in V$, $|\omega_{G_\theta}(u,v)| \le \theta^{-1}$, and

$$\deg_{G_\theta}(u) \le \deg_G(u) \quad \text{and} \quad \big|\omega_{G_\theta}(u,v)\big| \le \big|\omega_G(u,v)\big|. \tag{31}$$

If $\mathbf{g} \in \mathcal{G}_*$ and the network (G,o) is in the equivalence class $\mathbf{g}$, then $\mathbf{g}_\theta \in \mathcal{G}_*^{\theta}$ is defined as the equivalence class of $(G_\theta(o), o)$, where G_θ is defined by (30). This defines a map $\mathbf{g} \mapsto \mathbf{g}_\theta$ from $\mathcal{G}_*$ to $\mathcal{G}_*^{\theta}$. If $\rho \in \mathcal{P}(\mathcal{G}_*)$ and $\mathbf{g}$ has law ρ, the law of $\mathbf{g}_\theta$ defines a new measure $\rho_\theta \in \mathcal{P}(\mathcal{G}_*^{\theta})$.

The next lemma follows easily from the continuity of χ_θ, $\widetilde{\chi}_\theta$ and the fact that as $\theta \to 0$, for any for $x>0$, $\chi_\theta(x) \to 1$ and $\widetilde{\chi}_\theta(x) \to 1$.

LEMMA 3.5 (Continuity of projections).

(i) *For* $\theta > 0$, *the map* $\mathbf{g} \mapsto \mathbf{g}_\theta$ *from* $\mathcal{G}_* \to \mathcal{G}_*^\theta$ *is continuous for the local topology*;

(ii) *for* $\theta > 0$, *the map* $\rho \mapsto \rho_\theta$ *from* $\mathcal{P}(\mathcal{G}_*)$ *to* $\mathcal{P}(\mathcal{G}_*^\theta)$ *is continuous for the local weak topology*;

(iii) *as* $\theta \to 0$, *one has* $\mathbf{g}_\theta \stackrel{\text{loc}}{\to} \mathbf{g}$ *and* $\rho_\theta \stackrel{\text{loc}}{\rightsquigarrow} \rho$, *for any* $\mathbf{g} \in \mathcal{G}_*$ *and* $\rho \in \mathcal{P}(\mathcal{G}_*)$.

3.5. *Projective topology for locally finite rooted networks*. In order to circumvent the lack of compacity of $\mathcal{P}_s(\mathcal{G}_*)$ w.r.t. local weak topology, we now introduce a weaker topology, the *projective topology*. For integers $j \geq 1$, set

$$\theta_j = 2^{-j}.$$

Let $p_j : \mathcal{G}_* \to \mathcal{G}_*^{\theta_j}$ be defined by $p_j(\mathbf{g}) = \mathbf{g}_{\theta_j}$. Similarly, for $1 \leq i \leq j$, $p_{ij} : \mathcal{G}_*^{\theta_j} \to \mathcal{G}_*^{\theta_i}$ is the map $p_{ij}(\mathbf{g}) = \mathbf{g}_{\theta_i}$, $\mathbf{g} \in \mathcal{G}_*^{\theta_j}$. The collection $(p_{ij})_{1 \leq i \leq j}$ is a *projective system* in the sense that for any $1 \leq i \leq j \leq k$,

$$p_{ik} = p_{ij} \circ p_{jk}. \tag{32}$$

The latter follows from $2\theta_{j+1} \leq \theta_j$ and $\theta_j^{-2} \leq \theta_{j+1}^{-2} - 1$.

Define the projective space $\widetilde{\mathcal{G}}_* \subset \prod_{j \geq 1} \mathcal{G}_*^{\theta_j}$ as the set of $y = (y_1, y_2, \dots) \in \prod_{j \geq 1} \mathcal{G}_*^{\theta_j}$ such that for any $i \leq j$, $p_{ij}(y_j) = y_i$; see, for example, [14], Appendix B, for more details on projective spaces. One can identify $\mathcal{G}_*$ and $\widetilde{\mathcal{G}}_*$:

LEMMA 3.6. *The map* $\iota(\mathbf{g}) = (p_j(\mathbf{g}))_{j \geq 1}$ *from* $\mathcal{G}_*$ *to* $\widetilde{\mathcal{G}}_*$ *is bijective.*

PROOF. The fact that ι is injective is a consequence of Lemma 3.5 part (iii). It remains to prove that the map ι is surjective. Let $y = (y_j) \in \widetilde{\mathcal{G}}_*$. One can represent the y_j's by rooted networks $(G_j, o) = (V_j, \omega_j, o)$ such that $V_j \subset V_{j+1}$. Set $V := \bigcup_{j \geq 1} V_j$. By adding isolated points, one can view (G_j, o) as the connected component at the root of the network $\hat{G}_j = (V, \omega_j)$, where $\omega_j(u, v) = 0$ whenever either u or v (or both) belong to $V \setminus V_j$. Moreover, one has that $\hat{G}_i = (\hat{G}_j)_{\theta_i}$ for all $i < j$. This sequence of networks is monotone in the sense of (31).

For fixed $u, v \in V$, and $j \in \mathbb{N}$, if $\omega_j(u, v) \neq 0$ then the degree of u and v is bounded by 2^{2j} in any network $\hat{G}_k$, $k \geq j$ and, therefore, $\omega_k(u, v) = \omega_{j+1}(u, v)$ for all $k \geq j + 1$. In particular, for all $u, v \in V$ the limit

$$\omega(u, v) = \lim_{j \to \infty} \omega_j(u, v)$$

exists and is finite. The same argument shows that for any $u \in V$, $\lim_{j \to \infty} \deg_{\hat{G}_j}(u)$ exists and equals

$$\sum_{v \in V} |\omega(u, v)|^2 < \infty.$$

To prove surjectivity of the map ι, it suffices to take the network $G = (V, \omega)$, and observe that it satisfies $G_{\theta_j} = \hat{G}_j$ for all $j \in \mathbb{N}$. $\square$

With a slight abuse of notation, we will from now on write $\mathcal{G}_*$ in place of $\widetilde{\mathcal{G}}_*$. The projective topology on $\mathcal{G}_*$ is the topology induced by the metric

$$d_{\mathrm{proj}}(\mathbf{g}, \mathbf{g}') = \sum_{j \geq 1} 2^{-j} d_{\mathrm{loc}}(\mathbf{g}_{\theta_j}, \mathbf{g}'_{\theta_j}).$$

The metric space $(\mathcal{G}_*, d_{\mathrm{proj}})$ is complete and separable. Also, $\mathbf{g}_n \overset{\mathrm{proj}}{\to} \mathbf{g}$, that is, $d_{\mathrm{proj}}(\mathbf{g}_n, \mathbf{g}) \to 0$, if and only if for any $\theta > 0$, $(\mathbf{g}_n)_\theta \overset{\mathrm{loc}}{\to} \mathbf{g}_\theta$. The *projective weak topology* is the weak topology on $\mathcal{P}(\mathcal{G}_*)$ associated to continuous functions on $(\mathcal{G}_*, d_{\mathrm{proj}})$. We denote the associated convergence by $\overset{\mathrm{proj}}{\rightsquigarrow}$. Notice that $\rho_n \overset{\mathrm{proj}}{\rightsquigarrow} \rho$ if and only if for any $\theta > 0$, $(\rho_n)_\theta \overset{\mathrm{loc}}{\rightsquigarrow} \rho_\theta$. The topology generated by d_{proj} is coarser than the topology generated by d_{loc}, and the weak topology associated to $\overset{\mathrm{proj}}{\rightsquigarrow}$ is coarser than the weak topology associated to $\overset{\mathrm{loc}}{\rightsquigarrow}$.

EXAMPLE 3.7. Consider the star shaped rooted network $(G_n, 1) = (V_n, \omega_n, 1)$ where $V_n = \{1, \ldots, n\}$, with $\omega_n(u, v) = \omega_n(v, u) = 1$, if $u = 1$ and $v \neq 1$, and $\omega(u, v) = 0$ otherwise. Let $\mathbf{g}_n$ denote the associated equivalence class in $\mathcal{G}_*$. Then $\mathbf{g}_n$ does not converge in $(\mathcal{G}_*, d_{\mathrm{loc}})$ because of the diverging degree at the root. However, in $(\mathcal{G}_*, d_{\mathrm{proj}})$, $\mathbf{g}_n \overset{\mathrm{proj}}{\to} \mathbf{g}_\varnothing$ where $\mathbf{g}_\varnothing$ is the empty network. Moreover, $U(G_n)$ does not converge in $\mathcal{P}(\mathcal{G}_*)$ for $\overset{\mathrm{loc}}{\rightsquigarrow}$ however $U(G_n) \overset{\mathrm{proj}}{\rightsquigarrow} \delta_{\mathbf{g}_\varnothing}$.

LEMMA 3.8. (i) *$\mathcal{G}_*$ is compact for the projective topology.*

(ii) *$\mathcal{P}_s(\mathcal{G}_*)$ is compact for the projective weak topology.*

PROOF. Statement (i) is a consequence of Tychonoff theorem and Lemma 3.4(ii). It implies that $\mathcal{P}(\mathcal{G}_*)$ is compact for projective weak topology. Hence, to prove statement (ii), it is sufficient to check that $\mathcal{P}_s(\mathcal{G}_*)$ is closed. Assume that $\rho_n \in \mathcal{P}_s(\mathcal{G}_*)$ and $\rho_n \overset{\mathrm{proj}}{\rightsquigarrow} \rho$. Then for any $\theta > 0$, $(\rho_n)_\theta \in \mathcal{P}_s(\mathcal{G}_*)$ and $(\rho_n)_\theta \overset{\mathrm{loc}}{\rightsquigarrow} \rho_\theta$. By Lemma 3.4(i), we deduce that $\rho_\theta \in \mathcal{P}_s(\mathcal{G}_*)$. However, as $\theta \to 0$, using Lemma 3.5, we find $\rho_\theta \overset{\mathrm{loc}}{\rightsquigarrow} \rho$. By appealing to Lemma 3.4(i) again, we get $\rho \in \mathcal{P}_s(\mathcal{G}_*)$. $\square$

3.6. *Large deviations for the network G_n.* For a rooted network (G, o), $G = (V_G, \omega_G)$, define the functions

$$\psi(G, o) = |\omega_G(o)|^\alpha \quad \text{and} \quad \phi(G, o) = \frac{1}{2} \sum_{v \in V_G \backslash o} |\omega_G(o, v)|^\alpha. \tag{33}$$

Since these functions are invariant under rooted isomorphisms, one can take them as functions on $\mathcal{G}_*$. Then, if $\rho \in \mathcal{P}(\mathcal{G}_*)$ we write $\mathbb{E}_\rho \psi$, and $\mathbb{E}_\rho \phi$ to denote the corresponding expectations. We remark that for any $\theta > 0$, the restriction of ϕ, ψ to $(\mathcal{G}_*^\theta, d_{\mathrm{loc}})$ gives two bounded continuous functions. Therefore, as functions on $(\mathcal{G}_*, d_{\mathrm{proj}})$, ϕ and ψ are lower semicontinuous.

We now come back to the random matrix $C = C(n)$ defined in (11). For integer $n \geq 1$, consider the associated network

$$G_n = (V_n, \omega_n) \qquad \text{with } V_n = \{1, \dots, n\} \text{ and } \omega_n(i,j) = C_{ij}. \tag{34}$$

From the first Borel–Cantelli lemma, almost surely the matrix C has no nonzero entry for n large enough. Therefore, almost surely, $U(G_n) \overset{\mathrm{loc}}{\rightsquigarrow} \delta_{\mathbf{g}_\varnothing}$, the Dirac mass at the empty network.

For ease of notation, we define the random probability measure

$$\rho_n = U(G_n).$$

Notice that, by definition one has

$$\mathbb{E}_{\rho_n} \psi = \frac{1}{n^{1+\alpha/2}} \sum_{i=1}^{n} |X_{ii}|^\alpha \mathbf{1}\big(\varepsilon(n)\sqrt{n} \leq |X_{ii}| \leq \varepsilon(n)^{-1}\sqrt{n}\big) \tag{35}$$

and

$$\mathbb{E}_{\rho_n} \phi = \frac{1}{n^{1+\alpha/2}} \sum_{1 \leq i < j \leq n} |X_{ij}|^\alpha \mathbf{1}\big(\varepsilon(n)\sqrt{n} \leq |X_{ij}| \leq \varepsilon(n)^{-1}\sqrt{n}\big). \tag{36}$$

The next proposition gives the large deviation principle for $\rho_n = U(G_n)$ for the projective weak topology.

PROPOSITION 3.9. *$U(G_n)$ satisfies an LDP on $\mathcal{P}(\mathcal{G}_*)$ equipped with the projective weak topology, with speed $n^{1+\alpha/2}$ and good rate function $I : \mathcal{P}(\mathcal{G}_*) \mapsto [0, \infty]$ defined by*

$$I(\rho) = \begin{cases} b\mathbb{E}_\rho \psi + a\mathbb{E}_\rho \phi, & \text{if } \rho \in \mathcal{P}_s(\mathcal{G}_*), \\ +\infty, & \text{if } \rho \notin \mathcal{P}_s(\mathcal{G}_*). \end{cases} \tag{37}$$

If a or b is equal to ∞, the above formula holds with the convention $\infty \times 0 = 0$.

PROOF. By construction, $\rho_n = U(G_n) \in \mathcal{P}_s(\mathcal{G}_*)$; see Example 3.3. Since $\mathcal{P}_s(\mathcal{G}_*)$ is closed (see Lemma 3.4), it is sufficient to establish the LDP on the space $\mathcal{P}_s(\mathcal{G}_*)$ with good rate function $I(\rho) = b\mathbb{E}_\rho \psi + a\mathbb{E}_\rho \phi$, $\rho \in \mathcal{P}_s(\mathcal{G}_*)$.

Let $B_{\mathrm{proj}}(\rho, \delta)$ [resp., $B_{\mathrm{loc}}(\rho, \delta)$] denote the closed ball with radius $\delta > 0$ and center $\rho \in \mathcal{P}_s(\mathcal{G}_*)$ for the Lévy metric associated to the projective weak topology (resp., local weak topology).

Upper bound. By Lemma 3.8(ii), $\mathcal{P}_s(\mathcal{G}_*)$ is compact. Hence, it is sufficient to prove (see, e.g., [14]) that for any $\rho \in \mathcal{P}_s(\mathcal{G}_*)$

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n^{1+\alpha/2}}\log\mathbb{P}\big(\rho_n \in B_{\mathrm{proj}}(\rho,\delta)\big) \le -b\mathbb{E}_\rho\psi - a\mathbb{E}_\rho\phi. \tag{38}$$

Assume first that $\mathbb{E}_\rho\psi$ and $\mathbb{E}_\rho\phi$ are finite. From standard properties of weak convergence, and the fact that ϕ, ψ are lower semicontinuous on $(\mathcal{G}_*, d_{\mathrm{proj}})$, it follows that the maps $\mu \mapsto \mathbb{E}_\mu\psi$ and $\mu \mapsto \mathbb{E}_\mu\phi$ are lower semicontinuous on $\mathcal{P}_s(\mathcal{G}_*)$ w.r.t. the projective weak topology. Hence, we have for some continuous function $h(\cdot)$ with $h(0) = 0$,

$$\mathbb{P}\big(\rho_n \in B_{\mathrm{proj}}(\rho,\delta)\big) \le \mathbb{P}\big(\mathbb{E}_{\rho_n}\psi \ge \mathbb{E}_\rho\psi - h(\delta); \mathbb{E}_{\rho_n}\phi \ge \mathbb{E}_\rho\phi - h(\delta)\big).$$

Since (35) and (36) are independent random variables,

$$\begin{aligned}&\mathbb{P}\big(\rho_n \in B_{\mathrm{proj}}(\rho,\delta)\big)\\ &\qquad \le \mathbb{P}\big(\mathbb{E}_{\rho_n}\psi \ge \mathbb{E}_\rho\psi - h(\delta)\big)\mathbb{P}\big(\mathbb{E}_{\rho_n}\phi \ge \mathbb{E}_\rho\phi - h(\delta)\big).\end{aligned} \tag{39}$$

To prove the part of the bound involving ϕ, one may assume $\mathbb{E}_\rho\phi > 0$. Take δ small enough, so that $s := \mathbb{E}_\rho\phi - h(\delta) > 0$. From (36), using Markov's inequality, for any $a_1 > 0$,

$$\mathbb{P}(\mathbb{E}_{\rho_n}\phi \ge s) \le e^{-a_1 n^{1+\alpha/2}s}\big(\mathbb{E}\exp\big(a_1|X_{12}|^\alpha \mathbf{1}_{\varepsilon(n)\sqrt{n}\le|X_{12}|\le\varepsilon(n)^{-1}\sqrt{n}}\big)\big)^{n(n-1)/2}.$$

Take $0 < a_1 < a$. By assumption, there exists $a_2 \in (a_1, a)$, such that for all $t > 0$ large enough,

$$\mathbb{P}\big(|X_{12}| \ge t\big) \le \exp\big(-a_2 t^\alpha\big).$$

Using (16), one deduces that

$$\begin{aligned}&\mathbb{E}\exp\big(a_1|X_{12}|^\alpha \mathbf{1}_{\varepsilon(n)\sqrt{n}\le|X_{12}|\le\varepsilon(n)^{-1}\sqrt{n}}\big)\\ &\qquad \le 1 + e^{-(a_2-a_1)\varepsilon(n)^\alpha n^{\alpha/2}} + \alpha a_1 \int_{\varepsilon(n)\sqrt{n}}^{\varepsilon(n)^{-1}\sqrt{n}} x^{\alpha-1}e^{-(a_2-a_1)x^\alpha}\,dx\\ &\qquad \le 1 + \frac{a_2}{a_2-a_1}e^{-(a_2-a_1)\varepsilon(n)^\alpha n^{\alpha/2}}.\end{aligned}$$

Therefore,

$$\mathbb{P}(\mathbb{E}_{\rho_n}\phi \ge s) \le \exp\left(-a_1 n^{1+\alpha/2}s + \frac{a_2}{2(a_2-a_1)}n^2 e^{-(a_2-a_1)\varepsilon(n)^\alpha n^{\alpha/2}}\right).$$

We have thus proved that for δ small enough

$$\limsup_{n\to\infty}\frac{1}{n^{1+\alpha/2}}\log\mathbb{P}(\mathbb{E}_{\rho_n}\phi \ge s) \le -a_1\big(\mathbb{E}_\rho\phi - h(\delta)\big).$$

Since the above inequality is true for any $a_1 < a$, it also holds for $a_1 = a$. Similarly, one has

$$\limsup_{n\to\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}(\mathbb{E}_{\rho_n}\psi \geq s) \leq -b\big(\mathbb{E}_\rho\psi - h(\delta)\big).$$

From (39), it follows that (38) holds under the assumption that both $\mathbb{E}_\rho\psi, \mathbb{E}_\rho\phi$ are finite.

If, for example, $\mathbb{E}_\rho\psi$ is infinite, then the above argument can be repeated, replacing $\mathbb{E}_\rho\psi$ by a large number K, and then letting $K \to \infty$ at the end. The same reasoning applies to the case where $\mathbb{E}_\rho\phi = \infty$. Similarly, if, for example, $b = \infty$ and $\mathbb{E}_\rho\psi > 0$, one can replace b above by a large number K and then let $K \to \infty$ at the end. The same applies to the case $a = \infty$ and $\mathbb{E}_\rho\phi > 0$. In particular, in all these cases one has that the left-hand side of (38) is $-\infty$. $\square$

Lower bound. It is sufficient to prove that for any $\rho \in \mathcal{P}_s(\mathcal{G}_*)$ and any $\delta > 0$,

$$\liminf_{n\to\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}\big(\rho_n \in B_{\mathrm{proj}}(\rho,\delta)\big) \geq -b\mathbb{E}_\rho\psi - a\mathbb{E}_\rho\phi. \tag{40}$$

In order to prove (40), we may assume without loss of generality that $I(\rho) = b\mathbb{E}_\rho\psi + a\mathbb{E}_\rho\phi < \infty$. By monotonicity (31), one has that

$$\lim_{j\to\infty} I(\rho_{\theta_j}) = I(\rho).$$

Therefore, since the projective topology is generated from the product topology on $\prod_{j\geq 1} \mathcal{G}_*^{\theta_j}$, it is sufficient to prove (40) for all $\rho \in \mathcal{P}_s(\mathcal{G}_*^\theta)$, for all $0 < \theta < 1$. Finally, since the local weak topology is finer than the projective weak topology, it is enough to prove that for any $0 < \theta < 1$, $\rho \in \mathcal{P}_s(\mathcal{G}_*^\theta)$ and $\delta > 0$,

$$\liminf_{n\to\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}\big(\rho_n \in B_{\mathrm{loc}}(\rho,\delta)\big) \geq -b\mathbb{E}_\rho\psi - a\mathbb{E}_\rho\phi. \tag{41}$$

Let us start with some simple consequences of Assumption 1. From (4), there exists a positive sequence η_n converging to 0 such that, for any $s \geq \varepsilon(n) = 1/\log n$,

$$e^{-(a+\eta_n)s^\alpha n^{\alpha/2}} \leq \mathbb{P}\big(|X_{12}| \geq s\sqrt{n}\big) \leq e^{-(a-\eta_n)s^\alpha n^{\alpha/2}}. \tag{42}$$

In particular, if $s \geq \varepsilon(n)$, then for any $\gamma > 0$, for all n large enough,

$$\mathbb{P}\big(|X_{12}| \in [s, s+\gamma)\sqrt{n}\big) \geq \tfrac{1}{2} e^{-(a+\eta_n)s^\alpha n^{\alpha/2}}.$$

Therefore, using (5), one finds that there exists a sequence $a_n \to a$ such that for every $\gamma > 0$, for all n large enough, for every $z \in \mathbb{C}$, with $|z| \geq \varepsilon(n)$, $z/|z| \in S_a$,

$$\mathbb{P}\big(X_{12}/\sqrt{n} \in B_{\mathbb{C}}(z,\gamma)\big) \geq e^{-a_n|z|^\alpha n^{\alpha/2}}, \tag{43}$$

where S_a denotes the compact support of the measure $\vartheta_a \in \mathcal{P}(\mathbb{S}^1)$ associated to X_{12}, and $B_{\mathbb{C}}(z,\gamma)$ is the Euclidean ball in $\mathbb{C}$, with center z and radius $\gamma > 0$.

Similarly, there exists a sequence $b_n \to b$ such that for every $\gamma > 0$, for all n large enough, for every $x \in \mathbb{R}$, with $|x| \geq \varepsilon(n)$, $x/|x| \in S_b$,

$$\mathbb{P}\big(X_{11}/\sqrt{n} \in B_{\mathbb{R}}(x,\gamma)\big) \geq e^{-b_n|x|^\alpha n^{\alpha/2}}. \tag{44}$$

We remark that (43) and (44) are the only places where the assumption (5) is used in this work.

Since $\rho \in \mathcal{P}_s(\mathcal{G}_*^\theta)$, there exists a sequence of matrices $H_n \in \mathcal{A}_n$, such that the associated network as in (27) is in $\mathcal{G}^\theta$ and such that $U(H_n) \overset{\text{loc}}{\rightsquigarrow} \rho$. In particular, for n sufficiently large one has

$$U(H_n) \in B_{\text{loc}}(\rho, \delta/2).$$

From Lemma 3.10, there exists $\gamma = \gamma(\delta,\theta) > 0$ such that if $|\omega_{G_n}(i) - H_n(i,i)| \leq \gamma$ and $|\omega_{G_n}(i,j) - H_n(i,j)| \leq \gamma$ for all $1 \leq i \leq j \leq n$, then $\rho_n = U(G_n) \in B_{\text{loc}}(U(H_n), \delta/2)$. Then, by the triangle inequality, for all n large enough,

$$\begin{aligned}
&\mathbb{P}\big(\rho_n \in B_{\text{loc}}(\rho,\delta)\big)\\
&\quad \geq \mathbb{P}\big(\rho_n \in B_{\text{loc}}\big(U(H_n),\delta/2\big)\big)\\
&\quad \geq \mathbb{P}\Big(\max_{1\leq i\leq n}\big|\omega_{G_n}(i) - H_n(i,i)\big| \leq \gamma, \max_{1\leq i<j\leq n}\big|\omega_{G_n}(i,j) - H_n(i,j)\big| \leq \gamma\Big).
\end{aligned}$$

Independence of the weights $\omega_{G_n}(i,j) = C_{i,j}$, $1 \leq i \leq j \leq n$ then gives

$$\begin{aligned}
&\mathbb{P}\big(\rho_n \in B_{\text{loc}}(\rho,\delta)\big)\\
&\quad \geq \prod_{i=1}^n \mathbb{P}\big(\big|C_{ii} - H_n(i,i)\big| \leq \gamma\big) \prod_{1\leq i<j\leq n} \mathbb{P}\big(\big|C_{ij} - H_n(i,j)\big| \leq \gamma\big).
\end{aligned}$$

Notice that whenever $H_n(i,j) \neq 0$ one has $|H_n(i,j)| \geq \theta$, and thus using (42) and (44) one has for all $i = 1,\dots,n$:

$$\begin{aligned}
&\mathbb{P}\big(\big|C_{ii} - H_n(i,i)\big| \leq \gamma\big)\\
&\quad \geq e^{-b_n n^{\alpha/2}|H_n(i,i)|^\alpha}\big(\mathbf{1}\big(\big|H_n(i,i)\big| > 0\big)\\
&\qquad\qquad + \big(1 - e^{-c\varepsilon(n)^\alpha n^{\alpha/2}}\big)\mathbf{1}\big(\big|H_n(i,i)\big| = 0\big)\big)\\
&\quad \geq e^{-b_n n^{\alpha/2}|H_n(i,i)|^\alpha}\big(1 - e^{-c\varepsilon(n)^\alpha n^{\alpha/2}}\big),
\end{aligned}$$

where the constant c satisfies $c \geq b/2 > 0$. Similarly, using (43), for all $i \leq j$ and for some $c \geq a/2 > 0$:

$$\mathbb{P}\big(\big|C_{ij} - H_n(i,j)\big| \leq \gamma\big) \geq e^{-a_n n^{\alpha/2}|H_n(i,j)|^\alpha}\big(1 - e^{-c\varepsilon(n)^\alpha n^{\alpha/2}}\big).$$

Observe that

$$\frac{1}{n}\sum_{1\leq i\leq n}\big|H_n(i,i)\big|^\alpha = \mathbb{E}_{U(H_n)}\psi, \qquad \frac{1}{n}\sum_{1\leq i<j\leq n}\big|H_n(i,j)\big|^\alpha = \mathbb{E}_{U(H_n)}\phi.$$

Summarizing, using $(1 - e^{-c\varepsilon(n)^{\alpha} n^{\alpha/2}})^{n^2} \geq 1/2$ for n large enough, one finds

$$\mathbb{P}\big(\rho_n \in B_{\mathrm{loc}}(\rho, \delta)\big) \geq \tfrac{1}{2} e^{-b_n n^{1+\alpha/2} \mathbb{E}_{U(H_n)} \psi} e^{-a_n n^{1+\alpha/2} \mathbb{E}_{U(H_n)} \phi}. \tag{45}$$

Since ψ and ϕ are continuous and bounded on $\mathcal{G}_*^{\theta}$, one has $\mathbb{E}_{U(H_n)} \psi \to \mathbb{E}_{\rho} \psi$ and $\mathbb{E}_{U(H_n)} \phi \to \mathbb{E}_{\rho} \phi$, as $n \to \infty$. Moreover, $a_n \to a$ and $b_n \to b$. Therefore, (45) implies the desired bound (41). This concludes the proof of the lower bound. $\square$

The next lemma was used in the proof of the lower bound of Proposition 3.9. While the estimate is somewhat rough, it is crucial that it is uniform in the cardinality n of the vertex set.

LEMMA 3.10. *Let $0 < \theta < 1$ and $\delta > 0$. There exists $\gamma = \gamma(\delta, \theta) > 0$ such that for any integer $n \geq 1$, for any networks $G \in \mathcal{G}$, $H \in \mathcal{G}_\theta$ with common vertex set $V = \{1, \dots, n\}$ such that*

$$\max_{(u,v) \in V^2} \big|\omega_G(u, v) - \omega_H(u, v)\big| \leq \gamma, \tag{46}$$

then

$$\max_{u \in V} d_{\mathrm{loc}}\big((G(u), u), (H(u), u)\big) \leq \delta. \tag{47}$$

In particular,

$$U(G) \in B_{\mathrm{loc}}\big(U(H), \delta\big).$$

PROOF. Each edge of H has a weight bounded above by θ^{-1}. This implies that in H each path whose total length is bounded by $t > 0$, contains at most t^2/θ^2 edges. Moreover, H has at most θ^{-4} outgoing edges from any vertex. Hence, H has at most $m = \theta^{-4t^2/\theta^2}$ vertices at distance less than t from any given vertex. Fix the root $u \in V$ and $t > 0$. Therefore, there must exist $t_0 > 0$ such that $t/2 < t_0 < t$, and an interval $I = [t_0 - t/(8m), t_0 + t/(8m)]$, such that there is no vertex within distance $s \in I$ from u in H.

If $e_1, \dots, e_k$ are the edges on a path in H, then provided that $0 < \gamma < \theta/2$, one has

$$\begin{aligned} &\left[\left(\sum_{i=1}^{k} |\omega_H(e_i)|^{-2}\right)^{1/2} - \left(\sum_{i=1}^{k} |\omega_G(e_i)|^{-2}\right)^{1/2}\right]^2 \\ &\qquad \leq \sum_{i=1}^{k} \big(|\omega_H(e_i)|^{-1} - |\omega_G(e_i)|^{-1}\big)^2 \leq \frac{4\gamma^2 k}{\theta^4}. \end{aligned}$$

The first inequality follows from the convexity of $[0, \infty)^2 \ni (x, y) \mapsto (\sqrt{x} - \sqrt{y})^2$, which yields $((\sum_i u_i)^{1/2} - (\sum_i v_i)^{1/2})^2 \leq \sum_i (u_i^{1/2} - v_i^{1/2})^2$, for any $u, v \in \mathbb{R}_+^k$.

The second inequality follows from $|\omega_H(e_i)| \geq \theta$ and the assumption (46). In the worst possible case, one can take $k = t^2/\theta^2$ for the number of edges at distance t_0 from u. Together with the previous observation, this shows that if $2\gamma\sqrt{k}/\theta^2 \leq t/(8m)$, that is, $\gamma \leq \theta^3/(16m)$, then the neighborhood of u consisting of vertices within distance t_0 in G and in H have the same vertex set. From the definition of d_{loc}, this choice of γ in (46) implies that

$$d_{\mathrm{loc}}\big(\big(G(u), u\big), \big(H(u), u\big)\big) \leq \frac{1}{1 + \gamma^{-1} \wedge t_0} \leq \frac{2}{t}.$$

Thus, taking $t = 2/\delta$, one has (47), as soon as, for example, $\gamma \leq \theta^3/(16m) = \theta^{3+16/(\delta^2\theta^2)}/16$. From the definition of the Lévy distance, it immediately follows that $U(G) \in B_{\mathrm{loc}}(U(H), \delta)$. $\square$

REMARK 3.11. In the proof of Proposition 3.9, we have not appealed to general results, such as Dawson–Gärtner's theorem, that are available for projective topologies (see, e.g., [14], Section 4.6). We have, however, crucially used the compactness of $\mathcal{P}_s(\mathcal{G}^*)$ for the projective weak topology. It is not hard to check that the rate function $I(\rho)$ in (37) is not good for the weak topology (level sets are not compact).

3.7. *Spectral measure.* For a network $G = (V, \omega) \in \mathcal{G}^\theta$, we may define the bounded linear operator T on the Hilbert space $\ell^2(V)$ by

$$Te_v = \sum_{u \in V} \omega(u, v) e_u \tag{48}$$

for any $v \in V$, where $\{e_u, u \in V\}$ denotes the canonical orthonormal basis of $\ell^2(V)$. T is bounded since

$$\|Te_v\|_2^2 = \sum_{u \in V} \big|\omega(v, u)\big|^2 = \deg(v) \leq \theta^{-2}. \tag{49}$$

Also, since G is Hermitian, T is self-adjoint. We may thus define the spectral measure at vector e_v, see, e.g., [20], as the unique probability measure μ_T^v on $\mathbb{R}$ such that for any integer $k \geq 1$,

$$\int x^k \, d\mu_T^v = \langle e_v, T^k e_v \rangle. \tag{50}$$

Notice that for rooted networks (G, o) with $G \in \mathcal{G}^\theta$, then the associated spectral measure μ_T^o is constant on the equivalence class of (G, o), so that μ_T^o can be defined as a measurable map from $\mathcal{G}_*^\theta$ to $\mathcal{P}(\mathbb{R})$. Thus, if $\rho \in \mathcal{P}(\mathcal{G}_*^\theta)$ for some $\theta > 0$, one can define the spectral measure of ρ as

$$\mu_\rho = \mathbb{E}_\rho \mu_T^o. \tag{51}$$

In particular, consider a Hermitian matrix $H_n \in \mathcal{H}_n(\mathbb{C})$, let $G_n = G(H_n)$ be the associated network as in (27), and let $\rho_n = U(G_n)$. Then, if $(\psi_1, \dots, \psi_n)$ is an orthonormal basis of eigenvectors of H_n with associated eigenvalues $(\lambda_1, \dots, \lambda_n)$, by the spectral theorem, for any $v \in \{1, \dots, n\}$,

$$\mu_{H_n}^v = \sum_{i=1}^n |\langle \psi_i, e_v \rangle|^2 \delta_{\lambda_i},$$

where $\mu_{H_n}^v$ stands for the spectral measure at v; see (50). Moreover, the empirical distribution of the eigenvalues of H_n satisfies

$$\mu_{H_n} = \frac{1}{n}\sum_{i=1}^n \delta_{\lambda_i} = \frac{1}{n}\sum_{v=1}^n \mu_{H_n}^v = \mu_{\rho_n}. \tag{52}$$

Hence, our definition of spectral measure for a sofic distribution coincides for finite networks with the empirical distribution of the eigenvalues.

We turn to the definition of μ_ρ for the case where $\rho \in \mathcal{P}(\mathcal{G}_*)$ but there is no $\theta > 0$ such that $\rho \in \mathcal{P}(\mathcal{G}_*^\theta)$. In this case, (51) allows one to define the spectral measures μ_{ρ_θ}, where the truncated network ρ_θ is defined as in Lemma 3.5. Next, we shall define the spectral measure μ_ρ as the limit of μ_{ρ_θ} as $\theta \to 0$, provided some extra assumptions are satisfied. More precisely, for a rooted network (G, o), $G \in \mathcal{G}$, and for $\beta > 0$, let

$$\xi_\beta(G, o) = \sum_{v \in V_G} |\omega_G(o, v)|^\beta. \tag{53}$$

Since ξ_β is constant on the equivalence class of (G, o), it can be seen as a function on $\mathcal{G}_*$. For $\beta > 0$, $\tau > 0$, define

$$\mathcal{P}_{s,\beta,\tau}(\mathcal{G}_*) = \{\rho \in \mathcal{P}_s(\mathcal{G}_*) : \mathbb{E}_\rho \xi_\beta < \tau\}.$$

Lemma 3.12 below is an extension to the weighted case of analogous statements in [10, 11], where spectral measures are defined for random rooted graphs (with no weights). The first result allows one to define the spectral measure μ_ρ of any $\rho \in \mathcal{P}_{s,\beta,\tau}(\mathcal{G}_*)$.

LEMMA 3.12. *Let $0 < \beta < 2$, $\tau > 1$ and $\rho \in \mathcal{P}_{s,\beta,\tau}(\mathcal{G}_*)$. Then the weak limit*

$$\mu_\rho := \lim_{\theta \to 0} \mu_{\rho_\theta}$$

exists in $\mathcal{P}(\mathbb{R})$.

PROOF. To prove the lemma, we are going to show that the sequence μ_{ρ_θ}, $\theta \to 0$, is Cauchy w.r.t. the metric (12).

By assumption, there exists a sequence G_n of networks on $\{1, \dots, n\}$ such that $\rho_n \overset{\text{loc}}{\rightsquigarrow} \rho$, where $\rho_n = U(G_n)$. Call T_n the associated Hermitian matrix. The empirical distribution of the eigenvalues of T_n satisfies $\mu_{T_n} = \mu_{\rho_n}$ by (52) applied to $H_n = T_n$.

The truncations $(\rho_n)_\theta$ and ρ_θ satisfy $(\rho_n)_\theta \overset{\text{loc}}{\rightsquigarrow} \rho_\theta$ by Lemma 3.5(ii). Moreover, for all $\theta > 0$,

$$\mu_{(\rho_n)_\theta} \rightsquigarrow \mu_{\rho_\theta}. \tag{54}$$

To prove (54), let T^θ denote the random bounded self-adjoint operator associated to ρ_θ via (48) and let T_n^θ be the matrices associated to $(\rho_n)_\theta$. One can realize these operators on a common Hilbert space $\ell^2(V)$. Since $(\rho_n)_\theta \overset{\text{loc}}{\rightsquigarrow} \rho_\theta$, from the Skorokhod representation theorem one can define a common probability space such that the associated networks converge locally almost surely, so that a.s. $T_n^\theta e_v \to T^\theta e_v$, in $\ell^2(V)$, for any $v \in V$. This implies the strong resolvent convergence; see, for example, [20], Theorem VIII.25(a), and in particular that for any $v \in V$, a.s.

$$\mu^v_{T_n^\theta} \rightsquigarrow \mu^v_{T^\theta}.$$

Then (54) follows by applying this to $v = o$ and taking expectation.

Let $T_n^\theta, \widetilde{T}_n^\theta$ be the matrices associated to $(G_n)_\theta$ and $(\widetilde{G}_n)_\theta$, respectively, where $(\widetilde{G}_n)_\theta$ is defined according to (29), and $(G_n)_\theta$ according to (30). From (14), using the triangle inequality, Lemmas B.1 and B.2,

$$d(\mu_{T_n^\theta}, \mu_{T_n}) \le \frac{1}{n}\operatorname{rank}(\widetilde{T}_n^\theta - T_n) + \left(\frac{1}{n}\operatorname{tr}(\widetilde{T}_n^\theta - T_n^\theta)^2\right)^{1/2}.$$

From the definition (29), one has

$$\frac{1}{n}\operatorname{rank}(\widetilde{T}_n^\theta - T_n) \le \frac{2}{n}\sum_{i=1}^n \mathbf{1}(\deg_{G_n}(i) \ge \theta^{-2} - 1) = 2\mathbb{P}_{\rho_n}(\deg_G(o) \ge \theta^{-2} - 1).$$

From (30), one finds

$$\begin{aligned}\frac{1}{n}\operatorname{tr}(\widetilde{T}_n^\theta - T_n^\theta)^2 &\le \frac{1}{n}\sum_{i,j=1}^n |\omega_{G_n}(i,j)|^2 \mathbf{1}(|\omega_{G_n}(i,j)| \le 2\theta)\mathbf{1}(\deg_{G_n}(i) \le \theta^{-2}) \\ &= \mathbb{E}_{\rho_n}\mathbf{1}(\deg_G(o) \le \theta^{-2})\sum_v |\omega_G(o,v)|^2 \mathbf{1}(|\omega_G(o,v)| \le 2\theta).\end{aligned}$$

Letting n go to infinity, using $\mu_{T_n^\theta} = \mu_{(\rho_n)_\theta}$, and (54), one has $d(\mu_{T_n^\theta}, \mu_{T_n^{\theta'}}) \to d(\mu_{\rho_\theta}, \mu_{\rho_{\theta'}})$. Therefore, by the triangle inequality and the dominated convergence theorem, for any $0 < \theta' < \theta < 1/\sqrt{2}$,

$$\begin{aligned}d(\mu_{\rho_\theta}, \mu_{\rho_{\theta'}}) &\le 4\mathbb{P}_\rho(\deg_G(o) \ge \theta^{-2}/2) \\ &\quad + 2\left(\mathbb{E}_\rho \mathbf{1}(\deg_G(o) \le \theta^{-2})\sum_v |\omega_G(o,v)|^2 \mathbf{1}(|\omega_G(o,v)| \le 2\theta)\right)^{1/2}.\end{aligned}$$

Notice that, for $\beta \in (0,2)$

$$\deg_G(o)^{\beta/2} = \Bigl(\sum_v |\omega_G(o,v)|^2\Bigr)^{\beta/2} \le \sum_v |\omega_G(o,v)|^\beta = \xi_\beta(G,o), \tag{55}$$

where we use that $\sum_{i=1}^k a_i^r \le (\sum_{i=1}^k a_i)^r$ for all $a_i \ge 0$, $r \ge 1$ and $k \in \mathbb{N}$. Moreover,

$$\sum_v |\omega_G(o,v)|^2 \mathbf{1}(|\omega_G(o,v)| \le \theta) \le \theta^{2-\beta}\xi_\beta(G,o).$$

Hence, from Markov's inequality,

$$d(\mu_{\rho_\theta}, \mu_{\rho_{\theta'}}) \le 4\theta^\beta \mathbb{E}_\rho \xi_\beta + 2\theta^{1-\beta/2}(\mathbb{E}_\rho \xi_\beta)^{1/2}. \tag{56}$$

By assumption $\mathbb{E}_\rho \xi_\beta$ is finite. Hence, the sequence μ_{ρ_θ} is Cauchy. □

LEMMA 3.13. *For any $\beta \in (0,2)$, $\tau > 0$, the map $\rho \mapsto \mu_\rho$ from $\mathcal{P}_{s,\beta,\tau}(\mathcal{G}_*)$ to $\mathcal{P}(\mathbb{R})$ is continuous for the projective weak topology.*

PROOF. For any $\theta > 0$, from (56),

$$d(\mu_{\rho_\theta}, \mu_\rho) \le c(\theta^\beta + \theta^{1-\beta/2}), \tag{57}$$

with a constant $c = c(\tau) > 0$. Hence, from the triangle inequality, if $\rho, \rho' \in \mathcal{P}_{s,\beta,\tau}(\mathcal{G}_*)$,

$$d(\mu_\rho, \mu_{\rho'}) \le 2c(\theta^\beta + \theta^{1-\beta/2}) + d(\mu_{\rho_\theta}, \mu_{\rho'_\theta}).$$

Consider a sequence ρ' such that $\rho' \overset{\text{proj}}{\rightsquigarrow} \rho$. If $\rho' \overset{\text{proj}}{\rightsquigarrow} \rho$ then $\rho'_\theta \overset{\text{loc}}{\rightsquigarrow} \rho_\theta$ and, therefore, with the same argument used in the proof of (54) above one finds

$$\mu_{\rho'_\theta} \rightsquigarrow \mu_{\rho_\theta}.$$

We deduce that

$$\limsup_{\rho' \overset{\text{proj}}{\rightsquigarrow} \rho} d(\mu_\rho, \mu_{\rho'}) \le 2c(\theta^\beta + \theta^{1-\beta/2}).$$

Since $\theta > 0$ is arbitrarily small, the statement of the lemma follows. □

3.8. *Large deviations for the empirical spectral measure μ_C.* We can apply the previous results to the empirical spectral measure μ_C, where $C = C(n)$ is the random matrix defined in (11). So far, we have defined μ_ρ for every $\rho \in \bigcup_{0<\beta<2}\bigcup_{\tau>1}\mathcal{P}_{s,\beta,\tau}(\mathcal{G}_*)$. If $\rho \in \mathcal{P}_s(\mathcal{G}_*)$ but $\rho \notin \bigcup_{0<\beta<2}\bigcup_{\tau>1}\mathcal{P}_{s,\beta,\tau}(\mathcal{G}_*)$, then we set

$$\mu_\rho = \delta_0.$$

PROPOSITION 3.14. *The empirical spectral measures μ_C satisfy an LDP on $\mathcal{P}(\mathbb{R})$ equipped with the weak topology, with speed $n^{1+\alpha/2}$ and good rate function Φ given by*

$$\Phi(\nu) = \inf\{I(\rho), \rho \in \mathcal{P}_s(\mathcal{G}_*) : \mu_\rho = \nu\}, \tag{58}$$

where $I(\rho)$ is the good rate function in Proposition 3.9.

PROOF. Recall that by (52) the network G_n in (34) satisfies $\rho_n = U(G_n)$ and

$$\mu_{\rho_n} = \mu_C.$$

Notice that if $c = (\frac{a}{2} \wedge b)$, then

$$I(\rho) \geq c\mathbb{E}_\rho \xi_\alpha, \tag{59}$$

where ξ_α is defined by (53). Hence, by Lemma 3.13, the map $\rho \mapsto \mu_\rho$ is continuous on the domain of $I(\rho)$. We would like to apply a contraction principle to get the LDP for μ_{ρ_n} from the LDP for ρ_n; see, for example, [14], Theorem 4.2.1(a). However, a little care is needed here because $\rho \mapsto \mu_\rho$ is continuous on the set $I(\cdot) < \infty$ only.

We start with the lower bound. Assume that B is an open set in $\mathcal{P}(\mathbb{R})$. For each $\tau > 0$, by Lemma 3.13, the function $f_\tau : \rho \mapsto \mu_\rho$ from $\mathcal{P}_{s,\alpha,\tau}(\mathcal{G}_*) \to \mathcal{P}(\mathbb{R})$ is continuous. Hence, $f_\tau^{-1}(B)$ is an open subset of $\mathcal{P}_{s,\alpha,\tau}(\mathcal{G}_*)$, and

$$\mathbb{P}(\mu_{\rho_n} \in B) \geq \mathbb{P}(\rho_n \in f_\tau^{-1}(B)).$$

From Proposition 3.9, it follows that

$$-\inf_{\rho \in \mathcal{P}_{s,\alpha,\tau}(\mathcal{G}_*) : \mu_\rho \in B} I(\rho) \leq \liminf_{n\to\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}(\mu_{\rho_n} \in B).$$

Using (59), one has for some $c > 0$:

$$-\inf_{\rho \in \mathcal{P}_s(\mathcal{G}_*) : \mu_\rho \in B} I(\rho) \leq (-c\tau) \vee \liminf_{n\to\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}(\mu_{\rho_n} \in B).$$

Letting τ tend to infinity, we obtain the desired lower bound:

$$-\inf_{\nu \in B} \Phi(\nu) \leq \limsup_{n\to\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}(\mu_{\rho_n} \in B).$$

To prove the upper bound, assume that B is a closed set in $\mathcal{P}(\mathbb{R})$. By Lemma 3.13, $f_\tau^{-1}(B)$ is a closed subset of $\mathcal{P}_{s,\alpha,\tau}(\mathcal{G}_*)$. Write

$$\mathbb{P}(\mu_{\rho_n} \in B) \leq \mathbb{P}(\mu_{\rho_n} \in B; \rho_n \in \mathcal{P}_{s,\alpha,\tau}(\mathcal{G}_*)) + \mathbb{P}(\rho_n \notin \mathcal{P}_{s,\alpha,\tau}(\mathcal{G}_*)).$$

Proposition 3.9 yields

$$\limsup_{n\to\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}(\mu_{\rho_n} \in B; \rho_n \in \mathcal{P}_{s,\alpha,\tau}(\mathcal{G}_*)) \leq -\inf_{\rho \in \mathcal{P}_{s,\alpha,\tau}(\mathcal{G}_*) : \mu_\rho \in B} I(\rho),$$

and, for some $c > 0$:

$$\limsup_{n\to\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}(\rho_n \notin \mathcal{P}_{s,\alpha,\tau}(\mathcal{G}_*)) \leq -c\tau.$$

We have checked that

$$\limsup_{n\to\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}(\mu_{\rho_n} \in B) \leq -\Big[(c\tau) \wedge \inf_{\mu\in B} \Phi(\mu)\Big].$$

Letting τ tend to infinity, we obtain the desired upper bound. The fact that Φ is a good rate function can be seen as in [14], Theorem 4.2.1(a), or, more directly, it follows from Lemma 3.15 below. $\square$

3.9. *Proof of Theorem* 1.1. Thanks to Proposition 2.1, all we have to show is that is that the sequence of measures $\mu_{\mathrm{sc}} \boxplus \mu_C$ satisfies a LDP in $\mathcal{P}(\mathbb{R})$ with speed $n^{1+\alpha/2}$, with the good rate function Φ defined in Proposition 3.14. Since the map $\nu \mapsto \mu_{\mathrm{sc}} \boxplus \nu$ is continuous in $\mathcal{P}(\mathbb{R})$, the above is an immediate consequence of Proposition 3.14 and the standard contraction principle. This completes the proof of Theorem 1.1.

3.10. *On the rate function* Φ. We turn to a proof of the properties of the rate function listed in Theorems 1.2 and 1.3.

LEMMA 3.15. *For any* $\beta \in (0, 2)$, $\tau > 1$, *for any* $\rho \in \mathcal{P}_{s,\beta,\tau}(\mathcal{G}_*)$, *one has*

$$\int |x|^\beta \, d\mu_\rho(x) \leq \mathbb{E}_\rho \xi_\beta. \tag{60}$$

PROOF. We use the following Schatten bound: for all $0 < p \leq 2$,

$$\int |x|^p \, d\mu_A(x) \leq \frac{1}{n} \sum_{k=1}^n \left(\sum_{j=1}^n |A_{kj}|^2 \right)^{p/2} \tag{61}$$

for every Hermitian matrix $A \in \mathcal{H}_n(\mathbb{C})$. For a proof, see Zhan [21], proof of Theorem 3.32. For $\rho \in \mathcal{P}_{s,\beta,\tau}(\mathcal{G}_*)$, there exists a sequence of matrices H_n such that $\rho_n = U(H_n) \overset{\text{loc}}{\rightsquigarrow} \rho$. Let T_n^θ be the Hermitian matrix associated to $(H_n)_\theta$, the truncated network. From (61) and (55), one has for all $\theta > 0$:

$$\int |x|^\beta \, d\mu_{T_n^\theta}(x) \leq \mathbb{E}_{\rho_n}\left[\left(\theta^{-2} \wedge \sum_v |\omega(o, v)|^2 \right)^{\beta/2} \right] \leq \mathbb{E}_{\rho_n}(\theta^{-\beta} \wedge \xi_\beta(G, o)).$$

For $\theta > 0$, the spectral measures $\mu_{T_n^\theta} = \mu_{(\rho_n)_\theta}$ have compact support uniformly in n. Thus, letting n go to infinity, from (54) one has

$$\int |x|^\beta \, d\mu_{\rho_\theta}(x) \leq \mathbb{E}_\rho \xi_\beta. \tag{62}$$

On the other hand, by definition of μ_ρ (see Lemma 3.12), one has $\mu_{\rho_\theta} \rightsquigarrow \mu_\rho$, $\theta \to 0$ and, therefore,

$$\int |x|^\beta \, d\mu_\rho(x) \le \liminf_{\theta \to 0} \int |x|^\beta \, d\mu_{\rho_\theta}(x).$$

This proves the claim (60). $\square$

PROOF OF THEOREM 1.2(A). The proof is an immediate consequence of Lemma 3.15. Indeed, from (59) and the definition of Φ, it suffices to show that for any $\tau > 1$, for any $\rho \in \mathcal{P}_{s,\alpha,\tau}(\mathcal{G}_*)$, one has

$$\int |x|^\alpha \, d\mu_\rho(x) \le \mathbb{E}_\rho \xi_\alpha. \tag{63}$$

This is the case $\alpha = \beta$ in (60). $\square$

PROOF OF THEOREM 1.2(B). For $x \in \mathbb{R}$, let $\mathbf{g}_x \in \mathcal{G}_*$ denote the network consisting of a single vertex o with weight $\omega(o,o) = x$. If $\nu \in \mathcal{P}(\mathbb{R})$, let $\rho \in \mathcal{P}(\mathcal{G}_*)$ denote the law $\rho = \int_\mathbb{R} \delta_{\mathbf{g}_x} \, d\nu(x)$. Notice that

$$\mathbb{E}_\rho \xi_\alpha = \int_\mathbb{R} |x|^\alpha \, d\nu(x) = m_\alpha(\nu).$$

Thus, we can assume $\mathbb{E}_\rho \xi_\alpha < \infty$, otherwise there is nothing to prove. Since we assume $\mathrm{supp}(\vartheta_b) = \{-1, +1\}$, one has that ρ is admissible sofic; see Example 3.1, and $\rho \in \mathcal{P}_{s,\alpha,\tau}(\mathcal{G}_*)$ for some $\tau > 1$. The spectral measure μ_ρ of ρ, defined as in Lemma 3.12 is easily seen to be $\mu_\rho = \nu$. Then $\Phi(\nu) \le I(\rho) = b\mathbb{E}_\rho \xi_\alpha = b m_\alpha(\nu)$. $\square$

PROOF OF THEOREM 1.2(C). Thanks to parts (a) and (b), all we need to prove is that

$$\Phi(\nu) \le \frac{a}{2} m_\alpha(\nu), \tag{64}$$

for all symmetric probabilities ν on $\mathbb{R}$.

For $z \in \mathbb{C}$, let $\hat{\mathbf{g}}_z \in \mathcal{G}_*$ denote the equivalence class of the two vertex network (V, ω, o), with $V = \{o, 1\}$, $\omega(o,1) = z$, $\omega(1,o) = \bar{z}$ and $\omega(o,o) = \omega(1,1) = 0$. Fix some $e^{i\varphi} \in S_a = \mathrm{supp}(\vartheta_a)$, let T be a nonnegative random variable with some distribution μ_+ on $[0,\infty)$, and let $\mu \in \mathcal{P}(\mathbb{C})$ denote the law of $Te^{i\varphi}$. The law

$$\rho = \frac{1}{2} \int_\mathbb{C} (\delta_{\hat{\mathbf{g}}_z} + \delta_{\hat{\mathbf{g}}_{\bar{z}}}) \, d\mu(z),$$

is sofic; see Example 3.2. A simple computation shows that the spectral measure of ρ satisfies $\mu_\rho = \mu_{\mathrm{sym}}$, where μ_{sym} denotes the symmetric probability on $\mathbb{R}$ such that

$$\int_\mathbb{R} f(x) \, d\mu_{\mathrm{sym}}(x) = \frac{1}{2} \int_0^\infty \bigl(f(x) + f(-x)\bigr) \, d\mu_+(x)$$

for all bounded measurable f.

To prove (64), let $\nu \in \mathcal{P}_{\mathrm{sym}}(\mathbb{R})$ and write μ_+ for the law of $|X|$ when X has law ν. Then $\nu = \mu_{\mathrm{sym}}$ and the associated ρ satisfies $\mu_\rho = \nu$. Therefore,

$$\Phi(\nu) \le I(\rho) = \frac{a}{2}\int_0^\infty x^\alpha \, d\mu_+(x) = \frac{a}{2} m_\alpha(\nu). \qquad \square$$

PROOF OF THEOREM 1.3(a). We proceed as in the proof of Theorem 1.3(b). Here, $S_b = \{+1\}$, and thus the law $\rho = \int_{\mathbb{R}} \delta_{\mathbf{g}_x} \, d\nu(x)$ that we used there is not necessarily admissible sofic. However, it is so if one assumes $\mathrm{supp}(\nu) \subset \mathbb{R}_+$. The rest of the argument applies with no modifications. $\square$

For the remaining statements, we use the following observation.

LEMMA 3.16. *If $\rho \in \mathcal{P}_{s,\beta,\tau}(\mathcal{G}_*)$ for some $\beta \in (1,2)$, $\tau > 1$, then*

$$\int_{\mathbb{R}} x \, d\mu_\rho(x) = \mathbb{E}_\rho \omega_G(o). \tag{65}$$

PROOF. By definition of the spectral measure μ_{ρ_θ}[see (50)], for every $\theta > 0$ one has

$$\int_{\mathbb{R}} x \, d\mu_{\rho_\theta}(x) = \mathbb{E}_{\rho_\theta} \omega_G(o) = \mathbb{E}_\rho \omega_{G_\theta}(o),$$

where G_θ is the truncation of G; see (30). The weights $\omega_{G_\theta}(o)$ satisfy $|\omega_{G_\theta}(o)| \le |\omega_G(o)|$ and, since $\beta > 1$, $\mathbb{E}_\rho |\omega_G(o)| \le (\mathbb{E}_\rho \xi_\beta)^{1/\beta} < \tau^{1/\beta}$. Thus, by the dominated convergence theorem,

$$\lim_{\theta \to 0} \int_{\mathbb{R}} x \, d\mu_{\rho_\theta}(x) = \mathbb{E}_\rho \omega_G(o).$$

From (62), and the fact that $\beta > 1$, we know that the identity map $x \mapsto x$ is uniformly integrable for $(\mu_{\rho_\theta})_{\theta > 0}$. Therefore, by definition of μ_ρ (see Lemma 3.12), the limit above also equals $\int_{\mathbb{R}} x \, d\mu_\rho(x)$. $\square$

PROOF OF THEOREM 1.3(b). In view of the bound (64), it suffices to show that if $\rho \in \mathcal{P}_s(\mathcal{G}_*)$ with $\mu_\rho = \nu$, then

$$\frac{a}{2} \int |x|^\alpha \, d\mu_\rho(x) \le I(\rho). \tag{66}$$

Thanks to (59), one may assume that $\rho \in \mathcal{P}_{s,\alpha,\tau}(\mathcal{G}_*)$ for some $\tau > 1$. Moreover, by (59) and (63), we know that (66) holds if $b \ge a/2$. If $b < a/2$, we proceed as follows. Since $\alpha > 1$ here, we may apply Lemma 3.16, and obtain that

$$0 = \int_{\mathbb{R}} x \, d\nu(x) = \mathbb{E}_\rho \omega_G(o),$$

where we use the symmetry assumption on ν. Since $S_b = \{+1\}$, one has that $\omega_G(o) \geq 0$ and, therefore, $\omega_G(o) = 0$ ρ-a.s. In conclusion, $I(\rho) = a\mathbb{E}_\rho \phi = \frac{a}{2}\mathbb{E}_\rho \xi_\alpha$, and the claim (66) follows from (60). $\square$

PROOF OF THEOREM 1.3(c). Suppose that $I(\rho) < \infty$. Then by (59), one has $\rho \in \mathcal{P}_{s,\alpha,\tau}(\mathcal{G}_*)$ for some $\tau > 1$. Since $\alpha > 1$, Lemma 3.16 yields $\int_{\mathbb{R}} x\, d\nu(x) = \mathbb{E}_\rho \omega_G(o)$ which, together with the assumption $\int_{\mathbb{R}} x\, d\nu(x) < 0$, implies

$$\mathbb{E}_\rho \omega_G(o) < 0.$$

However, $S_b = \{+1\}$ implies that $\mathbb{E}_\rho \omega_G(o) \geq 0$, a contradiction. Thus, $I(\rho) = +\infty$, for all $\rho \in \mathcal{P}_s(\mathcal{G}_*)$ such that $\mu_\rho = \nu$. $\square$

APPENDIX A: UNIFORM ASYMPTOTIC FREENESS

A.1. Proof of Theorem 2.6. Recall the definition (13) of the function $g_\mu : \mathbb{C}_+ \mapsto \mathbb{C}_+$, for a given $\mu \in \mathcal{P}(\mathbb{R})$. Theorem 2.6 is a consequence of the following result.

THEOREM A.1 (Uniform bound in subordination formula). *Let $Y = (Y_{ij})_{1\leq i,j\leq n} \in \mathcal{H}_n(\mathbb{C})$ be a Wigner random matrix with* $\mathrm{Var}(Y_{12}) = 1$, $\mathbb{E}|Y_{12}|^3 < \infty$ *and $\mathbb{E}|Y_{11}|^2 < \infty$. There exists a universal constant $c > 0$, such that for any integer $n \geq 1$, any $M \in \mathcal{H}_n(\mathbb{C})$, any $z \in \mathbb{C}_+$, $\mathfrak{Im}(z) \geq 1$,*

$$\left|\overline{g}(z) - g_{\mu_M}\big(z + \overline{g}(z)\big)\right| \leq c\frac{(\mathbb{E}|Y_{11}|^2)^{1/2} + \mathbb{E}|Y_{12}|^3}{n^{1/2}},$$

where $\overline{g}(z) = \mathbb{E}g_{\mu_{Y/\sqrt{n}+M}}(z)$.

Theorem A.1 is a small generalization of Pastur and Shcherbina [19], Theorem 18.3.1: the main difference here is that we do not assume that the real and imaginary parts of Y_{ij} are independent. We also allow the mean of the entries to be nonzero. Note that the rate $1/\sqrt{n}$ in Theorem A.1 is not necessarily optimal with stronger assumptions; see, for example, [12], equation (3.8). We postpone the proof of Theorem A.1 to the next subsection. We first check that it implies Theorem 2.6. This is done by a simple contraction argument. For $z \in \mathbb{C}_+$, we define the $\mathbb{C}_+ \to \mathbb{C}_+$ map,

$$\phi_z : h \mapsto g_{\mu_M}(z + h). \tag{67}$$

It is Lipschitz with constant $1/\mathfrak{Im}(z)^2$. In particular, if $\mathfrak{Im}(z) \geq 2$, ϕ_z is a contraction with Lipschitz constant $1/4$. Now, it is well known that if $\mu = \mu_M \boxplus \mu_{\mathrm{sc}}$, we have for all $z \in \mathbb{C}_+$ the subordination formula,

$$g_\mu(z) = g_{\mu_M}\big(z + g_\mu(z)\big) = \phi_z\big(g_\mu(z)\big),$$

see Biane [8]. In particular, if for some probability measure $\nu \in \mathcal{P}(\mathbb{R})$ and $\varepsilon \ge 0$,

$$\left|g_\nu(z) - g_{\mu_M}\big(z + g_\nu(z)\big)\right| \le \varepsilon,$$

then

$$|g_\mu(z) - g_\nu(z)| \le \varepsilon + \left|\phi_z\big(g_\mu(z)\big) - \phi_z\big(g_\nu(z)\big)\right| \le \varepsilon + \frac{1}{\mathfrak{Im}(z)^2}\left|g_\mu(z) - g_\nu(z)\right|.$$

So that, if $\mathfrak{Im}(z) \ge 2$,

$$\left|g_\mu(z) - g_\nu(z)\right| \le \tfrac{4}{3}\varepsilon.$$

Hence, from the definition of the distance $d(\mu, \nu)$ in (12), we see that Theorem 2.6 is a corollary of Theorem A.1.

A.2. Proof of Theorem A.1: The Gaussian case. In this subsection, we assume that:

(1) $G = (\mathfrak{Re}(Y_{12}), \mathfrak{Im}(Y_{12}))$ is a centered Gaussian vector in $\mathbb{R}^2$ with covariance $K \in \mathcal{H}_2(\mathbb{R})$, $\mathrm{tr}(K) = 1$.

(2) Y_{11} is a centered Gaussian in $\mathbb{R}$ with variance 1.

The proof is a variant of Pastur and Shcherbina [19], Lemma 2.2.3 (the main difference is that in [19], Lemma 2.2.3, the covariance matrix K is diagonal). We first recall the Gaussian integration by part formula (see, e.g., [19]): for any continuously differentiable function $F : \mathbb{R}^2 \mapsto \mathbb{R}$, with $\mathbb{E}\|\nabla F(G)\|_2 < \infty$,

$$\mathbb{E}F(G)G = K\mathbb{E}\nabla F(G). \tag{68}$$

We identify $\mathcal{H}_n(\mathbb{C})$ with $\mathbb{R}^{n^2}$. Then, if $\Phi : \mathcal{H}_n(\mathbb{C}) \mapsto \mathbb{C}$ is a continuously differentiable function, we define $D_{jk}\Phi(X)$ as the derivative with respect to $\mathfrak{Re}(X_{jk})$, and for $1 \le j \ne k \le n$, $D'_{jk}\Phi(X)$ as the derivative with respect to $\mathfrak{Im}(X_{jk})$.

Define the resolvent $R(X) = (X - z)^{-1}$, $z \in \mathbb{C}_+$. From the resolvent formula,

$$R(X + A) - R(X) = -R(X + A)AR(X), \tag{69}$$

valid for any matrix $A \in \mathcal{H}_n(\mathbb{C})$, a standard computation shows that if $1 \le j, k \le n$, and $1 \le a \ne b \le n$, then

$$D_{ab}R_{jk} = -(R_{ja}R_{bk} + R_{jb}R_{ak}) \quad \text{and} \quad D'_{ab}R_{jk} = -i(R_{ja}R_{bk} - R_{jb}R_{ak}),$$

while if $1 \le a \le n$, then

$$D_{aa}R_{jk} = -R_{ja}R_{ak}.$$

Set $X = Y/\sqrt{n} + M$, so that

$$R = (Y/\sqrt{n} + M - z)^{-1}.$$

Using (68) we get, for $0 \le a \ne b \le n$, and all j, k:

$$\begin{aligned}\mathbb{E} R_{jk} Y_{ab} &= \frac{1}{\sqrt{n}} \mathbb{E}\big[K_{11} D_{ab} R_{jk} + K_{12} D'_{ab} R_{jk} + i K_{21} D_{ab} R_{jk} + i K_{22} D'_{ab} R_{jk}\big] \\ &= -\frac{1}{\sqrt{n}} \mathbb{E}\big[(K_{11} - K_{22} + i K_{12} + i K_{21}) R_{ja} R_{bk} \\ &\qquad + (K_{11} + K_{22} - i K_{12} + i K_{21}) R_{jb} R_{ak}\big] \\ &= -\frac{1}{\sqrt{n}} \mathbb{E}(\gamma R_{ja} R_{bk} + R_{jb} R_{ak}),\end{aligned} \tag{70}$$

where at the last line, we have used the symmetry of K and $\mathrm{tr}(K) = 1$, together with the notation

$$\gamma = K_{11} - K_{22} + 2i K_{12} = \mathbb{E} Y_{ab}^2.$$

Notice that $|\gamma| \le 1$. Similarly, for $a = b$ one has

$$\mathbb{E} R_{jk} Y_{aa} = -\frac{1}{\sqrt{n}} \mathbb{E} R_{ja} R_{ak}. \tag{71}$$

Next, set

$$G(z) = (M - z)^{-1}.$$

Notice that in this case the dependency of $G(z)$ on z is explicit in our notation. From the resolvent formula (69),

$$R = G(z) - \frac{1}{\sqrt{n}} R Y G(z).$$

Hence, for $1 \le j, k \le n$, using (70)–(71),

$$\begin{aligned}\mathbb{E} R_{jk} &= G(z)_{jk} - \frac{1}{\sqrt{n}} \sum_{1 \le a, b \le n} \mathbb{E}[R_{ja} Y_{ab}] G(z)_{bk} \\ &= G(z)_{jk} + \frac{\gamma}{n} \sum_{1 \le a \ne b \le n} \mathbb{E}[R_{ja} R_{ba}] G(z)_{bk} + \frac{1}{n} \sum_{1 \le a, b \le n} \mathbb{E}[R_{jb} R_{aa}] G(z)_{bk}.\end{aligned}$$

We set

$$g = g_{\mu_{Y/\sqrt{n}+M}}(z) = \frac{1}{n} \sum_{a=1}^{n} R_{aa}, \qquad \overline{g} = \mathbb{E} g, \qquad \underline{g} = g - \mathbb{E} g,$$

and consider the diagonal matrix D with $D_{jk} = \mathbf{1}_{j=k} R_{jk}$. We find

$$\mathbb{E} R = G(z) + \mathbb{E}[g R] G(z) + \frac{\gamma}{n} \mathbb{E}\big[R\big(R^{\top} - D\big)\big] G(z).$$

Multiplying on the right-hand side by $G(z)^{-1} = M - z$ and subtracting $\overline{g}R$, one has

$$\mathbb{E}R(M - z - \overline{g}) = I + \mathbb{E}\underline{g}R + \frac{\gamma}{n}\mathbb{E}R(R^\top - D).$$

Multiplying on the right-hand side by $G(z + \overline{g})$,

$$\mathbb{E}R = G(z + \overline{g}) + \mathbb{E}\underline{g}RG(z + \overline{g}) + \frac{\gamma}{n}\mathbb{E}R(R^\top - D)G(z + \overline{g}).$$

Finally, multiplying by $\frac{1}{n}$ and taking the trace,

$$\overline{g} = g_{\mu_M}(z + \overline{g}) + \frac{1}{n}\mathbb{E}\underline{g}\,\mathrm{tr}[RG(z + \overline{g})] + \frac{\gamma}{n^2}\mathbb{E}\,\mathrm{tr}[R(R^\top - D)G(z + \overline{g})].$$

As a function of the entries of Y, g has Lipschitz constant $O(n^{-1}\,\mathfrak{Im}(z)^{-2})$. This fact can be seen, for example, as in [3], Lemma 2.3.1. Since the entries of Y satisfy a Poincaré inequality, a standard concentration bound [18] implies

$$\mathbb{E}|\underline{g}| = O(n^{-1}\,\mathfrak{Im}(z)^{-2}).$$

Also, since $|\mathrm{tr}(AB)| \le n\|A\|\|B\|$, we find

$$\left|\frac{1}{n}\,\mathrm{tr}\,RG(z + \overline{g})\right| \le \mathfrak{Im}(z)^{-2} \quad \text{and} \quad \left|\mathrm{tr}\,R(R^\top - D)G(z + \overline{g})\right| \le 2n\,\mathfrak{Im}(z)^{-3}.$$

This concludes the proof of Theorem A.1 in the Gaussian case.

A.3. Proof of Theorem A.1: The general case. Let $\underline{Y}_{ij} = Y_{ij} - \mathbb{E}Y_{12}$. Then $\underline{Y} - Y$ has rank at most 1. Hence, by Lemma B.1,

$$|g_{\mu_{Y/\sqrt{n}+M}}(z) - g_{\mu_{\underline{Y}/\sqrt{n}+M}}(z)| \le O((n\,\mathfrak{Im}(z))^{-1}),$$

where we have used (14) and the fact that $f(x) = (x - z)^{-1}$ has a bounded variation norm of order $\mathfrak{Im}(z)^{-1}$. Also, we recall that the map ϕ_z defined by (67) is Lipschitz with constant $1/\mathfrak{Im}(z)^2$. Hence, in order to prove Theorem A.1, we assume without loss of generality that the off-diagonal entries of the matrix are centered: $\mathbb{E}Y_{12} = 0$.

We now check that the diagonal entries of Y are negligible. Let Y' be the matrix obtained from Y by setting the diagonal equal to zero: $Y'_{ij} = \mathbf{1}_{i\ne j}Y_{ij}$.

LEMMA A.2 (Diagonal entries are negligible). *For $z \in \mathbb{C}_+$, $\mathfrak{Im}\,z \ge 1$,*

$$|\mathbb{E}g_{\mu_{Y/\sqrt{n}+M}}(z) - \mathbb{E}g_{\mu_{Y'/\sqrt{n}+M}}(z)| = O((\mathbb{E}|Y_{11}|^2/n)^{1/2}).$$

PROOF. From (77), we find

$$\begin{aligned}|\mathbb{E}g_{\mu_{Y/\sqrt{n}+M}}(z) - \mathbb{E}g_{\mu_{Y'/\sqrt{n}+M}}(z)| &\le \frac{\mathbb{E}W_1(\mu_{Y/\sqrt{n}+M}, \mu_{Y'/\sqrt{n}+M})}{(\mathfrak{Im}\,z)^2}\\ &\le \frac{\mathbb{E}W_2(\mu_{Y/\sqrt{n}+M}, \mu_{Y'/\sqrt{n}+M})}{(\mathfrak{Im}\,z)^2}.\end{aligned}$$

Then by Lemma B.2 using Jensen inequality,

$$\mathbb{E}W_2(\mu_{Y/\sqrt{n}+M},\mu_{Y'/\sqrt{n}+M}) \le \frac{1}{n}\left(\sum_{i=1}^{n}\mathbb{E}|Y_{ii}|^2\right)^{1/2} = \frac{1}{\sqrt{n}}(\mathbb{E}|Y_{11}|^2)^{1/2}. \qquad \square$$

As a consequence of Lemma A.2, we can assume without loss of generality that the diagonal entries of Y are independent centered Gaussian with variance 1. By Section A.2, the conclusion of Theorem A.1 holds for the matrix $\widehat{Y}$ whose off-diagonal entries are centered Gaussian random variables with covariance is K, where K is the covariance of Y, and with diagonal entries centered Gaussian with variance 1. Therefore, since the map ϕ_z defined by (67) is Lipschitz, in order to prove Theorem A.1, it is sufficient to establish that

$$\left|\mathbb{E}g_{\mu_{Y/\sqrt{n}+M}}(z) - \mathbb{E}g_{\mu_{\widehat{Y}/\sqrt{n}+M}}(z)\right| \le c\frac{\mathbb{E}|Y_{12}|^3}{n^{1/2}}. \tag{72}$$

We may repeat verbatim the interpolation trick in Pastur and Shcherbina [19], Theorem 18.3.1. Consider the random matrix $\widehat{Y}$, independent of Y, and for $0 \le t \le 1$, define the matrix

$$Y(t) = \sqrt{t}Y + \sqrt{1-t}\widehat{Y}.$$

Set $R(t) = (Y(t)/\sqrt{n} + M - zI)^{-1}$. Then, using the resolvent equation (69)

$$\begin{aligned}
&g_{\mu_{Y/\sqrt{n}+M}}(z) - g_{\mu_{\widehat{Y}/\sqrt{n}+M}}(z)\\
&\quad = \frac{1}{n}\int_0^1 \frac{d}{dt}\operatorname{tr} R(t)\,dt\\
&\quad = -\frac{1}{n^{3/2}}\int_0^1 \operatorname{tr} R(t)Y'(t)R(t)\,dt \qquad (73)\\
&\quad = -\frac{1}{2n^{3/2}}\int_0^1 \operatorname{tr} R(t)\left(\frac{Y}{\sqrt{t}} - \frac{\widehat{Y}}{\sqrt{1-t}}\right)R(t)\,dt\\
&\quad = -\frac{1}{2n^{3/2}}\int_0^1 \left[\operatorname{tr} R^2(t)\frac{Y}{\sqrt{t}} - \operatorname{tr} R^2(t)\frac{\widehat{Y}}{\sqrt{1-t}}\right]dt.
\end{aligned}$$

Next, consider the extension of (68) to arbitrary centered random variable G with covariance K. Namely, for any twice continuously differentiable function $F:\mathbb{R}^2 \mapsto \mathbb{R}$, with $\mathbb{E}\|\nabla F(G)\|_2 < \infty$ and $\sup_{x\in\mathbb{R}^2}\|\operatorname{Hess} F(x)\| < \infty$, a Taylor expansion gives

$$\mathbb{E}F(G)G = K\mathbb{E}\nabla F(G) + O\Big(\mathbb{E}\|G\|_2^3 \sup_{x\in\mathbb{R}^2}\|\operatorname{Hess} F(x)\|\Big).$$

Since Y and $\widehat{Y}$ have the same first two moments, we get for all $t \in [0, 1]$

$$\begin{aligned}
&\mathbb{E}\operatorname{tr} R^2(t)\frac{Y}{\sqrt{t}} - \mathbb{E}\operatorname{tr} R^2(t)\frac{\widehat{Y}}{\sqrt{1-t}} \\
&\quad= \sum_{1\le j,k\le n} \mathbb{E}R^2(t)_{kj}\frac{Y_{jk}}{\sqrt{t}} - \mathbb{E}R^2(t)_{kj}\frac{\widehat{Y}_{jk}}{\sqrt{1-t}} \\
&\quad\le c\frac{\mathbb{E}|Y_{12}|^3}{n} \sum_{1\le j,k\le n} \sup_{X\in\mathcal{H}_n(\mathbb{C}),\varepsilon,\varepsilon'} \bigl|D^{\varepsilon}_{jk}D^{\varepsilon'}_{jk}\bigl(R(X)^2\bigr)_{kj}\bigr|,
\end{aligned}$$

where $c > 0$ is a constant, and $D^{\varepsilon}_{jk}D^{\varepsilon'}_{jk}$ ranges over D^2_{jk}, D'^2_{jk} and $D_{jk}D'_{jk}$. However, it follows from (70)–(71) that

$$\bigl|D^{\varepsilon}_{jk}D^{\varepsilon'}_{jk}\bigl(R(X)^2\bigr)_{kj}\bigr|$$

is a finite linear combination of products of 4 resolvent entries of the form $\prod_{i=1}^4 R(X)_{u_i v_i}$. Since for any $X \in \mathcal{H}_n(\mathbb{C})$, $|R(X)_{jk}| \le (\Im\mathfrak{m}\, z)^{-1}$, one has for some new constant $c > 0$ and for all $t \in [0, 1]$:

$$\left|\mathbb{E}\operatorname{tr} R^2(t)\frac{Y}{\sqrt{t}} - \mathbb{E}\operatorname{tr} R^2(t)\frac{\widehat{Y}}{\sqrt{1-t}}\right| \le cn\frac{\mathbb{E}|Y_{12}|^3}{(\Im\mathfrak{m}\, z)^4}.$$

Plugging this last upper bound in (73) concludes the proof (72) and of Theorem A.1.

APPENDIX B

In this section, we collect some standard facts that are repeatedly used in the main text. For probability measures $\mu, \mu' \in \mathcal{P}(\mathbb{R})$, the *Kolmogorov–Smirnov* (KS) distance is defined by

$$d_{\mathrm{KS}}(\mu, \mu') = \sup_{t\in\mathbb{R}}\bigl|\mu(-\infty, t] - \mu'(-\infty, t]\bigr|. \tag{74}$$

The KS distance is closely related to functions with bounded variations. More precisely, for $f : \mathbb{R} \mapsto \mathbb{R}$ the bounded variation norm is defined as

$$\|f\|_{\mathrm{BV}} = \sup \sum_{k\in\mathbb{Z}} \bigl|f(x_{k+1}) - f(x_k)\bigr|,$$

where the supremum is over all sequence $(x_k)_{k\in\mathbb{Z}}$ with $x_n \le x_{n+1}$. If $f = \mathbf{1}((-\infty, t))$, then $\|f\|_{BV} = 1$ while if the derivative of f is in $L^1(\mathbb{R})$, we have $\|f\|_{\mathrm{BV}} = \int |f'(x)|\, dx$. The KS distance is also given by the variational formula

$$d_{\mathrm{KS}}(\mu, \mu') = \sup\left\{\int f\, d\mu - \int f\, d\mu' : \|f\|_{\mathrm{BV}} \le 1\right\}. \tag{75}$$

[Indeed, the functions $H_t = \mathbf{1}((-\infty, t))$, $t \in \mathbb{R}$, are the extremal points of the convex set of functions f with $\|f\|_{\mathrm{BV}} \leq 1$ and the map $f \to \int f\, d\mu - \int f\, d\mu'$ is linear].

For $p \geq 1$ and $\mu, \mu' \in \mathcal{P}(\mathbb{R})$ such that $\int |x|^p\, d\mu(x)$ and $\int |x|^p\, d\mu'(x)$ are finite, their L^p-Wasserstein distance is defined as

$$W_p(\mu, \mu') = \left(\inf_{\pi} \int_{\mathbb{R}\times\mathbb{R}} |x - y|^p\, d\pi(x, y)\right)^{1/p}, \tag{76}$$

where the infimum is over all coupling π of μ and μ' (i.e., π is probability measure on $\mathbb{R} \times \mathbb{R}$ whose first marginal is equal to μ and second marginal is equal to μ'). Hölder's inequality implies that for $1 \leq p \leq p'$, $W_p \leq W_{p'}$.

For any $p \geq 1$, if $W_p(\mu_n, \mu)$ converges to 0 then $\mu_n \rightsquigarrow \mu$. This follows, for example, from the Kantorovich–Rubinstein duality

$$W_1(\mu, \mu') = \sup\left\{\int f\, d\mu - \int f\, d\mu' : \|f\|_{\mathrm{Lip}} \leq 1\right\}, \tag{77}$$

where $\|f\|_{\mathrm{Lip}}$ denotes the Lipschitz constant of f (see, e.g., Dudley [15], Theorem 11.8.2).

The following inequality is a standard consequence of interlacing; see, for example, [4], Theorem A.43.

LEMMA B.1 (Rank inequality). *If A, B in $\mathcal{H}_n(\mathbb{C})$, then*

$$d_{\mathrm{KS}}(\mu_A, \mu_B) \leq \frac{1}{n}\operatorname{rank}(A - B).$$

Next, we recall a very useful estimate which allows one to bound eigenvalue differences in terms of matrix entries. For a proof see, for example, [3], Lemma 2.1.19.

LEMMA B.2 (Hoffman–Wielandt inequality). *If A, B in $\mathcal{H}_n(\mathbb{C})$, then*

$$W_2(\mu_A, \mu_B) \leq \sqrt{\frac{1}{n}\operatorname{tr}\bigl[(A - B)^2\bigr]}.$$

Acknowledgments. The idea of studying this model came from a discussion with Manjunath Krishnapur. We also thank the anonymous referees for several helpful comments.

REFERENCES

[1] ALDOUS, D. and LYONS, R. (2007). Processes on unimodular random networks. *Electron. J. Probab.* **12** 1454–1508. MR2354165

[2] ALDOUS, D. and STEELE, J. M. (2004). The objective method: Probabilistic combinatorial optimization and local weak convergence. In *Probability on Discrete Structures*. *Encyclopaedia Math. Sci.* **110** 1–72. Springer, Berlin. MR2023650

[3] ANDERSON, G. W., GUIONNET, A. and ZEITOUNI, O. (2010). *An Introduction to Random Matrices*. *Cambridge Studies in Advanced Mathematics* **118**. Cambridge Univ. Press, Cambridge. MR2760897

[4] BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer, New York. MR2567175

[5] BENJAMINI, I. and SCHRAMM, O. (2001). Recurrence of distributional limits of finite planar graphs. *Electron*. *J*. *Probab*. **6** 13 pp. (electronic). MR1873300

[6] BENNETT, G. (1962). Probability inequalities for the sum of independent random variables. *J*. *Amer*. *Statist*. *Assoc*. **57** 33–45.

[7] BEN AROUS, G. and GUIONNET, A. (1997). Large deviations for Wigner's law and Voiculescu's noncommutative entropy. *Probab*. *Theory Related Fields* **108** 517–542. MR1465640

[8] BIANE, P. (1997). On the free convolution with a semi-circular distribution. *Indiana Univ*. *Math*. *J*. **46** 705–718. MR1488333

[9] BORDENAVE, C., CAPUTO, P. and CHAFAÏ, D. (2011). Spectrum of large random reversible Markov chains: Heavy-tailed weights on the complete graph. *Ann*. *Probab*. **39** 1544–1590. MR2857250

[10] BORDENAVE, C. and LELARGE, M. (2010). Resolvent of large random graphs. *Random Structures Algorithms* **37** 332–352. MR2724665

[11] BORDENAVE, C., LELARGE, M. and SALEZ, J. (2011). The rank of diluted random graphs. *Ann*. *Probab*. **39** 1097–1121. MR2789584

[12] CAPITAINE, M., DONATI-MARTIN, C., FÉRAL, D. and FÉVRIER, M. (2011). Free convolution with a semicircular distribution and eigenvalues of spiked deformations of Wigner matrices. *Electron*. *J*. *Probab*. **16** 1750–1792. MR2835253

[13] CHATTERJEE, S. and VARADHAN, S. R. S. (2012). Large deviations for random matrices. *Commun*. *Stoch*. *Anal*. **6** 1–13. MR2890846

[14] DEMBO, A. and ZEITOUNI, O. (2010). *Large Deviations Techniques and Applications*. *Stochastic Modelling and Applied Probability* **38**. Springer, Berlin. MR2571413

[15] DUDLEY, R. M. (2002). *Real Analysis and Probability*. *Cambridge Studies in Advanced Mathematics* **74**. Cambridge Univ. Press, Cambridge. MR1932358

[16] EICHELSBACHER, P., SOMMERAUER, J. and STOLZ, M. (2011). Large deviations for disordered bosons and multiple orthogonal polynomial ensembles. *J*. *Math*. *Phys*. **52** 073510, 16. MR2849043

[17] GUIONNET, A. and ZEITOUNI, O. (2000). Concentration of the spectral measure for large matrices. *Electron*. *Commun*. *Probab*. **5** 119–136 (electronic). MR1781846

[18] LEDOUX, M. (2001). *The Concentration of Measure Phenomenon*. *Mathematical Surveys and Monographs* **89**. Amer. Math. Soc., Providence, RI. MR1849347

[19] PASTUR, L. and SHCHERBINA, M. (2011). *Eigenvalue Distribution of Large Random Matrices*. *Mathematical Surveys and Monographs* **171**. Amer. Math. Soc., Providence, RI. MR2808038

[20] REED, M. and SIMON, B. (1980). *Methods of Modern Mathematical Physics*. *I*: *Functional Analysis*, 2nd ed. Academic Press, New York. MR0751959

[21] ZHAN, X. (2002). *Matrix Inequalities*. *Lecture Notes in Math*. **1790**. Springer, Berlin. MR1927396

INSTITUT DE MATHÉMATIQUES DE TOULOUSE
CNRS AND UNIVERSITÉ TOULOUSE III
118 ROUTE DE NARBONNE
31062 TOULOUSE
FRANCE
E-MAIL: charles.bordenave@math.univ-toulouse.fr
URL: http://www.math.univ-toulouse.fr/~bordenave

DIPARTIMENTO DI MATEMATICA E FISICA
UNIVERSITÀ ROMA TRE
LARGO SAN MURIALDO 1
00146 ROMA
ITALY
E-MAIL: caputo@mat.uniroma3.it
URL: http://www.mat.uniroma3.it/users/caputo