

Submitted to the *Annals of Probability*

LIMITING ENTROPY OF DETERMINANTAL PROCESSES

BY ANDRÁS MÉSZAROS¹,¹Central European University, Budapest Meszaros_Andras@phd.ceu.edu

We extend Lyons's tree entropy theorem to general determinantal measures. As a byproduct we show that the sofic entropy of an invariant determinantal measure does not depend on the chosen sofic approximation.

1. Introduction. Let $P = (p_{ij})$ be an orthogonal projection matrix, where rows and columns are both indexed with a finite set V . Then there is a unique probability measure η_P on the subsets of V such that for every $F \subset V$ we have

$$\eta_P(\{B|F \subset B \subset V\}) = \det(p_{ij})_{i,j \in F}.$$

The measure η_P is called the *determinantal measure* corresponding to P [13]. Let B^P be a random subset of V with distribution η_P . In this paper we investigate the asymptotic behavior of the *Shannon-entropy* of B^P defined as

$$H(B^P) = \sum_{A \subset V} -\mathbb{P}(B^P = A) \log \mathbb{P}(B^P = A).$$

Let P_1, P_2, \dots be a sequence of orthogonal projection matrices. Assume that rows and columns of P_n are both indexed with the finite set V_n . Let G_n be a graph on the vertex set V_n . Throughout the paper we assume that the degrees of graphs are at most D for some fixed finite D . Our main theorem is the following.

THEOREM. *Assume that the sequence of pairs (G_n, P_n) is Benjamini-Schramm convergent and tight. Then*

$$\lim_{n \rightarrow \infty} \frac{H(B^{P_n})}{|V_n|}$$

exists.

Note that this theorem will be restated in a slightly more general and precise form as Theorem 2.5 in the next section. We will also give a formula for the limit.

We define *Benjamini-Schramm convergence* of (G_n, P_n) along the lines of [6] and [2] via the following local sampling procedure. Fix any positive integer r , this will be our radius of sight. For a vertex $o \in V_n$ let $B_r(G_n, o)$ be the r -neighborhood of o in the graph G_n , and let $M_{n,r,o}$ be the submatrix of P_n determined by rows and columns with indices in $B_r(G_n, o)$. Then the outcome of the local sampling at o is the pair $(B_r(G_n, o), M_{n,r,o})$. Of course, we are only interested in the outcome up to rooted isomorphism. Now if we pick o as a uniform random element of V_n , we get a probability measure $\mu_{n,r}$ on the set of isomorphism classes of pairs (H, M) , where H is a rooted r -neighborhood and M is a matrix where rows and columns are indexed with the vertices of H . We say that the sequence (G_n, P_n) converges if for any fixed r the measures $\mu_{n,r}$ converge weakly as n tends to infinity. See the next section for more details including the description of the limit object.

MSC 2010 subject classifications: Primary 60C05, 37A35; secondary 60K99, 60B99

Keywords and phrases: determinantal processes, spanning trees, tree entropy, sofic entropy, weak convergence

To define the notion of tightness, we introduce a measure ν_n on $\mathbb{N} \cup \{\infty\}$ for each pair (G_n, P_n) as follows. Given $k \in \mathbb{N} \cup \{\infty\}$ we set

$$\nu_n(\{k\}) = |V_n|^{-1} \sum_{\substack{u, v \in V_n \\ d_n(u, v) = k}} |P_n(u, v)|^2,$$

where d_n is the graph metric on $V_n = V(G_n)$. Then the sequence (G_n, P_n) is tight if the family of measures ν_n is tight, that is, for each $\varepsilon > 0$ we have a finite R such that

$$\nu_n(\{R+1, R+2, \dots\} \cup \{\infty\}) < \varepsilon$$

for all n . Tightness makes sure that the local sampling procedure from the previous paragraph detects most of the significant matrix entries for large enough r .

Note that a related convergence notion of operators was introduced by Lyons and Thom [15]. We expect that their notion is slightly stronger, but were unable to clarify this.

The idea of the proof of the main theorem is the following. Consider a uniform random ordering of V_n . Then using the chain rule for conditional entropy we can write $H(B^{P_n})$ as the sum of $|V_n|$ conditional entropies. We show that in the limit we can control these conditional entropies. This method in the context of local convergence first appeared in [7].

Now we describe a special case of our theorem. Consider a finite connected graph G , and consider the uniform measure on the set of spanning trees of G . This measure turns out to be a determinantal measure, the corresponding projection matrix $P_\star(G)$ is called the transfer-current matrix [9]. Since this is a uniform measure, the Shannon-entropy is simply $\log \tau(G)$, where $\tau(G)$ is the number of spanning trees in G . A theorem of Lyons [14] states that if G_n is a Benjamini-Schramm convergent sequence of finite connected graphs then

$$\lim_{n \rightarrow \infty} \frac{\log \tau(G_n)}{|V(G_n)|}$$

exists. This theorem now follows from our results, because it is easy to see that the sequence $(L(G_n), P_\star(G_n))$ is convergent and tight in our sense, where $L(G_n)$ is the line graph of G_n . See Section 7. Note that we need to take the line graph of G_n , because the uniform spanning tree measure is defined on the edges of G_n rather than the vertices of G_n . We also obtain a formula for limit which is different from Lyons's original formula. However, in practice it seems easier to evaluate Lyons's original formula.

Another application comes from ergodic theory. Let Γ be a finitely generated countable group, and let T be an invariant positive contraction on $\ell^2(\Gamma)$. Here a linear operator is called a positive contraction if it is positive semidefinite and has operator norm at most 1. Invariance means that for any $\gamma, g_1, g_2 \in \Gamma$ we have

$$\langle Tg_1, g_2 \rangle = \langle T(\gamma^{-1}g_1), \gamma^{-1}g_2 \rangle.$$

Note that here we identify elements of Γ with their characteristic vectors. Then the determinantal measure μ_T corresponding to T gives us an invariant measure on $\{0, 1\}^\Gamma$. Note that there is a natural graph structure on Γ . Namely, we can fix a finite generating set S , and consider the corresponding Cayley-graph $\text{Cay}(\Gamma, S)$. When Γ belongs to the class of sofic groups, one can define the so-called sofic entropy of this invariant measure [1]. This is done by first considering an approximation of $\text{Cay}(\Gamma, S)$ by a sequence of finite graphs G_n , and then investigating how we can model μ_T on these finite graphs. In general it is not known whether sofic entropy depends on the chosen approximating sequence G_n or not, apart from certain trivial examples. However, in our special case, our results allow us to give a formula

for the sofic entropy, which only depends on the measure μ_T , but not on the finite approximations. This shows that in this case the sofic entropy does not depend on the chosen sofic approximation.

Observe that in our main theorem the graphs G_n do not play any role in the definition of the random subsets B^{P_n} or the Shannon entropy $H(B^{P_n})$, they are only there to help us define our convergence notion. This suggests that there might be a notion of convergence of orthogonal projection matrices without any additional graph structure such that the normalized Shannon entropy of B^{P_n} is continuous.

Structure of the paper. In Section 2 we explain the basic definitions and state our results. In Section 3 we investigate what happens if we condition a Benjamini-Schramm convergence sequence of determinantal measures in a Benjamini-Schramm convergent way. In Sections 4, 5 and 6 we prove the theorems stated in Section 2. In Section 7 we explain the connections of our results and Lyons's tree entropy theorem. The proof of a technical lemma about the measurability of the polar decomposition is given in the Appendix.

2. Definitions and statements of the results.

2.1. The space of rooted graphs and sofic groups. Fix a degree bound D . A *rooted graph* is a pair (G, o) where G is a (possibly infinite) connected graph with degrees at most D , $o \in V(G)$ is a distinguished vertex of G called the root. Given two rooted graphs (G_1, o_1) and (G_2, o_2) their distance is defined to be the infimum over all $\varepsilon > 0$ such that for $r = \lfloor \varepsilon^{-1} \rfloor$ there is a root preserving graph isomorphism from $B_r(G_1, o_1)$ to $B_r(G_2, o_2)$. Let \mathcal{G} be the set of isomorphism classes of rooted graphs. With the above defined distance \mathcal{G} is a compact metric space. Therefore, the set of probability measures $\mathcal{P}(\mathcal{G})$ endowed with the weak* topology is also compact. A sequence of random rooted graphs (G_n, o_n) *Benjamini-Schramm converges* to the random rooted graph (G, o) , if their distributions converge in $\mathcal{P}(\mathcal{G})$. Given any finite graph G , we can turn it into a random rooted graph $U(G) = (G_o, o)$ by considering a uniform random vertex o of G and its connected component G_o . A sequence of finite graphs G_n Benjamini-Schramm converges to the random rooted graph (G, o) if the sequence $U(G_n)$ Benjamini-Schramm converges to (G, o) .

Let S be a finite set, an *S -labeled Schreier graph* is a graph where each edge is oriented and labeled with an element from S , moreover for every vertex v of the graph and every $s \in S$ there is exactly one edge labeled with s entering v and there is exactly one edge labeled with s leaving v . For example, if Γ is a group with generating set S , then its Cayley-graph $\text{Cay}(\Gamma, S)$ is an S -labeled Schreier-graph. The notion of Benjamini-Schramm convergence can be extended to the class of S -labeled Schreier-graphs with the modification that graph isomorphisms are required to respect the orientation and labeling of the edges. Let Γ be a finitely generated group. Fix a finite generating set S , and consider the Cayley-graph $G_\Gamma = \text{Cay}(\Gamma, S)$. Let e_Γ be the identity of Γ . We say that Γ is *sofic* if there is a sequence of finite S -labeled Schreier-graphs G_n , such that G_n Benjamini-Schramm converges to (G_Γ, e_Γ) .

2.2. The space of rooted graph-operators. Fix a degree bound D , and let K be a non-empty finite set.

A *rooted graph-operator* (RGO) is a triple (G, o, T) , where (G, o) is a rooted graph and T is a bounded operator on $\ell^2(V(G) \times K)$. In this paper we will use real Hilbert spaces, but the results can be generalized to the complex case as well. Note that to prove our main theorem it suffices to only consider the case $|K| = 1$. The usefulness of allowing $|K| > 1$ will be only clear in Section 5, where we extend our results to positive contractions.

Given two RGOs (G_1, o_1, T_1) and (G_2, o_2, T_2) their distance $d((G_1, o_1, T_1), (G_2, o_2, T_2))$ is defined as the infimum over all $\varepsilon > 0$ such that for $r = \lfloor \varepsilon^{-1} \rfloor$ there is a root preserving

graph isomorphism ψ from $B_r(G_1, o_1)$ to $B_r(G_2, o_2)$ with the property that

$$(1) \quad |\langle T_1(v, k), (v', k') \rangle - \langle T_2(\psi(v), k), (\psi(v'), k') \rangle| < \varepsilon$$

for every $v, v' \in V(B_r(G_1, o_1))$ and $k, k' \in K$. Here we identified elements of $V(G_i) \times K$ with their characteristic vectors in $\ell^2(V(G_i) \times K)$.

Two RGOs (G_1, o_1, T_1) and (G_2, o_2, T_2) are called isomorphic if their distance is 0, or equivalently if there is a root preserving graph isomorphism ψ from (G_1, o_1) to (G_2, o_2) such that

$$\langle T_1(v, k), (v', k') \rangle = \langle T_2(\psi(v), k), (\psi(v'), k') \rangle$$

for every $v, v' \in V(G_1)$ and $k, k' \in K$. Let $\mathcal{RG}\mathcal{O}$ be the set of isomorphism classes of RGOs. For any $0 < B < \infty$, we define

$$\mathcal{RG}\mathcal{O}(B) = \{(G, o, T) \in \mathcal{RG}\mathcal{O} \mid \|T\| \leq B\}.$$

One can prove that $\mathcal{RG}\mathcal{O}(B)$ is a compact metric space with the above defined distance d . Let $\mathcal{P}(\mathcal{RG}\mathcal{O}(B))$ be the set of probability measures on $\mathcal{RG}\mathcal{O}(B)$ endowed with the weak* topology, this is again a compact space. Often it will be more convenient to consider an element $\mathcal{P}(\mathcal{RG}\mathcal{O})$ as a random RGO.

A RGO (G, o, T) is called a *rooted graph-positive-contraction* (RGPC) if T is a self-adjoint positive operator with norm at most 1. Then the set $\mathcal{RG}\mathcal{PC}$ of isomorphism classes of RGPCs is a compact metric space. Therefore, $\mathcal{P}(\mathcal{RG}\mathcal{PC})$ with the weak* topology is compact.

We need a slight generalization of the notion of RGO. An *h-decorated RGO* is a tuple $(G, o, T, A^{(1)}, A^{(2)}, \dots, A^{(h)})$, where G, o and T are like above, $A^{(1)}, A^{(2)}, \dots, A^{(h)}$ are subsets of $V(G) \times K$. Given two *h-decorated RGOs* $(G_1, o_1, T_1, A_1^{(1)}, A_1^{(2)}, \dots, A_1^{(h)})$ and $(G_2, o_2, T_2, A_2^{(1)}, A_2^{(2)}, \dots, A_2^{(h)})$ their distance is defined as the infimum over all $\varepsilon > 0$ such that for $r = \lfloor \varepsilon^{-1} \rfloor$ there is a root preserving graph isomorphism ψ from $B_r(G_1, o_1)$ to $B_r(G_2, o_2)$ satisfying the property given in (1), and for $i = 1, 2, \dots, h$ we have

$$\bar{\psi}(A_1^{(i)} \cap (B_r(G_1, o_1) \times K)) = A_2^{(i)} \cap (B_r(G_2, o_2) \times K),$$

where $\bar{\psi}(v, k) = (\psi(v), k)$.

Two *h-decorated RGOs* $(G_1, o_1, T_1, A_1^{(1)}, \dots, A_1^{(h)})$ and $(G_2, o_2, T_2, A_2^{(1)}, \dots, A_2^{(h)})$ are called isomorphic if their distance is 0. Let $\mathcal{RG}\mathcal{O}_h$ be the set of isomorphism classes of *h-decorated RGOs*. We also define $\mathcal{RG}\mathcal{O}_h(B)$ and $\mathcal{RG}\mathcal{PC}_h$ the same way as their non-decorated versions were defined. With the above defined distance they are compact metric spaces. Similarly as before, $\mathcal{P}(\mathcal{RG}\mathcal{O}_h(B))$ and $\mathcal{P}(\mathcal{RG}\mathcal{PC}_h)$, endowed with the weak* topology, are compact spaces. Whenever the value of h is clear from the context, we omit it and simply use the term "decorated RGO".

A *finite graph-positive-contraction* is a pair (G, T) , where G is finite graph with degrees at most D , and T is a positive contraction on $\ell^2(V(G) \times K)$. It can be turned into a random RGPC

$$U(G, T) = (G_o, o, T_o)$$

by choosing o as a uniform random vertex of G .

Note that all the definitions above depend on the choice of the finite set K . In most of the paper we can keep K as fixed. Whenever we need to emphasize the specific choice of K , we will refer to K as the support set of RGOs. Unless stated otherwise the support set is always assumed to be K . Let $L \subset K$ and let (G, o, T) be a RGO with support set K . Let P_L be the orthogonal projection from $\ell^2(V(G) \times K)$ to $\ell^2(V(G) \times L) \subset \ell^2(V(G) \times K)$. We define the

operator $\text{rest}_L(T)$ on $\ell^2(V(G) \times L)$ as $\text{rest}_L(T) = P_L T \upharpoonright_{\ell^2(V(G) \times L)}$. So $(G, o, \text{rest}_L(T))$ is an RGO with support set L .

Sometimes we need to consider more than one operator on a rooted graph. A *double RGO* will mean a tuple (G, o, T_1, T_2) where (G, o) is a rooted graph and T_1, T_2 are bounded operators on $\ell^2(V(G) \times K)$. We omit the details how the set of isomorphism classes of double RGOs can be turned into a metric space. It is also clear what we mean by a decorated double RGO, or a triple RGO, or a double RGPC.

2.3. Determinantal processes. Let E be a countable set, and T be a positive contraction of $\ell^2(E)$. Then there is a random subset B^T of E with the property that for each finite subset F of E we have

$$\mathbb{P}[F \subset B^T] = \det(\langle Tx, y \rangle)_{x, y \in F},$$

where we identify an element $x \in E$ with its characteristic vector in $\ell^2(E)$. The distribution of B^T is uniquely determined by these constraints, and it is called the *determinantal measure* corresponding to T [13].

Using the definition of the random subset B^T , we can define a map $\tau : \mathcal{RGPC} \rightarrow \mathcal{P}(\mathcal{RGPC}_1)$ by $\tau(G, o, T) = (G, o, T, B^T)$. This induces a map $\tau_* : \mathcal{P}(\mathcal{RGPC}) \rightarrow \mathcal{P}(\mathcal{P}(\mathcal{RGPC}_1))$. Taking expectation we get the map $\mathbb{E}\tau_* : \mathcal{P}(\mathcal{RGPC}) \rightarrow \mathcal{P}(\mathcal{RGPC}_1)$. So given a random RGPC (G, o, T) the meaning of (G, o, T, B^T) is ambiguous. Unless stated otherwise (G, o, T, B^T) will mean a random decorated RGPC, i.e., its distribution is an element of $\mathcal{P}(\mathcal{RGPC}_1)$.

PROPOSITION 2.1. The maps τ, τ_* and $\mathbb{E}\tau_*$ are continuous.

2.4. Trace and spectral measure. Given a random RGO (G, o, T) we define

$$\text{Tr}(G, o, T) = \mathbb{E} \sum_{k \in K} \langle T(o, k), (o, k) \rangle.$$

We extend the definition to the decorated case in the obvious way.

Given a random RGPC (G, o, T) its *spectral measure* is the unique measure $\mu = \mu_{(G, o, T)}$ on $[0, 1]$ with the property that, for any integer $n \geq 0$ we have

$$\text{Tr}(G, o, T^n) = \int_0^1 x^n d\mu.$$

Note that $\mu([0, 1]) = |K|$. Also if T is a projection with probability 1, then we have

$$\mu = \text{Tr}(G, o, T)\delta_1 + (|K| - \text{Tr}(G, o, T))\delta_0.$$

If (G, T) is a finite graph-positive-contraction, then the spectral measure of $U(G, T)$ can be obtained as

$$\frac{1}{|V(G)|} \sum_{i=1}^{|V(G) \times K|} \delta_{\lambda_i},$$

where $\lambda_1, \lambda_2, \dots, \lambda_{|V(G) \times K|}$ are the eigenvalues of T with multiplicity.

2.5. *An equivalent characterization of tightness.* We already defined the notion of tightness in the Introduction. Here we repeat the definition in a slightly more general setting. For a finite graph-positive-contraction (G, T) we define the measure $\nu_{(G,T)}$ on $\mathbb{N} \cup \{\infty\}$ by setting

$$\nu_{(G,T)}(\{t\}) = |V(G)|^{-1} \sum_{\substack{(v_1, k_1), (v_2, k_2) \in V(G) \times K \\ d_G(v_1, v_2) = t}} |\langle T(v_1, k_1), (v_2, k_2) \rangle|^2,$$

for all $t \in \mathbb{N} \cup \{\infty\}$. A sequence (G_n, T_n) of finite graph-positive-contractions is tight if the family of measures $\nu_{(G_n, T_n)}$ is tight, that is, for each $\varepsilon > 0$ we have a finite R such that

$$\nu_{(G_n, T_n)}(\{R+1, R+2, \dots\} \cup \{\infty\}) < \varepsilon$$

for all n . The next lemma gives an equivalent characterization of tightness.

LEMMA 2.2. *Let (G_n, P_n) be a Benjamini-Schramm convergent sequence of finite graph-positive-contractions with limit (G, o, T) . Assume that P_1, P_2, \dots are orthogonal projections. Then the following are equivalent*

- i) *The sequence (G_n, P_n) is tight.*
- ii) *The limit T is an orthogonal projection with probability 1 and $\nu_{(G_n, P_n)}(\{\infty\}) = 0$ for every n .*

PROOF. i) \Rightarrow ii): Recall the following well-known result.

PROPOSITION 2.3. Let E be a countable set, and let T be a positive contraction on $\ell^2(E)$. Then for all $e \in E$ we have $\langle T^2 e, e \rangle \leq \langle T e, e \rangle$. Moreover, if for all $e \in E$ we have $\langle T^2 e, e \rangle = \langle T e, e \rangle$, then T is an orthogonal projection.

Let $(H_n, o_n, T_n) = U(G_n, P_n)$. Then

$$\nu_{(G_n, P_n)}(\mathbb{N} \cup \{\infty\}) = |V(G_n)|^{-1} \text{Tr}(P_n^* P_n) = |V(G_n)|^{-1} \text{Tr}(P_n) = \text{Tr}(H_n, o_n, T_n).$$

Combining this with the definition of tightness we get that for any $\varepsilon > 0$ we have an R such that

$$(2) \quad \mathbb{E} \sum_{k \in K} \sum_{(v, k') \in B_R(H_n, o_n) \times K} |\langle T_n(o_n, k), (v, k') \rangle|^2 > \text{Tr}(H_n, o_n, T_n) - \varepsilon$$

for every n .

Using the convergence of (H_n, o_n, T_n) we get that

$$\lim_{n \rightarrow \infty} \text{Tr}(H_n, o_n, T_n) = \text{Tr}(G, o, T),$$

and

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E} \sum_{k \in K} \sum_{(v, k') \in B_R(H_n, o_n) \times K} |\langle T_n(o_n, k), (v, k') \rangle|^2 \\ = \mathbb{E} \sum_{k \in K} \sum_{(v, k') \in B_R(G, o) \times K} |\langle T(o, k), (v, k') \rangle|^2. \end{aligned}$$

Combining these with inequality (2) we get that

$$\begin{aligned} \text{Tr}(G, o, T^2) &= \mathbb{E} \sum_{k \in K} \sum_{(v, k') \in V(G) \times K} |\langle T(o, k), (v, k') \rangle|^2 \\ &\geq \mathbb{E} \sum_{k \in K} \sum_{(v, k') \in B_R(G, o) \times K} |\langle T(o, k), (v, k') \rangle|^2 \geq \text{Tr}(G, o, T) - \varepsilon. \end{aligned}$$

Tending to 0 with ε we get that

$$\mathrm{Tr}(G, o, T^2) \geq \mathrm{Tr}(G, o, T).$$

Combining this with the first statement of Proposition 2.3 we get that with probability 1 we have $\langle T^2(o, k), (o, k) \rangle = \langle T(o, k), (o, k) \rangle$ for every $k \in K$. But then it follows from the unimodularity of (G, o, T) that with probability 1 we have $\langle T^2(v, k), (v, k) \rangle = \langle T(v, k), (v, k) \rangle$ for any $(v, k) \in V(G) \times K$. See [2, Lemma 2.3 (Everything Shows at the Root)] and Section 3. Then Proposition 2.3 gives us that T is a projection with probability 1. From the definition of tightness it is clear that $\nu_n(\{\infty\}) = 0$ for every n .

ii) \Rightarrow i): Pick any $\varepsilon > 0$. From the monotone convergence theorem and the fact that T is a projection with probability 1 we have

$$\begin{aligned} \mathrm{Tr}(G, o, T) &= \mathrm{Tr}(G, o, T^2) = \mathbb{E} \sum_{k \in K} \sum_{(v, k') \in V(G) \times K} |\langle T(o, k), (v, k') \rangle|^2 \\ &= \lim_{R \rightarrow \infty} \mathbb{E} \sum_{k \in K} \sum_{(v, k') \in B_R(G, o) \times K} |\langle T(o, k), (v, k') \rangle|^2 \end{aligned}$$

Thus, if we choose a large enough R_0 , then we have

$$\mathrm{Tr}(G, o, T) - \mathbb{E} \sum_{k \in K} \sum_{(v, k') \in B_{R_0}(G, o) \times K} |\langle T(o, k), (v, k') \rangle|^2 < \frac{\varepsilon}{2}.$$

Then from the convergence of (H_n, o_n, T_n) we get that there is an N such that if $n > N$ we have

$$\begin{aligned} \nu_{(G_n, P_n)}(\{R_0 + 1, R_0 + 2, \dots\} \cup \{\infty\}) &= \\ \mathrm{Tr}(H_n, o_n, T_n) - \mathbb{E} \sum_{k \in K} \sum_{(v, k') \in B_{R_0}(H_n, o_n) \times K} |\langle T_n(o_n, k), (v, k') \rangle|^2 &< \varepsilon. \end{aligned}$$

Using the condition that $\nu_{(G_n, P_n)}(\{\infty\}) = 0$ for all n and the definition of $\nu_{(G_n, P_n)}$ we get that the support of the measure $\nu_{(G_n, P_n)}$ is contained in $\{0, 1, \dots, |V(G_n)|\}$. Thus, the choice $R = \max(R_0, |V(G_1)|, |V(G_2)|, \dots, |V(G_N)|)$ is good for ε . \square

2.6. Sofic entropy. Let C be a finite set and let Γ be a finitely generated group. Let f be a random coloring of Γ with C , that is a random element of C^Γ . (The measurable structure of C^Γ comes from the product topology on C^Γ .) Given a coloring $f \in C^\Gamma$ and $\gamma \in \Gamma$ we define the coloring f_γ by $f_\gamma(g) = f(\gamma^{-1}g)$ for all $g \in \Gamma$. This notation extends to random colorings in the obvious way. A random coloring f is invariant if for every $\gamma \in \Gamma$ the distribution of f_γ is the same as the distribution of f .

Now assume that Γ is a finitely generated sofic group, and f is an invariant random coloring of Γ . Let S be a finite generating set, and let G_1, G_2, \dots be a sequence of S -labeled Schreier-graphs Benjamini-Schramm converging to the Cayley-graph $G_\Gamma = \mathrm{Cay}(\Gamma, S)$. Now we define the so called *sofic entropy* of f . There are many slightly different versions of this notion [8, 3], we will follow Abért and Weiss [1]. Let G be a finite S -labeled Schreier graph and g be a random coloring of $V(G)$. Given $\varepsilon > 0$ and a positive integer r , we say that g is an (ε, r) approximation of f on the graph G , if there are at least $(1 - \varepsilon)|V(G)|$ vertices $v \in V(G)$, such that $B_r(G, v)$ is isomorphic to $B_r(G_\Gamma, e_\Gamma)$, moreover $d_{TV}(f \upharpoonright B_r(G_\Gamma, e_\Gamma), g \upharpoonright B_r(G, v)) < \varepsilon$, where d_{TV} is the total variational distance, and it is meant that we identify $B_r(G_\Gamma, e_\Gamma)$ and $B_r(G, v)$. Let us define

$$H(G, \varepsilon, r) = \sup \left\{ \frac{H(g)}{|V(G)|} \mid g \text{ is an } (\varepsilon, r) \text{ approximation of } f \text{ on } G \right\}.$$

Here $H(g)$ is the Shannon-entropy of g . Let $H(\varepsilon, r)$ be the supremum of $H(G, \varepsilon, r)$, over all finite S -labeled Schreier graphs G . We define two versions of sofic entropy. The first one

$$h(f) = \inf_{\varepsilon, r} \limsup_{n \rightarrow \infty} H(G_n, \varepsilon, r).$$

Note that this might depend on the chosen sofic approximation. Another option is to define sofic entropy as

$$h'(f) = \inf_{\varepsilon, r} H(\varepsilon, r).$$

Observe that $h'(f) \geq h(f)$. It is open whether $h'(f) = h(f)$ for any sofic approximation apart from trivial counterexamples. We can also express these quantities as

$$h(f) = \inf_{\varepsilon} \limsup_{n \rightarrow \infty} H(G_n, \varepsilon, \lfloor \varepsilon^{-1} \rfloor) \text{ and } h'(f) = \inf_{\varepsilon} H(\varepsilon, \lfloor \varepsilon^{-1} \rfloor).$$

The quantities $h(f)$ and $h'(f)$ are isomorphism invariants in the abstract ergodic theoretic sense.

Remark. Sofic entropy can be defined in a more general setting. Namely, let Q be a locally finite vertex transitive graph. Let o be any vertex of it. Assume that (Q, o) is a Benjamini-Schramm limit of finite graphs. Let f be a random coloring of $V(Q)$ with C such that the distribution of f is invariant under all automorphisms of Q . We would like to define the sofic entropy of f the same way as above. The only problematic point is that in the definition of (ε, r) -approximation we need to identify $B_r(G, v)$ with $B_r(Q, o)$. But $B_r(Q, o)$ might have non-trivial automorphisms, in which case there are more than one possible identifications and it is not clear which we should choose. If all the automorphisms $B_r(Q, o)$ can be extended to an automorphism of Q , then we can choose any identification, because they all give the same total variation distance. But if $B_r(Q, o)$ has other automorphisms then things get more complicated. However, one can overcome these difficulties and get a sensible notion of sofic entropy [1]. Here we do not give the details, we just mention that Theorem 2.6 stated in the next subsection can be extended to this more general setting.

2.7. Our main theorems. Let E be a countable set, and T be a positive contraction on $\ell^2(E)$. Let c be a $[0, 1]$ labeling of E . For $e \in E$ let $I(e)$ be the indicator of the event that $e \in B^T$. For $e \in E$ we define

$$\bar{h}(e, c, T) = H(I(e) | \{I(f) | c(f) < c(e)\}).$$

Here, H is the conditional entropy, that is, with the notation

$$g(x) = -x \log x - (1 - x) \log(1 - x),$$

we have

$$H(I(e) | \{I(f) | c(f) < c(e)\}) = \mathbb{E}g(\mathbb{E}[I(e) | \{I(f) | c(f) < c(e)\}]).$$

Moreover, we define

$$\bar{h}(e, T) = \mathbb{E}\bar{h}(e, c, T),$$

where c is an i.i.d. uniform $[0, 1]$ labeling of E .

For a random RGPC (G, o, T) we define

$$\bar{h}(G, o, T) = \mathbb{E} \sum_{k \in K} \bar{h}((o, k), T).$$

If $L \subset K$ and (G, T) is a finite graph-positive-contraction we define $h_L(G, T)$ to be the Shannon entropy of $B^T \cap (V(G) \times L)$.

THEOREM 2.4. *Let (G_n, P_n) be a sequence of finite graph-positive-contractions, such that $\lim_{n \rightarrow \infty} U(G_n, P_n) = (G, o, P)$ for some random RGPC (G, o, P) . Assume that P_1, P_2, \dots are orthogonal projections, and P is an orthogonal projection with probability 1. Let $L \subset K$. Then*

$$\lim_{n \rightarrow \infty} \frac{h_L(G_n, P_n)}{|V(G_n)|} = \bar{h}(G, o, \text{rest}_L(P)).$$

Using Lemma 2.2 we immediately get the following theorem.

THEOREM 2.5. *Let (G_n, P_n) be a **tight** sequence of finite graph-positive-contractions, such that $\lim_{n \rightarrow \infty} U(G_n, P_n) = (G, o, P)$ for some random RGPC (G, o, P) . Assume that P_1, P_2, \dots are orthogonal projections. Let $L \subset K$. Then*

$$\lim_{n \rightarrow \infty} \frac{h_L(G_n, P_n)}{|V(G_n)|} = \bar{h}(G, o, \text{rest}_L(P)).$$

Let Γ be a finitely generated sofic group. A positive contraction T on $\ell^2(\Gamma \times K)$ is called invariant, if for any $\gamma, g_1, g_2 \in \Gamma$ and $k_1, k_2 \in K$ we have

$$\langle T(g_1, g_1), (g_2, k_2) \rangle = \langle T(\gamma^{-1}g_1, k_1), (\gamma^{-1}g_2, k_2) \rangle.$$

For an invariant positive contraction if we regard the random subset B^T as a random coloring with $\{0, 1\}^K$, we see that B^T is an invariant coloring. Thus we can speak about its sofic entropy.

As before let S be a finite generating set of Γ , let e_Γ be the identity of Γ , and $G_\Gamma = \text{Cay}(\Gamma, S)$ be the Cayley-graph of Γ .

THEOREM 2.6. *Let Γ be a finitely generated sofic group. If T is an invariant positive contraction on $\ell^2(\Gamma \times K)$ then we have*

$$h(B^T) = h'(B^T) = \bar{h}(G_\Gamma, e_\Gamma, T)$$

for any sofic approximation of Γ .

Note that we can easily generalize the definition of \bar{h} to any invariant random coloring f . It is known that even in this more general setting \bar{h} is an upper bound on the sofic entropy. However, \bar{h} is not an isomorphism invariant in the ergodic theoretic sense. See [18].

The random ordering idea above was used by Borgs, Chayes, Kahn and Lovász [7] to give the growth of the partition function and entropy of certain Gibbs measures at high temperature on Benjamini-Schramm convergent graph sequences. See also [4].

2.8. An example: Why tightness is necessary? We consider two connected graphs H_1 and H_2 . Let H_1 be the complete graph on 4 vertices, and let H_2 be the graph that is obtained from a star with 3 edges by doubling each edge. Both have 4 vertices and 6 edges. Let T_i be a uniform random spanning tree of H_i , and let P_i be the corresponding 6×6 transfer-current matrix. It is straightforward to check that for any $e \in E(H_i)$ we have $\mathbb{P}(e \in T_i) = \frac{1}{2}$. Thus, in both P_1 and P_2 all the diagonal entries are equal to $\frac{1}{2}$. Now let G_i be the empty graph on the vertex set $E(H_i)$. Then the pairs (G_1, P_1) and (G_2, P_2) are indistinguishable by local sampling, that is, $U(G_1, P_1)$ and $U(G_2, P_2)$ have the same distribution. On the other hand H_1 has 16 spanning trees, and H_2 has only 8 spanning trees. So $|V(G_1)|^{-1}H(B^{P_1}) \neq |V(G_2)|^{-1}H(B^{P_2})$. This shows that the condition of tightness can not be omitted in Theorem 2.5. One could think that this only works, because the graphs G_1 and

G_2 are not connected. But Theorem 2.5 still fails without the assumption of tightness, even if we assume that all the graphs are connected. We sketch the main idea. Let $i \in \{1, 2\}$. For each n we consider a block diagonal matrix $B_{i,n}$, where we have n diagonal blocks each of which equal to P_i . Then we take a connected graph $G_{i,n}$ on $V_{i,n}$ (the set of columns of $B_{i,n}$) in such a way that if two columns are in the same block, then they must be at least at distance $d(n)$ in the graph $G_{i,n}$ for some $d(n)$ tending to infinity. Moreover, we can choose $G_{i,n}$ such that the sequences $(G_{1,n})$ and $(G_{2,n})$ have the same Benjamini-Schramm limit (G, o) . Then both of the sequences $(G_{1,n}, B_{1,n})$ and $(G_{2,n}, B_{2,n})$ have the same limit, namely, $(G, o, \frac{1}{2}I)$. But their asymptotic entropy is different.

3. Unimodularity and conditional determinantal processes.

3.1. *Unimodularity.* We define bi-rooted graph-operators as tuples (G, o, o', T) , where G is a connected graph with degree bound D , $o, o' \in V(G)$ and T is a bounded operator on $\ell^2(V(G) \times K)$. Let $bi\mathcal{RG}\mathcal{O}$ be the set of isomorphism classes of bi-rooted graph-operators. We omit the details how to endow this space with a measurable structure. A random RGO (G, o, T) is called *unimodular*, if for any non-negative measurable function $f : bi\mathcal{RG}\mathcal{O} \rightarrow \mathbb{R}$ we have

$$\mathbb{E} \sum_{v \in V(G)} f(G, o, v, T) = \mathbb{E} \sum_{v \in V(G)} f(G, v, o, T).$$

The next lemma gives some examples of unimodular random RGOs. The proof goes like the one given in [6].

LEMMA 3.1. *If (G, T) is a finite graph-positive-contraction, then $U(G, T)$ is unimodular. The limit of unimodular random RGOs is unimodular.*

Of course the notion of unimodularity can be extended to double/triple (decorated) RGOs. We will use the following consequence of unimodularity.

LEMMA 3.2. *Let (G, o, T, S) be a unimodular random double RGO. Assume that there is a finite B such that $\|T\|, \|S\| < B$ with probability 1. Then*

$$\text{Tr}(G, o, TS) = \text{Tr}(G, o, ST).$$

PROOF. The proof is the same as in [2, Section 5]. □

It has the following consequences.

LEMMA 3.3. *In the following statements we always assume that P and P_i are all orthogonal projections with probability 1.*

1. *Let (G, o, P_1, P_2, U) be a unimodular random triple RGO, such that with probability 1 we have $U \upharpoonright \ker P_1 \equiv 0$ and $U \upharpoonright \text{Im } P_1$ is an isomorphism between $\text{Im } P_1$ and $\text{Im } P_2$. Then*

$$\text{Tr}(G, o, P_1) = \text{Tr}(G, o, P_2).$$

2. *Let (G, o, P_1, P_2, T) be a unimodular random triple RGO, such that with probability 1 we have $\overline{\text{Im } TP_1} = \text{Im } P_2$ and T is injective on $\text{Im } P_1$. Then*

$$\text{Tr}(G, o, P_1) = \text{Tr}(G, o, P_2).$$

3. (*rank-nullity theorem*) Let (G, o, P, P_1, P_2, T) be a unimodular random quadruple RGO, such that with probability 1 we have that $\overline{P_1}$ is the orthogonal projection to $\ker(T \upharpoonright \text{Im } P)$ and P_2 is the orthogonal projection to $\overline{\text{Im}(T \upharpoonright \text{Im } P)}$. Then

$$\text{Tr}(G, o, P) = \text{Tr}(G, o, P_1) + \text{Tr}(G, o, P_2).$$

PROOF. To prove part 1 observe that $P_1 U^* U = P_1$ and $U P_1 U^* = P_2$. Note that all operators have norm at most 1, so from Lemma 3.2

$$\text{Tr}(G, o, P_1) = \text{Tr}(G, o, (P_1 U^*) U) = \text{Tr}(G, o, U (P_1 U^*)) = \text{Tr}(G, o, P_2).$$

To prove part 2 let $T P_1 = U H$ be the unique polar decomposition of $T P_1$, then $(G, o, P_1, P_2, U P_1)$ satisfies the conditions in part 1, so the statement follows. The rather technical details why the polar decomposition is measurable are given in the Appendix. Note that once we established the measurability of U , unimodularity follows from the uniqueness of the decomposition.

To prove part 3 let $H = \text{Im } P \cap (\ker T \upharpoonright \text{Im } P)^\perp$. Let P_H be the orthogonal projection to H , then we have $P = P_1 + P_H$. Therefore, $\text{Tr}(G, o, P) = \text{Tr}(G, o, P_1) + \text{Tr}(G, o, P_H)$. It is also clear that $\overline{\text{Im } T P} = \overline{\text{Im}(T \upharpoonright H)}$ and T is injective on H . Thus part 2 gives us $\text{Tr}(G, o, P_H) = \text{Tr}(G, o, P_2)$. Putting everything together we obtain that

$$\text{Tr}(G, o, P) = \text{Tr}(G, o, P_1) + \text{Tr}(G, o, P_H) = \text{Tr}(G, o, P_1) + \text{Tr}(G, o, P_2).$$

□

3.2. *Conditional determinantal processes.* Let P be an orthogonal projection to a closed subspace H of $\ell^2(E)$. Given $C \subset E$, let $[C]$ be the closed subspace generated by $e \in C$, and let $[C]^\perp$ be the orthogonal complement of it. Note that $[C]^\perp = [E \setminus C]$. We define $P_{/C}$ as the orthogonal projection to the closed subspace $(H \cap [C]^\perp) + [C]$, and $P_{\times C}$ as the orthogonal projection to the closed subspace $H \cap [C]^\perp$. We also define $P_{-C} = I - (I - P)_{/C}$.

PROPOSITION 3.4. We have $P_{/C} = P_{\times C} + P_{[C]}$, where $P_{[C]}$ is the orthogonal projection to $[C]$. In other words $P_{/C} e = e$ for $e \in C$ and $P_{/C} e = P_{\times C} e$ for $e \in E \setminus C$. Moreover, if C_n is an increasing sequence of subsets of E and $C = \cup C_n$, then $P_{/C_n}$ converges to $P_{/C}$ in the strong operator topology. Furthermore, the sequence $\langle P_{\times C_n} e, e \rangle$ is monotone decreasing.

PROOF. The first statement is trivial. To prove the second statement, observe that $P_{\times C_n}$ is a sequence of orthogonal projections to a monotone decreasing sequence of closed subspaces with intersection $\text{Im } P_{\times C}$, so $P_{\times C_n}$ converge to $P_{\times C}$ in the strong operator topology. It is also clear that $P_{[C_n]}$ converge to $P_{[C]}$, so from $P_{/C_n} = P_{\times C_n} + P_{[C_n]}$ the statement follows. To prove the third statement observe that $\langle P_{\times C_n} e, e \rangle = \|P_{\times C_n} e\|_2^2$. So the statement follows again from the fact that $P_{\times C_n}$ is a sequence of orthogonal projections to a monotone decreasing sequence of closed subspaces. □

For $C, D \subset E$ we define $P_{/C-D} = (P_{/C})_{-D}$, and we define $P_{-D/C} = (P_{-D})_{/C}$. We only include the next lemma here to make it easier to compare formulas in [13] with our formulas.

LEMMA 3.5. Let P be an orthogonal projection to a closed subspace H . Then for any $D \subset E$ we have

$$\text{Im } P_{-D} = \overline{H + [D]} \cap [D]^\perp.$$

Moreover, if C and D are disjoint subsets of E , then

$$\text{Im } P_{/C-D} = \overline{(H \cap [C]^\perp) + [C \cup D]} \cap [D]^\perp$$

and

$$\text{Im } P_{-D/C} = \overline{(H + [D]) \cap [C \cup D]^\perp} + [C].$$

If C and D are finite, then the above formulas are true even if we omit the closures.

PROOF. We only prove the first statement. The other statements can be easily deduced from it. Unpacking the definitions we need to prove that

$$((H^\perp \cap [D]^\perp) + [D])^\perp = \overline{H + [D]} \cap [D]^\perp.$$

As a first step observe that $\overline{H + [D]} \cap [D]^\perp = \overline{\text{Im}(P_{[D]^\perp} \upharpoonright H)}$. Indeed, if $x \in \overline{\text{Im}(P_{[D]^\perp} \upharpoonright H)}$, then $x = \lim x_n$, where for all n we have $x_n \in [D]^\perp$ and there is an $y_n \in [D]$ such that $x_n + y_n \in H$. But then $x_n = (x_n + y_n) - y_n \in H + [D]$, which implies that $x \in \overline{H + [D]}$. Clearly $x \in [D]^\perp$, so $x \in \overline{H + [D]} \cap [D]^\perp$.

To prove the other containment let $x \in \overline{H + [D]} \cap [D]^\perp$, then $x = \lim x_n$ where $x_n = y_n + z_n$ with $y_n \in H$ and $z_n \in [D]$. Since $P_{[D]^\perp}$ is continuous, we have

$$x = P_{[D]^\perp} x = \lim P_{[D]^\perp}(y_n + z_n) = \lim P_{[D]^\perp} y_n \in \overline{\text{Im}(P_{[D]^\perp} \upharpoonright H)}.$$

Now it is easy to see that we need to prove that

$$(H^\perp \cap [D]^\perp) + [D] = (\text{Im } P_{[D]^\perp} \upharpoonright H)^\perp.$$

First let $x \in (\text{Im } P_{[D]^\perp} \upharpoonright H)^\perp$. Then for any $h \in H$ we have

$$0 = \langle x, P_{[D]^\perp} h \rangle = \langle P_{[D]^\perp} x, h \rangle,$$

which implies that $P_{[D]^\perp} x \in H^\perp \cap [D]^\perp$. Thus, $x = P_{[D]^\perp} x + P_{[D]} x \in (H^\perp \cap [D]^\perp) + [D]$. To show the other containment let us consider $x = y + z$ such that $y \in H^\perp \cap [D]^\perp$ and $z \in [D]$. Then for any $h \in H$ we have

$$\langle x, P_{[D]^\perp} h \rangle = \langle P_{[D]^\perp} x, h \rangle = \langle y, h \rangle = 0,$$

because $y \in H^\perp$.

For the last statement, see the discussion in the paper [13] after the proof of Corollary 6.4. \square

We have the following lemma. See [13, Equation (6.5)].

LEMMA 3.6. *Let C and D be disjoint finite subsets of E such that $\mathbb{P}[B^P \cap (C \cup D) = C] > 0$. Then $P_{/C-D} = P_{-D/C}$ and conditioned on the event $B^P \cap (C \cup D) = C$, the distribution of B^P is the same as that of $B^{P/C-D}$.*

The lemma above shows why the pairs (C, D) of finite disjoint sets with the property that $\mathbb{P}[B^P \cap (C \cup D) = C] > 0$ are interesting for us. The next proposition gives an equivalent characterization of these pairs.

PROPOSITION 3.7. *Let C and D be disjoint finite subsets of E . Then we have $\mathbb{P}[B^P \cap (C \cup D) = C] > 0$ if and only if $\text{Im } P_{[C]} P = [C]$ and $\text{Im } P_{[D]}(I - P) = [D]$.*

This motivates the following definitions. A (not necessary finite) subset C of E is called *independent* (with respect to P) if $\overline{\text{Im } P_{[C]} P} = [C]$. A subset D of E is called *dually independent* (with respect to P) if $\overline{\text{Im } P_{[D]}(I - P)} = [D]$. A pair (C, D) of subsets of E is called *permitted* (with respect to P) if C and D are disjoint, C is independent and D is dually independent.

We will need the following theorem of Lyons [13, Theorem 7.2].

THEOREM 3.8. *The pair $(B^P, E \setminus B^P)$ is permitted with probability 1.*

We will also need the following statements.

PROPOSITION 3.9. *If (C, D) is permitted, $C' \subset C$ and $D' \subset D$, then (C', D') is permitted.*

PROPOSITION 3.10. *Assume (C, D) is a permitted pair. Then D is dually independent with respect to $P_{/C}$, or equivalently, D is independent with respect to $I - P_{/C}$.*

PROOF. By the definition of a permitted pair $\overline{\text{Im } P_{[D]}(I - P)} = [D]$, so it is enough to show that $\text{Im } P_{[D]}(I - P) \subset \text{Im } P_{[D]}(I - P_{/C})$. Take any $r \in \text{Im } P_{[D]}(I - P)$, then there is x such that $r = P_{[D]}(I - P)x$. Let $y = P_{[C]^\perp}(I - P)x$. We claim that $y \in \text{Im}(I - P_{/C})$, or in other words, y is orthogonal to any element $w \in \text{Im } P_{/C}$. We can write w as $w = w_0 + w_1$, where $w_0 \in \text{Im } P \cap [C]^\perp$ and $w_1 \in [C]$. We have

$$\langle y, w_0 \rangle = \langle P_{[C]^\perp}(I - P)x, w_0 \rangle = \langle (I - P)x, P_{[C]^\perp} w_0 \rangle = \langle (I - P)x, w_0 \rangle = 0,$$

since $w_0 \in \text{Im } P$. Moreover $\langle y, w_1 \rangle = 0$, because $y \in [C]^\perp$ and $w_1 \in [C]$. Thus, $\langle y, w \rangle = 0$, so y is indeed in the image of $I - P_{/C}$, then $P_{[D]}y$ is in the image of $P_{[D]}(I - P_{/C})$. Using that C and D are disjoint $P_{[D]}y = P_{[D]}P_{[C]^\perp}(I - P)x = P_{[D]}(I - P)x = r$. \square

Assume for a moment that E is finite, then $|B^P| = \dim \text{Im } P$ with probability 1. If (C, D) is a permitted pair, then the distribution of $B^{P_{/C-D}}$ is the same as that of B^P conditioned on the event that $B^P \cap (C \cup D) = C$. So $|B^{P_{/C-D}}| = \dim \text{Im } P$ with probability 1. In particular, $\mathbb{E}|B^P| = \mathbb{E}|B^{P_{/C-D}}|$. The next lemma extends this statement to the more general unimodular setting.

LEMMA 3.11. *Let (G, o, P, C, D) be a unimodular random decorated RGPC where P is an orthogonal projection and the pair (C, D) is permitted with probability 1. Then*

$$\text{Tr}(G, o, P) = \text{Tr}(G, o, P_{/C-D}) = \text{Tr}(G, o, P_{-D/C}).$$

This can be obtained from combining Proposition 3.10 and the following lemma.

LEMMA 3.12. *Let (G, o, P, C) be a unimodular random decorated RGPC where P is an orthogonal projection and C is independent with probability 1. Then*

$$\text{Tr}(G, o, P) = \text{Tr}(G, o, P_{/C}).$$

We also have the corresponding dual statement, that is, let (G, o, P, D) be a unimodular random decorated RGPC where P is an orthogonal projection and D is dually independent with probability 1. Then

$$\text{Tr}(G, o, P) = \text{Tr}(G, o, P_{-D}).$$

PROOF. We only need to prove the first statement, because the second one can be obtained by applying the first statement to $I - P$.

Observe that $\ker(P_{[C]} \upharpoonright \operatorname{Im} P) = \operatorname{Im} P_{\times C}$ from the definition of $P_{\times C}$, moreover, $\overline{\operatorname{Im}(P_{[C]} \upharpoonright \operatorname{Im} P)} = [C]$, because C is independent. Applying the rank nullity theorem (Lemma 3.3.3) and then using the fact $P_{/C} = P_{\times C} + P_{[C]}$ from Proposition 3.4 we get that

$$\operatorname{Tr}(G, o, P) = \operatorname{Tr}(G, o, P_{\times C}) + \operatorname{Tr}(G, o, P_{[C]}) = \operatorname{Tr}(G, o, P_{\times C} + P_{[C]}) = \operatorname{Tr}(G, o, P_{/C}).$$

□

The next lemma gives an extension of Lemma 3.6.

LEMMA 3.13. *Let $F \subset E$, and assume that*

$$\langle P_{/B^P \cap F - F \setminus B^P} e, e \rangle = \langle P_{-F \setminus B^P / B^P \cap F} e, e \rangle$$

for all $e \in E$ with probability 1. Then for any finite $A \subset E$ we have

$$\mathbb{P}(A \subset B^P | B^P \upharpoonright F) = \mathbb{P}(A \subset B^{P_{/B^P \cap F - F \setminus B^P}}).$$

PROOF. Let F_1, F_2, \dots be an increasing sequence of finite sets such that their union is F . The crucial step in the proof is the following lemma.

LEMMA 3.14. *Let (C, D) be a permitted pair, such that $C \cup D = F$. Then $\langle P_{/C-D} e, e \rangle \leq \langle P_{-D/C} e, e \rangle$ for all $e \in E$. Now assume that $\langle P_{/C-D} e, e \rangle = \langle P_{-D/C} e, e \rangle$ for all $e \in E$. Let us define $P_n = P_{/C \cap F_n - D \cap F_n}$. Then $B^{P_{/C-D}}$ is the weak limit of B^{P_n} .*

PROOF. Let A be a finite set such that, $A \cap F = \emptyset$, moreover let \mathcal{A} be an upwardly closed subset of 2^A , i.e. if $X \subset Y \subset A$ and $X \in \mathcal{A}$, then $Y \in \mathcal{A}$. Using that determinantal measures have negative associations ([13, Theorem 6.5]) we get the following inequality for $m > n$

$$\mathbb{P}[B^{P_n} \cap A \in \mathcal{A}] = \mathbb{P}[B^{P_{/C \cap F_m - D \cap F_m}} \cap A \in \mathcal{A}] \geq \mathbb{P}[B^{P_{/C \cap F_m - D \cap F_m}} \cap A \in \mathcal{A}].$$

Tending to infinity with m , we get that

$$(3) \quad \mathbb{P}[B^{P_n} \cap A \in \mathcal{A}] \geq \mathbb{P}[B^{P_{/C-D \cap F_n}} \cap A \in \mathcal{A}].$$

To justify this last statement, let \mathcal{U} be the set of orthogonal projections R such that $D \cap F_n$ is dually independent with respect to R . Combining Proposition 3.9 and Proposition 3.10, we obtain that $P_{/C \cap F_m}$ and $P_{/C}$ are all contained in \mathcal{U} . For $R \in \mathcal{U}$, the probability $\mathbb{P}[B^{R-D \cap F_n} \cap A \in \mathcal{A}]$ is a continuous function of $(\langle Re, f \rangle)_{e, f \in A \cup (D \cap F_n)}$. As we proved in Proposition 3.4, $P_{/C \cap F_m}$ tends to $P_{/C}$ in the strong operator topology. Thus,

$$\lim_{m \rightarrow \infty} \mathbb{P}[B^{P_{/C \cap F_m - D \cap F_m}} \cap A \in \mathcal{A}] = \mathbb{P}[B^{P_{/C-D \cap F_n}} \cap A \in \mathcal{A}].$$

This gives us Inequality (3).

Tending to infinity with n we get that

$$\liminf_{n \rightarrow \infty} \mathbb{P}[B^{P_n} \cap A \in \mathcal{A}] \geq \lim_{n \rightarrow \infty} \mathbb{P}[B^{P_{/C-D \cap F_n}} \cap A \in \mathcal{A}] = \mathbb{P}[B^{P_{/C-D}} \cap A \in \mathcal{A}].$$

A similar argument gives that

$$\limsup_{n \rightarrow \infty} \mathbb{P}[B^{P_n} \cap A \in \mathcal{A}] \leq \mathbb{P}[B^{P_{-D/C}} \cap A \in \mathcal{A}]$$

Therefore,

$$(4) \quad \begin{aligned} \mathbb{P}[B^{P-D/C} \cap A \in \mathcal{A}] &\geq \limsup_{n \rightarrow \infty} \mathbb{P}[B^{P_n} \cap A \in \mathcal{A}] \\ &\geq \liminf_{n \rightarrow \infty} \mathbb{P}[B^{P_n} \cap A \in \mathcal{A}] \geq \mathbb{P}[B^{P/C-D} \cap A \in \mathcal{A}]. \end{aligned}$$

These inequalities are in fact true without the assumption $A \cap F = \emptyset$. Indeed, let $A \subset E$ finite and \mathcal{A} be an upwardly closed subset of 2^A . We define $A' = A \setminus F$ and

$$\mathcal{A}' = \{X \subset A' \mid X \cup (A \cap C) \in \mathcal{A}\}.$$

Note that \mathcal{A}' is upwardly closed subset of $2^{A'}$.

Then $\mathbb{P}[B^{P/C-D} \cap A \in \mathcal{A}] = \mathbb{P}[B^{P/C-D} \cap A' \in \mathcal{A}']$. Moreover, for any large enough n , we have $\mathbb{P}[B^{P_n} \cap A \in \mathcal{A}] = \mathbb{P}[B^{P_n} \cap A' \in \mathcal{A}']$. Clearly $A' \cap F = \emptyset$, so we reduced the problem to the already established case.

Choosing $A = \{e\}$ and $\mathcal{A} = \{\{e\}\}$ in (4), we get that $\langle P_{/C-D}e, e \rangle \leq \langle P_{-D/C}e, e \rangle$ for all $e \in E$. Inequality (4) tells us that $B^{P-D/C}$ stochastically dominates $B^{P/C-D}$. But if $\langle P_{/C-D}e, e \rangle = \langle P_{-D/C}e, e \rangle$ for all $e \in E$, then the distribution of $B^{P/C-D}$ and $B^{P-D/C}$ must be the same. Then inequality (4) gives the statement. \square

Let A be any finite set. We define the martingale X_n by

$$X_n = \mathbb{P}[A \subset B^P \mid B^P \upharpoonright F_n] = \mathbb{P}[A \subset B^{P/B^P \cap F_n - F_n \setminus B^P}].$$

Combining the previous lemma with our assumptions on B^P we get that with probability 1 we have $\lim X_n = \mathbb{P}[A \subset B^{P/B^P \cap F - F \setminus B^P}]$. On the other hand we have

$$\lim X_n = \mathbb{P}[A \subset B^P \mid B^P \upharpoonright F].$$

The statement follows. \square

LEMMA 3.15. *Let (G, o, P, F) be a unimodular random decorated RGPC where P is an orthogonal projection with probability 1. Then with probability 1, we have that for any finite set $A \subset V(G) \times K$*

$$\mathbb{P}(A \subset B^P \mid B^P \upharpoonright F) = \mathbb{P}(A \subset B^{P/B^P \cap F - F \setminus B^P}).$$

PROOF. From Lemma 3.14, we have that for all $e \in V(G) \times K$ we have

$$\langle P_{/B^P \cap F - F \setminus B^P}e, e \rangle \leq \langle P_{-F \setminus B^P / B^P \cap F}e, e \rangle.$$

From Lemma 3.11, we have $\text{Tr}(G, o, P_{/B^P \cap F - F \setminus B^P}) = \text{Tr}(G, o, P_{-F \setminus B^P / B^P \cap F})$, which imply that with probability 1 we have $\langle P_{/B^P \cap F - F \setminus B^P}e, e \rangle = \langle P_{-F \setminus B^P / B^P \cap F}e, e \rangle$ for any $e \in \{o\} \times K$, but then it is true for any e from unimodularity. (See [2, Lemma 2.3 (Everything Shows at the Root)].) Therefore, Lemma 3.13 can be applied to get the statement. \square

The lemma above establishes Conjecture 9.1 of [13] in the special unimodular case. Note that this conjecture is false in general as it was pointed out to the author by Russel Lyons. Indeed, it follows from the results of Hecklen and Lyons [11] that for the WUSF on certain trees, conditioning on all edges but one does not (a.s.) give a measure corresponding to an orthogonal projection, because the probability of the remaining edge to be present is in $(0, 1)$ a.s.

3.3. Limit of conditional determinantal processes.

THEOREM 3.16. *Let $(G_n, o_n, P_n, C_n, D_n)$ be a convergent sequence of unimodular random decorated RGPCs with limit (G, o, P, C, D) . Assume that P_n and P are orthogonal projections and (C_n, D_n) and (C, D) are all permitted with probability 1. Then $(G_n, o_n, (P_n)_{/C_n - D_n})$ converges to $(G, o, P_{/C - D})$.*

This will follow from applying the next lemma twice, first for the sequence P_n , then for $I - (P_n)_{/C}$ with D_n in place of C_n . At the second time we need to use Proposition 3.10 to show that the conditions of the lemma are satisfied.

LEMMA 3.17. *Let $(G_n, o_n, P_n, C_n, D_n)$ be a convergent sequence of unimodular random decorated RGPCs with limit (G, o, P, C, D) . Assume that P_n and P are orthogonal projections and C_n, C are all independent with probability 1. Then $(G_n, o_n, (P_n)_{/C_n}, D_n)$ converges to $(G, o, P_{/C}, D)$.*

PROOF. The presence of D_n does not add any extra difficulty to the problem, so for simplicity of notation we will prove the following statement instead:

Let (G_n, o_n, P_n, C_n) be a convergent sequence of unimodular random decorated RGPCs with limit (G, o, P, C) . Assume that P_n and P are orthogonal projections, C_n and C are all independent with probability 1. Then $(G_n, o_n, (P_n)_{/C_n})$ converges to $(G, o, P_{/C})$.

We start by the following lemma.

LEMMA 3.18. *Let (G_n, o_n, P_n, C_n) be a convergent sequence of decorated RGPCs with limit (G, o, P, C) . Assume that P_n and P are orthogonal projections, C_n and C are all independent, and there is an r such that $C_n \subset V(B_r(G_n, o_n)) \times K$ and $C \subset V(B_r(G, o)) \times K$. Then $(G_n, o_n, (P_n)_{\times C_n})$ converges to $(G, o, P_{\times C})$.*

PROOF. Let us choose an orthogonal projection Π from a small neighborhood U of P . If this neighborhood is small enough, then C is independent with respect to Π . For $c \in C$, we have $\Pi_{\times \{c\}} e = \Pi e - \frac{\langle \Pi e, c \rangle}{\langle \Pi c, c \rangle} \Pi c$. Indeed, clearly $\Pi e - \frac{\langle \Pi e, c \rangle}{\langle \Pi c, c \rangle} \Pi c \in \text{Im } \Pi \cap [\{c\}]^\perp$, moreover with the notation $\alpha = \frac{\langle \Pi e, c \rangle}{\langle \Pi c, c \rangle}$ for any $w \in \text{Im } \Pi \cap [\{c\}]^\perp$ we have

$$\langle w, e - (\Pi e - \alpha \Pi c) \rangle = \langle w, (I - \Pi)e \rangle + \langle w, \alpha \Pi c \rangle = \langle \Pi w, \alpha c \rangle = \langle w, \alpha c \rangle = 0.$$

By induction we get that

$$\Pi_{\times C} e = \Pi e - \sum_{c \in C} \alpha_{c,e} \Pi c.$$

Here $\alpha_{c,e}$ is a continuous function of $(\langle \Pi x, y \rangle)_{x,y \in C \cup \{e\}}$ in the neighborhood U . The statement can be deduced using this. \square

From compactness every subsequence of $(G_n, o_n, P_n, (P_n)_{/C_n}, C_n)$ has a convergent subsequence, so it is enough to prove the following lemma.

LEMMA 3.19. *Let (G_n, o_n, P_n, C_n) be a convergent sequence of unimodular random decorated RGPCs with limit (G, o, P, C) . Assume that P_n and P are orthogonal projections, C_n and C are all independent with probability 1. If $(G_n, o_n, P_n, (P_n)_{/C_n}, C_n)$ converges to (G, o, P, Q, C) , then (G, o, Q) has the same distribution as $(G, o, P_{/C})$.*

PROOF. Using Skorokhod's representation theorem we can find a coupling of $(G_n, o_n, P_n, (P_n)_{/C_n}, C_n)$ and (G, o, P, Q, C) such that $\lim_{n \rightarrow \infty} (G_n, o_n, P_n, (P_n)_{/C_n}, C_n) = (G, o, P, Q, C)$ with probability 1. By definition there is a random sequence r_1, r_2, \dots such that $\lim_{n \rightarrow \infty} r_n = \infty$ with probability 1, and there is a root preserving graph isomorphism ψ_n from $B_{r_n}(G, o)$ to $B_{r_n}(G_n, o_n)$ such that $\bar{\psi}_n(C \cap (B_{r_n}(G, o) \times K)) = C_n \cap (B_{r_n}(G_n, o_n) \times K)$, where $\bar{\psi}_n(v, k) = (\psi_n(v), k)$ and with probability 1 for each $e, f \in V(G) \times K$ we have

$$\lim_{n \rightarrow \infty} \langle P_n \bar{\psi}_n e, \bar{\psi}_n f \rangle = \langle P e, f \rangle,$$

and

$$\lim_{n \rightarrow \infty} \langle (P_n)_{/C_n} \bar{\psi}_n e, \bar{\psi}_n f \rangle = \langle Q e, f \rangle.$$

Of course, $\bar{\psi}_n e$ only makes sense if n is large enough.

Let us define $C_n(r) = C_n \cap (B_r(G_n, o_n) \times K)$ and $C(r) = C \cap (B_r(G, o) \times K)$.

Lemma 3.18 gives us that for any r we have

$$(5) \quad \lim_{n \rightarrow \infty} \langle (P_n)_{\times C_n(r)} \bar{\psi}_n(e), \bar{\psi}_n(f) \rangle = \langle P_{\times C(r)} e, f \rangle.$$

Note that $\text{Im } P_{\times C(r)}$ is a decreasing sequence of subspaces with intersection $\text{Im } P_{\times C}$. So $P_{\times C(r)}$ converges to $P_{\times C}$ in the strong operator topology.

In particular, for any $e, f \in V(G) \times K$, we have

$$(6) \quad \lim_{r \rightarrow \infty} \langle P_{\times C(r)} e, f \rangle = \langle P_{\times C} e, f \rangle,$$

and

$$(7) \quad \lim_{r \rightarrow \infty} \langle (P_n)_{\times C_n(r)} \bar{\psi}_n(e), \bar{\psi}_n(f) \rangle = \langle (P_n)_{\times C_n} \bar{\psi}_n(e), \bar{\psi}_n(f) \rangle.$$

We need the following elementary fact.

LEMMA 3.20. *Let $a(r, n)$ be non-negative real numbers, such that for any fixed n , the sequence $a(r, n)$ is monotone decreasing in r . Let $A_n = \lim_{r \rightarrow \infty} a(r, n)$, assume that for any fixed r the limit $B_r = \lim_{n \rightarrow \infty} a(r, n)$ exists. Then $\lim_{n \rightarrow \infty} A_n \leq \lim_{r \rightarrow \infty} B_r$ if these limits exist.*

Note that if $e = f$ then the limits in (6) and (7) are decreasing limits as we observed in Proposition 3.4. So the previous lemma combined with equation (5) gives us that for any $e \in V(G) \times K$ we have

$$\lim_{n \rightarrow \infty} \langle (P_n)_{\times C_n} \bar{\psi}_n e, \bar{\psi}_n e \rangle \leq \langle P_{\times C} e, e \rangle.$$

Combining this with Proposition 3.4, we get that

$$(8) \quad \langle Q e, e \rangle = \lim_{n \rightarrow \infty} \langle (P_n)_{/C_n} \bar{\psi}_n e, \bar{\psi}_n e \rangle \leq \langle P_{/C} e, e \rangle.$$

On the other hand, from Lemma 3.12, we know that

$$\begin{aligned} \mathbb{E} \sum_{k \in K} \langle Q(o, k), (o, k) \rangle &= \text{Tr}(G, o, Q) = \lim_{n \rightarrow \infty} \text{Tr}(G_n, o_n, (P_n)_{/C_n}) \\ &= \lim_{n \rightarrow \infty} \text{Tr}(G_n, o_n, P_n) = \text{Tr}(G, o, P) = \text{Tr}(G, o, P_{/C}) \\ &= \mathbb{E} \sum_{k \in K} \langle P_{/C}(o, k), (o, k) \rangle. \end{aligned}$$

From this and inequality (8) we get that $\langle Q(o, k), (o, k) \rangle = \langle P_{/C}(o, k), (o, k) \rangle$ for all $k \in K$ with probability 1, so from unimodularity ([2, Lemma 2.3 (Everything shows at the root)]) it follows that

$$(9) \quad \langle Qe, e \rangle = \lim_{n \rightarrow \infty} \langle (P_n)_{/C_n} \bar{\psi}_n e, \bar{\psi}_n e \rangle = \langle P_{/C} e, e \rangle$$

for all $e \in V(G) \times K$ with probability 1.

Now we prove that with probability 1 for every $e, f \in V(G) \times K$ we have $\langle Qe, f \rangle = \langle P_{/C} e, f \rangle$. This is clear if $e \in C$, because in that case $P_{/C} e = Qe = e$. So assume that $e \notin C$, then

$$\begin{aligned} |\langle P_{/C} e, f \rangle - \langle Qe, f \rangle| &= |\langle P_{\times C} e, f \rangle - \langle Qe, f \rangle| \\ &\leq |\langle P_{\times C} e, f \rangle - \langle P_{\times C(r)} e, f \rangle| \\ &\quad + |\langle P_{\times C(r)} e, f \rangle - \langle (P_n)_{\times C_n(r)} \bar{\psi}_n e, \bar{\psi}_n f \rangle| \\ &\quad + |\langle (P_n)_{\times C_n(r)} \bar{\psi}_n e, \bar{\psi}_n f \rangle - \langle (P_n)_{\times C_n} \bar{\psi}_n e, \bar{\psi}_n f \rangle| \\ &\quad + |\langle (P_n)_{\times C_n} \bar{\psi}_n e, \bar{\psi}_n f \rangle - \langle Qe, f \rangle|. \end{aligned}$$

Pick any $\varepsilon > 0$. If we choose a large enough r , then $|\langle P_{\times C} e, f \rangle - \langle P_{\times C(r)} e, f \rangle| < \varepsilon$ and $|\langle P_{\times C(r)} e, e \rangle - \langle P_{\times C} e, e \rangle| < \varepsilon$ from equation (6). Fix such an r . Then if n is large enough $|\langle P_{\times C(r)} e, f \rangle - \langle (P_n)_{\times C_n(r)} \bar{\psi}_n e, \bar{\psi}_n f \rangle| < \varepsilon$ from equation (5), and also $|\langle (P_n)_{\times C_n} \bar{\psi}_n e, \bar{\psi}_n f \rangle - \langle Qe, f \rangle| < \varepsilon$, because of Proposition 3.4 and the fact that $e \notin C$. Finally, observing that $(P_n)_{\times C_n(r)} - (P_n)_{\times C_n}$ is an orthogonal projection, we have

$$\begin{aligned} &|\langle (P_n)_{\times C_n(r)} \bar{\psi}_n e, \bar{\psi}_n f \rangle - \langle (P_n)_{\times C_n} \bar{\psi}_n e, \bar{\psi}_n f \rangle| \\ &\leq \| (P_n)_{\times C_n(r)} \bar{\psi}_n e - (P_n)_{\times C_n} \bar{\psi}_n e \|_2 \\ &= \sqrt{\langle (P_n)_{\times C_n(r)} \bar{\psi}_n e - (P_n)_{\times C_n} \bar{\psi}_n e, \bar{\psi}_n e \rangle} \\ &\leq \left(|\langle (P_n)_{\times C_n(r)} \bar{\psi}_n e, \bar{\psi}_n e \rangle - \langle P_{\times C(r)} e, e \rangle| \right. \\ &\quad \left. + |\langle P_{\times C(r)} e, e \rangle - \langle P_{\times C} e, e \rangle| \right. \\ &\quad \left. + |\langle P_{\times C} e, e \rangle - \langle (P_n)_{\times C_n} \bar{\psi}_n e, \bar{\psi}_n e \rangle| \right)^{\frac{1}{2}} \end{aligned}$$

Now, for a large enough n we have $|\langle (P_n)_{\times C_n(r)} \bar{\psi}_n e, \bar{\psi}_n e \rangle - \langle P_{\times C(r)} e, e \rangle| < \varepsilon$ from equation (5) and $|\langle P_{\times C} e, e \rangle - \langle (P_n)_{\times C_n} \bar{\psi}_n e, \bar{\psi}_n e \rangle| = |\langle P_{/C} e, e \rangle - \langle (P_n)_{/C_n} \bar{\psi}_n e, \bar{\psi}_n e \rangle| < \varepsilon$ from line (9). Finally, $|\langle P_{\times C(r)} e, e \rangle - \langle P_{\times C} e, e \rangle| < \varepsilon$ follows from the choice of r . Putting everything together, $|\langle P_{/C} e, f \rangle - \langle Qe, f \rangle| < 3\varepsilon + \sqrt{3\varepsilon}$, so Lemma 3.19 follows. \square

This completes the proof of Lemma 3.17 and Theorem 3.16. \square

4. The proof of Theorem 2.4. First we observe that we may assume that $|V(G_n)| \rightarrow \infty$. If not, then we can take a large $m = m(n)$ and replace G_n with m disjoint copies of G_n , and P_n with the m fold direct sum of copies of P_n .

Let (G, P) be a finite graph-positive-contraction, where P is an orthogonal projection. Let $m = |V(G) \times L|$. Fix an ordering e_1, e_2, \dots, e_m of the element of $V(G) \times L$. Let $E_i = \{e_1, e_2, \dots, e_i\}$. For $e \in V(G) \times L$ let $I(e)$ be the indicator of the event that $e \in B^P$. Let $g(x) = -x \log x - (1-x) \log(1-x)$. Using the chain rule for the conditional entropy and Lemma 3.6 we obtain that

$$\begin{aligned}
h_L(G, P) &= H(I(e_1), I(e_2), \dots, I(e_m)) \\
&= \sum_{i=0}^{m-1} H(I(e_{i+1}) | I(e_1), I(e_2), \dots, I(e_i)) \\
&= \sum_{i=0}^{m-1} \sum_{C \subset E_i} \mathbb{P}[B^P \cap E_i = C] g(\mathbb{P}[e_{i+1} \in B^P | B^P \cap E_i = C]) \\
&= \sum_{i=0}^{m-1} \mathbb{E} g(\mathbb{P}[e_{i+1} \in B^P /_{(E_i \cap B^P) - (E_i \setminus B^P)}]) \\
&= \sum_{i=0}^{m-1} \mathbb{E} g(\langle P /_{(E_i \cap B^P) - (E_i \setminus B^P)} e_{i+1}, e_{i+1} \rangle).
\end{aligned}$$

Here expectation is over the random choice of B^P .

Instead of a fixed ordering of $V(G) \times L$ we can choose a uniform random ordering. Taking expectation we get that

$$h_L(G, P) = \sum_{i=0}^{m-1} \mathbb{E} g(\langle P /_{(E_i \cap B^P) - (E_i \setminus B^P)} e_{i+1}, e_{i+1} \rangle),$$

where expectation is over the random choice of $E_i = \{e_1, e_2, \dots, e_i\}$ and B^P . Note that $g(0) = g(1) = 0$, so

$$g(\langle P /_{(E_i \cap B^P) - (E_i \setminus B^P)} e, e \rangle) = 0$$

whenever $e \in E_i$. Also note that e_{i+1} is a uniform random element of $(V(G) \times L) \setminus E_i$. From these it follows that if e is a uniform random element of $V(G) \times L$ independent of E_i , then

$$(10) \quad \frac{m}{m-i} \mathbb{E} g(\langle P /_{(E_i \cap B^P) - (E_i \setminus B^P)} e, e \rangle) = \mathbb{E} g(\langle P /_{(E_i \cap B^P) - (E_i \setminus B^P)} e_{i+1}, e_{i+1} \rangle) \leq \log 2.$$

Thus,

$$h_L(G, P) = \sum_{i=0}^{m-1} \frac{m}{m-i} \mathbb{E} g(\langle P /_{(E_i \cap B^P) - (E_i \setminus B^P)} e, e \rangle).$$

Let $(G, o, P) = U(G, P)$. Then

$$h_L(G, P) = \sum_{i=0}^{m-1} \frac{m}{m-i} \frac{1}{|L|} \mathbb{E} \sum_{\ell \in L} g(\langle P /_{(E_i \cap B^P) - (E_i \setminus B^P)} (o, \ell), (o, \ell) \rangle).$$

So

$$\frac{h_L(G, P)}{|V(G)|} = \sum_{i=0}^{m-1} \frac{1}{m-i} \mathbb{E} \sum_{\ell \in L} g(\langle P /_{(E_i \cap B^P) - (E_i \setminus B^P)} (o, \ell), (o, \ell) \rangle).$$

For $t \in [0, 1)$ we define

$$H_t(G, P) = \mathbb{E} \sum_{\ell \in L} g(\langle P /_{(E_i \cap B^P) - (E_i \setminus B^P)} (o, \ell), (o, \ell) \rangle),$$

where $i = \lfloor tm \rfloor$, and E_i is a uniform random i element subset of $V(G) \times L$ independent of B^P and o . For $i = 0, 1, \dots, m-1$, we have

$$\frac{1}{m-i} \mathbb{E} \sum_{\ell \in L} g(\langle P_{/(E_i \cap B^P) - (E_i \setminus B^P)}(o, \ell), (o, \ell) \rangle) = \int_{i/m}^{(i+1)/m} \frac{m}{m - \lfloor tm \rfloor} H_t(G, P) dt.$$

Therefore

$$(11) \quad \frac{h_L(G, P)}{|V(G)|} = \int_0^1 \frac{m}{m - \lfloor tm \rfloor} H_t(G, P) dt.$$

Let $m_n = |V(G_n) \times L|$. Recall that we observed at the beginning of the proof that we may assume that $|V(G_n)| \rightarrow \infty$. So we assume this.

LEMMA 4.1. *Let (G_n, P_n) be the sequence given in the statement of the theorem. For any $t \in [0, 1)$ we have*

$$\lim_{n \rightarrow \infty} H_t(G_n, P_n) = \mathbb{E} \sum_{\ell \in L} g(\langle P_{/(E_t \cap B^P) - (E_t \setminus B^P)}(o, \ell), (o, \ell) \rangle),$$

where E_t is a Bernoulli(t) percolation of the set $V(G) \times L$ independent of B^P . Consequently,

$$\lim_{n \rightarrow \infty} \frac{m_n}{m_n - \lfloor tm_n \rfloor} H_t(G_n, P_n) = \frac{1}{1-t} \mathbb{E} \sum_{\ell \in L} g(\langle P_{/(E_t \cap B^P) - (E_t \setminus B^P)}(o, \ell), (o, \ell) \rangle).$$

PROOF. From Proposition 2.1 we have $(G_n, o_n, P_n, B^{P_n}) \rightarrow (G, o, P, B^P)$. It is straightforward to show that $(G_n, o_n, P_n, E_{\lfloor tm_n \rfloor}) \rightarrow (G, o, P, E_t)$, here $m_n = |V(G) \times L|$ and $E_{\lfloor tm_n \rfloor}$ is a uniform $\lfloor tm_n \rfloor$ element subset of $V(G) \times L$ independent of B^{P_n} . Then it follows that $(G_n, o_n, P_n, E_{\lfloor tm_n \rfloor}, B^{P_n}) \rightarrow (G, o, P, E_t, B^P)$. But then with the notations $C_n = E_{\lfloor tm_n \rfloor} \cap B^{P_n}$, $C = E_t \cap B^P$, $D_n = E_{\lfloor tm_n \rfloor} \setminus B^{P_n}$ and $D = E_t \setminus B^P$ we have $(G_n, o_n, P_n, C_n, D_n) \rightarrow (G, o, P, C, D)$. It follows from Theorem 3.8 and Proposition 3.9, that (C_n, D_n) and (C, D) are all permitted with probability 1. It is also clear that $(G_n, o_n, P_n, C_n, D_n)$ are unimodular. Thus applying Theorem 3.16 we get that $(G_n, o_n, (P_n)_{/C_n - D_n})$ converge to $(G, o, P_{/C - D})$. We define the continuous map $f : \mathcal{RGP}C \rightarrow \mathbb{R}$ as $f(G, o, P) = \sum_{\ell \in L} g(\langle P(o, \ell), (o, \ell) \rangle)$. Then from the definition of weak* convergence we get that

$$\lim_{n \rightarrow \infty} \mathbb{E} f(G_n, o_n, (P_n)_{/C_n - D_n}) = \mathbb{E} f(G, o, P_{/C - D})$$

and this is exactly what we needed to prove. \square

From (10) we have $\frac{m_n}{m_n - \lfloor tm_n \rfloor} H_t(G, P) \leq \log 2$ for any n and t . So combining equation (11) and Lemma 4.1 with the dominated convergence theorem we get that

$$(12) \quad \begin{aligned} \lim_{n \rightarrow \infty} \frac{h_L(G_n, P_n)}{|V(G_n)|} &= \lim_{n \rightarrow \infty} \int_0^1 \frac{m_n}{m_n - \lfloor tm_n \rfloor} H_t(G_n, P_n) dt \\ &= \int_0^1 \lim_{n \rightarrow \infty} \frac{m_n}{m_n - \lfloor tm_n \rfloor} H_t(G_n, P_n) dt \\ &= \int_0^1 \frac{1}{1-t} \mathbb{E} \sum_{\ell \in L} g(\langle P_{/(E_t \cap B^P) - (E_t \setminus B^P)}(o, \ell), (o, \ell) \rangle) dt \end{aligned}$$

$$\begin{aligned}
&= \int_0^1 \frac{1}{1-t} \sum_{\ell \in L} \mathbb{P}[(o, \ell) \notin E_t] \mathbb{E} [g(\langle P_{/(E_t \cap B^P) - (E_t \setminus B^P)}(o, \ell), (o, \ell) \rangle) | (o, \ell) \notin E_t] dt \\
&= \int_0^1 \sum_{\ell \in L} \mathbb{E} [g(\langle P_{/(E_t \cap B^P) - (E_t \setminus B^P)}(o, \ell), (o, \ell) \rangle) | (o, \ell) \notin E_t] dt.
\end{aligned}$$

Here we used the law of total expectation and the fact that $g(\langle B^P_{/(E_t \cap B^P) - (E_t \setminus B^P)}(o, \ell), (o, \ell) \rangle) = 0$ whenever $(o, \ell) \in E_t$. Let c be an i.i.d. uniform $[0, 1]$ labeling of $V(G) \times L$. Observe that conditioned on the event $(o, \ell) \notin E_t$ the distribution of E_t is the same as the distribution of $\{e \in V(G) \times L | c(e) < c(o, \ell)\}$ conditioned on $c(o, \ell) = t$. Let $I(e)$ be the indicator of the event $e \in B^{\text{rest}_L P}$. From Lemma 3.15 we get for $\ell \in L$

$$\begin{aligned}
&\int_0^1 \mathbb{E} [g(\langle P_{/(E_t \cap B^P) - (E_t \setminus B^P)}(o, \ell), (o, \ell) \rangle) | (o, \ell) \notin E_t] dt \\
&= \int_0^1 \mathbb{E} [g(\mathbb{E}(I(o, \ell) | \{I(f) | f \in E_t\})) | (o, \ell) \notin E_t] dt \\
&= \int_0^1 \mathbb{E} [g(\mathbb{E}(I(o, \ell) | \{I(f) | c(f) < c(o, \ell)\})) | c(o, \ell) = t] dt \\
&= \mathbb{E} [g(\mathbb{E}(I(o, \ell) | \{I(f) | c(f) < c(o, \ell)\}))] = \mathbb{E} \bar{h}((o, \ell), \text{rest}_L P).
\end{aligned}$$

Combining this with equation (12) we get Theorem 2.4.

5. Extension of Theorem 2.4 to positive contractions. To state the extension of Theorem 2.4 we need another tightness notion. Let $K_0 \supset K$ be finite. A random RGPC (G_0, o_0, T_0) with support set K_0 is called a K_0 -extension of the random RGPC (G, o, T) with support set K , if $(G_0, o_0, \text{rest}_K(T_0))$ has the same distribution as (G, o, T) . We say that the extension is *tight* if T_0 is an orthogonal projection with probability 1. A finite graph-positive-contraction (G_0, T_0) with support set K_0 is called a K_0 -extension of the finite graph-positive-contraction (G, T) with support set K , if $G = G_0$ and $\text{rest}_K T_0 = T$. We say that the extension is *tight*, if T_0 is an orthogonal projection.

Given a sequence of finite graph-positive-contractions (G_n, T_n) with support K and a random RGPC (G, o, T) with support set K , we say that $\lim U(G_n, T_n) = (G, o, T)$ *p-tightly*, if there is a finite $K_0 \supset K$ and there are tight K_0 -extensions (G_n, P_n) of (G_n, T_n) and a tight K_0 -extension (G, o, P) of (G, o, T) such that $\lim U(G_n, P_n) = (G, o, P)$.

With these definitions we have the following extension of Theorem 2.4.

THEOREM 5.1. *Let (G_n, T_n) be a sequence of finite graph-positive-contractions such that $\lim U(G_n, T_n) = (G, o, T)$ p-tightly for some random RGPC (G, o, T) . Then*

$$\lim_{n \rightarrow \infty} \frac{h_L(G_n, T_n)}{|V(G_n)|} = \bar{h}_L(G, o, T).$$

PROOF. By the definition of tight convergence, there is a finite $K_0 \supset K$ and there are tight K_0 -extensions (G_n, P_n) of (G_n, T_n) and a tight K_0 -extension (G, o, P) of (G, o, T) such that $\lim U(G_n, P_n) = (G, o, P)$. Note that the distribution of B^{T_n} is the same as $B^{P_n} \cap (V(G) \times K)$. So $h_L(G_n, T_n) = h_L(G_n, P_n)$. Similarly, $\bar{h}_L(G, o, T) = \bar{h}_L(G, o, P)$. So from Theorem 2.4

$$\lim_{n \rightarrow \infty} \frac{h_L(G_n, T_n)}{|V(G_n)|} = \lim_{n \rightarrow \infty} \frac{h_L(G_n, P_n)}{|V(G_n)|} = \bar{h}_L(G, o, P) = \bar{h}_L(G, o, T).$$

□

We do not know whether the condition of p -tightness can be replaced with tightness in the theorem above.

Later we will need the following proposition.

PROPOSITION 5.2. Let $K \subset K_0$, such that $|K_0| = 2|K|$. Any finite graph-positive-contraction (G, T) has a tight K_0 -extension (G, P) .

PROOF. This is well known, see for example [13, Chapter 9]. We include the proof for the reader's convenience. Let $q(x) = \sqrt{x(1-x)}$ on the interval $[0, 1]$ and 0 otherwise. Using functional calculus we can define $q(T)$, for every positive contraction. Then the block matrix

$$P = \begin{pmatrix} T & q(T) \\ q(T) & I - T \end{pmatrix}$$

gives the desired operator. \square

The K_0 -extension given in the previous lemma will be called the *standard K_0 -extension* of (G, T) . The standard K_0 -extension of a random RGPC is defined in the analogous way.

6. Sofic entropy: The proof of Theorem 2.6. Note that for any graph G the set of random $\{0, 1\}^K$ colorings of $V(G)$ can be identified with the set of random subsets of $V(G) \times K$. In this proof we use the random subset terminology.

As we mentioned in Subsection 2.7, the inequality $h'(B^T) \leq \bar{h}(G_\Gamma, e_\Gamma, T)$ is well known, but we give the proof for completeness.

Let G be a graph, and F be a random subset of $V(G) \times K$. Let c be a $[0, 1]$ labeling of $V(G) \times K$. For $e \in V(G) \times K$ let $I(e)$ be the indicator of the event that $e \in F$. For $(v, k) \in V(G) \times K$ we define

$$\bar{h}((v, k), c, F) = H(I(v, k) | \{I(v', k') | c(v', k') < c(v, k)\}).$$

We also define

$$\bar{h}((v, k), F) = \mathbb{E} \bar{h}((v, k), c, F),$$

where c is an i.i.d. uniform $[0, 1]$ labeling of $V(G) \times K$.

Moreover, if r is an integer, then we define

$$\bar{h}_r((v, k), c, F) = H(I(v, k) | \{I(v', k') | c(v', k') < c(v, k) \text{ and } (v', k') \in B_r(G, v) \times K\})$$

and

$$\bar{h}_r((v, k), F) = \mathbb{E} \bar{h}_r((v, k), c, F),$$

where c is an i.i.d. uniform $[0, 1]$ labeling of $V(G) \times K$.

Comparing these definitions with the definitions given in Subsection 2.7, we see that if $F = B^T$ for some positive contraction T , then $\bar{h}((v, k), F) = \bar{h}((v, k), T)$. Thus, it is justified the use the same symbol in both cases.

If c is a $[0, 1]$ -labeling such that the labels are pairwise distinct and G is finite, the chain rule of conditional entropy gives us

$$H(F) = \sum_{(v, k) \in V(G) \times K} \bar{h}((v, k), c, F).$$

Taking expectation over c we get that

$$H(F) = \sum_{(v, k) \in V(G) \times K} \bar{h}((v, k), F).$$

Or, alternatively,

$$\frac{H(F)}{|V(G)|} = \mathbb{E} \sum_{k \in K} \bar{h}((o, k), F)$$

where o is a uniform random vertex of $V(G)$.

Combining this with the well known monotonicity properties of conditional entropies, for any integer r , we have

$$\frac{H(F)}{|V(G)|} = \mathbb{E} \sum_{k \in K} \bar{h}((o, k), F) \leq \mathbb{E} \sum_{k \in K} \bar{h}_r((o, k), F).$$

Note that $\bar{h}_r((o, k), F)$ only depends on the distribution of $F \cap (B_r(G, o) \times K)$. Therefore, if F is an (ε, r) approximation, then we have

$$\frac{H(F)}{|V(G)|} \leq \mathbb{E} \sum_{k \in K} \bar{h}_r((o, k), F) \leq \sum_{k \in K} \bar{h}_r((e_\Gamma, k), B^T) + \eta_r(\varepsilon),$$

where $\eta_r(\varepsilon)$ does not depend on G , and $\eta_r(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$. In particular,

$$H(\varepsilon, r) \leq \sum_{k \in K} \bar{h}_r((e_\Gamma, k), B^T) + \eta_r(\varepsilon)$$

tending to 0 with ε we obtain that

$$\inf_{\varepsilon} H(\varepsilon, r) \leq \sum_{k \in K} \bar{h}_r((e_\Gamma, k), B^T).$$

But we have

$$\lim_{r \rightarrow \infty} \sum_{k \in K} \bar{h}_r((e_\Gamma, k), B^T) = \sum_{k \in K} \bar{h}((e_\Gamma, k), B^T).$$

Thus tending to infinity with r we get

$$h'(B^T) \leq \sum_{k \in K} \bar{h}((e_\Gamma, k), B^T) = \bar{h}(G_\Gamma, e_\Gamma, T).$$

Now let G_1, G_2, \dots be a sequence of finite S -labeled Schreier graphs Benjamini-Schramm converging to (G_Γ, e_Γ) . Let $K \subset K_0$, such that $|K_0| = 2|K|$. Let P be the standard K_0 -extension of T . Then it is clear that P is an invariant operator on $\ell^2(V(G_\Gamma) \times K_0)$.

LEMMA 6.1. *There is a sequence of positive contractions R_n on $\ell^2(V(G_n) \times K_0)$ such that $\lim_{n \rightarrow \infty} U(G_n, R_n) = (G_\Gamma, e_\Gamma, P)$. Moreover, the spectral measures $\mu_n = \mu_{U(G_n, R_n)}$ weakly converge to $\mu = \mu_{(G_\Gamma, e_\Gamma, P)} = |K|(\delta_0 + \delta_1)$.*

PROOF. One can easily define a metric d' on $\mathcal{P}(\mathcal{RGPC})$ such that for any sequence of positive contractions R_n on $\ell^2(V(G_n) \times K_0)$, we have that $\lim_{n \rightarrow \infty} d'(U(G_n, R_n), (G_\Gamma, e_\Gamma, P)) = 0$ if and only if $\lim_{n \rightarrow \infty} U(G_n, R_n) = (G_\Gamma, e_\Gamma, P)$ and μ_n weakly converge to μ .

Thus if the required sequence does not exist, then there is an $\varepsilon > 0$ and an infinite sequence $n_1 < n_2 < \dots$ such that $d'(U(G_{n_i}, R_{n_i}), (G_\Gamma, e_\Gamma, P)) \geq \varepsilon$ for any i and any positive contractions R_{n_i} on $\ell^2(V(G_{n_i}) \times K_0)$.

We will now use the results of Lyons and Thom [15]. In their paper they are using ultralimits. However, by passing to a subsequence we may replace ultralimits by actual limits. Thus [15, Proposition 4.4, Lemma 4.7 and Remark 4.3] provide us a subsequence (m_i) of

(n_i) and positive contractions R_{m_i} on $\ell^2(V(G_{m_i}) \times K_0)$, such that $\lim_{i \rightarrow \infty} U(G_{m_i}, R_{m_i}) = (G_\Gamma, e_\Gamma, P)$ and μ_{m_i} weakly converge to μ . Indeed, [15, Proposition 4.4] gives us the convergence $\lim_{i \rightarrow \infty} U(G_{m_i}, R_{m_i}) = (G_\Gamma, e_\Gamma, P)$ and [15, Proposition 4.7] is used to make sure R_{m_i} is indeed a positive contraction. Finally, the convergence of spectral measures follows from [15, Remark 4.3].

Then $\lim_{i \rightarrow \infty} d'(U(G_{m_i}, R_{m_i}), (G_\Gamma, e_\Gamma, P)) = 0$, which contradicts to the choice of the subsequence (n_i) .

Finally, observe that

$$\mathrm{Tr}(G_\Gamma, e_\Gamma, P) = \mathrm{Tr}(G_\Gamma, e_\Gamma, T) + \mathrm{Tr}(G_\Gamma, e_\Gamma, I - T) = |K|,$$

so the spectral measure μ is indeed equal to $|K|(\delta_0 + \delta_1)$. \square

Note that R_n is not necessary an orthogonal projection. Now we modify R_n slightly to get an orthogonal projection. Let us define

$$w(x) = \begin{cases} x & \text{for } 0 \leq x < \frac{1}{2}, \\ x - 1 & \text{for } \frac{1}{2} \leq x \leq 1 \end{cases}$$

Note that w is not continuous, but w^2 is continuous. Let $(v_i)_{i=1}^{|V(G_n) \times K_0|}$ be an orthonormal basis of $\ell^2(V(G_n) \times K_0)$ consisting of eigenvectors of R_n , such that $R_n v_i = \lambda_i v_i$. Let $w(R_n)$ be the unique operator, such that $w(R_n)v_i = w(\lambda_i)v_i$ for $i = 1, 2, \dots, |V(G_n) \times K_0|$.

Then $P_n = R_n - w(R_n)$ will be the orthogonal projection to the span of $\{v_i | \lambda_i \geq \frac{1}{2}\}$. Moreover,

$$\begin{aligned} (13) \quad \lim_{n \rightarrow \infty} \mathbb{E} \sum_{k \in K_0} \|w(R_n)(o, k)\|_2^2 &= \lim_{n \rightarrow \infty} \mathbb{E} \sum_{k \in K_0} \langle w(R_n)^2(o, k), (o, k) \rangle \\ &= \lim_{n \rightarrow \infty} \int_0^1 w^2 d\mu_n = \int_0^1 w^2 d\mu \\ &= |K|(w^2(0) + w^2(1)) = 0 \end{aligned}$$

Here the expectation is over a uniform random vertex o of $V(G_n)$. This easily implies that $U(G_n, R_n)$ and $U(G_n, P_n)$ have the same limit, that is, $\lim U(G_n, P_n) = (G_\Gamma, e_\Gamma, P)$. (Note that in the language of [15] the vanishing limit in (13) means that (R_n) and (P_n) represent the same operator.) Now using Theorem 2.4 we get that

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{H(B^{\mathrm{rest}_K(P_n)})}{|V(G_n)|} &= \lim_{n \rightarrow \infty} h_K(G_n, P_n) \\ &= \bar{h}(G_\Gamma, e_\Gamma, \mathrm{rest}_K(P)) = \bar{h}(G_\Gamma, e_\Gamma, T). \end{aligned}$$

Now for any ε and r for large enough n we have that $B^{\mathrm{rest}_K(P_n)}$ is an (ε, r) -approximation of B^T , because $\lim_{n \rightarrow \infty} U(G_n, \mathrm{rest}(P_n)) = (G_\Gamma, e_\Gamma, T)$. So $\bar{h}(G_\Gamma, e_\Gamma, T) \leq h(B^T)$ follows.

Putting everything together we get that $\bar{h}(G_\Gamma, e_\Gamma, T) \leq h(B^T) \leq h'(B^T) \leq \bar{h}(G_\Gamma, e_\Gamma, T)$. So Theorem 2.6 follows.

7. Tree entropy. Let $G = (V, E)$ be a locally finite connected graph. Choose an orientation of each edge to obtain the oriented graph \vec{G} . The *vertex-edge incidence matrix* $A = (a_{ve})$ of \vec{G} is a $V \times E$ matrix such that

$$a_{ve} = \begin{cases} 1 & \text{if } e \text{ enters } v, \\ -1 & \text{if } e \text{ leaves } v, \\ 0 & \text{otherwise.} \end{cases}$$

Let $\star = \star(\vec{G})$ be the closed subspace of $\ell^2(E)$ generated by the rows of A , and let P_\star be the orthogonal projection from $\ell^2(E)$ to \star . If G is finite, then the determinantal measure corresponding to P_\star is the uniform measure on the spanning trees of G [9]. Let $\tau(G)$ be the number of spanning trees of G , then $H(B^{P_\star}) = \log \tau(G)$. If G is infinite, the corresponding determinantal measure is the so-called *wired uniform spanning forest* (WUSF) [17, 10, 5, 12]. Note that in both cases, the resulting measure does not depend on the chosen orientation of G .

Given a rooted graph (G, o) and a non-negative integer k , let $p_k(G, o)$ be the probability that a simple random walk starting at o is back at o after k steps.

The following theorem was proved by Lyons [14].

THEOREM 7.1. *Let G_n be a sequence of finite connected graphs, such that $|V(G_n)| \rightarrow \infty$ and their Benjamini-Schramm limit is a random rooted graph (G, o) . Then*

$$\lim_{n \rightarrow \infty} \frac{\log \tau(G_n)}{|V(G_n)|} = \mathbb{E} \left(\log \deg(o) - \sum_{k=1}^{\infty} \frac{1}{k} p_k(G, o) \right).$$

Using our results we can give another expression for the limiting quantity. Let G be a connected locally finite infinite graph, let \mathfrak{F} be the WUSF of G . For $e \in E(G)$ let $I(e)$ be the indicator of the event that $e \in \mathfrak{F}$. Given a $[0, 1]$ labeling c of $E(G)$ and an edge $e \in E(G)$ we define

$$\bar{h}(G, e, c) = H(I(e) | \{I(f) | c(f) < c(e)\}),$$

and

$$\bar{h}(G, e) = \mathbb{E} \bar{h}(G, e, c),$$

where the expectation is over the i.i.d. uniform random $[0, 1]$ labeling of G . Now we state our version of the tree entropy theorem.

THEOREM 7.2. *Let G_n be a sequence of finite connected graphs, such that $|V(G_n)| \rightarrow \infty$ and their Benjamini-Schramm limit is a random rooted graph (G, o) . Then*

$$\lim_{n \rightarrow \infty} \frac{\log \tau(G_n)}{|V(G_n)|} = \frac{1}{2} \mathbb{E} \sum_{e \sim o} \bar{h}(G, e),$$

where the summation is over the edges e incident to the root o .

PROOF. Let (\vec{G}, o) be the random rooted oriented graph obtained from (G, o) by orienting each edge independently and uniformly to one of the two possible directions. Let $L(\vec{G})$ be the line graph of \vec{G} , that is the vertex set of $L(\vec{G})$ is $V(\vec{G})$ and two vertices of $L(\vec{G})$ are connected if the corresponding edges in \vec{G} are adjacent. Let (\vec{G}', o') be obtained from (\vec{G}, o) by biasing by the degree of the root. Let e be a uniform random edge incident to o' . Then $(L(\vec{G}'), e, P_{\star(\vec{G}')})$ will be a random RGPC, which we denote by (L, e, P) . (Here the support set K of (L, e, P) is a one element set.) Now there is an orientation \vec{G}_n of G_n such that the Benjamini-Schramm limit of \vec{G}_n is (\vec{G}, o) . This can be proved by choosing random orientations, and using concentration results. We omit the details. Let (L_n, P_n) be the finite-graph-contraction $(L(\vec{G}_n), P_{\star(\vec{G}_n)})$. We have the following lemma.

LEMMA 7.3. *We have $\lim_{n \rightarrow \infty} U(L_n, P_n) = (L, e, P)$.*

PROOF. This can be proved by slightly modifying the argument of the proof of [2, Proposition 7.1]. \square

The proof can be finished using Theorem 2.4. \square

Both Lyons's and our theorem can be extended to edge weighted graphs, but in this case they are about two different quantities. However, these two quantities are closely related as we explain now. Let G be a connected finite graph, and assume that each edge e has a positive weight $w(e)$. The weight of a spanning tree T is defined as $w(T) = \prod_{e \in T} w(e)$. Let

$$Z(G, w) = \sum_{T \text{ is a spanning tree}} w(T)$$

be the sum of the weights of the spanning trees of G . Let \mathfrak{F} be a random spanning tree of G , such that for any spanning tree T we have $\mathbb{P}(\mathfrak{F} = T) = Z(G, w)^{-1}w(T)$. This is again a determinantal process, the only difference compared to the uniform case is that for each edge e we need to multiply the corresponding column of the vertex-edge incidence matrix by $\sqrt{w(e)}$. In fact, this is the way we define the weighted version of the WUSF for infinite graphs. The Shannon entropy $H(\mathfrak{F})$ of \mathfrak{F} is related to $Z(G, w)$ by the identity

$$(14) \quad H(\mathfrak{F}) = \log Z(G, w) - \mathbb{E} \log w(\mathfrak{F}).$$

Let (G_n, w_n) be a Benjamini-Schramm convergent sequence of weighted connected graphs, such that $|V(G_n)| \rightarrow \infty$ and their Benjamini-Schramm limit is a random rooted weighted graph (G, o, w) . Assume that the weights are uniformly bounded away from zero and infinity, that is, there are $0 < C_1 < C_2 < \infty$ such that all the weight are from the interval $[C_1, C_2]$. Then the generalization of Lyons's theorem states that

$$\lim_{n \rightarrow \infty} \frac{\log Z(G_n, w_n)}{|V(G_n)|} = \mathbb{E} \left(\log \pi(o) - \sum_{k=1}^{\infty} \frac{1}{k} p_{k,w}(G, o) \right),$$

where $\pi(v)$ is total weight of the edges incident to v , and $p_{k,w}(G, o)$ is defined using the random walk with transition probabilities $p(x, y) = \pi(x)^{-1}w(xy)$ instead of the simple random walk. On the other hand our theorem states that

$$\lim_{n \rightarrow \infty} \frac{H(\mathfrak{F}_n)}{|V(G_n)|} = \frac{1}{2} \mathbb{E} \sum_{e \sim o} \bar{h}(G, e, w),$$

where $\bar{h}(G, e, w)$ is defined as above, but using the weighted version of the WUSF.

These two statements above together with equation (14) of course imply that $\lim_{n \rightarrow \infty} |V(G_n)|^{-1} \mathbb{E} \log w(\mathfrak{F}_n)$ exists. However, there is a more direct proof. It is based on the observation that

$$\begin{aligned} \frac{\mathbb{E} \log w(\mathfrak{F}_n)}{|V(G_n)|} &= \frac{1}{|V(G_n)|} \sum_{e \in E(G_n)} \mathbb{P}(e \in \mathfrak{F}_n) \log w(e) \\ &= \frac{1}{2} \mathbb{E} \sum_{e \sim o} \mathbb{P}(e \in \mathfrak{F}_n) \log w(e), \end{aligned}$$

where the last expectation is over a uniform random $o \in V(G_n)$. Since we know that the limit of \mathfrak{F}_n is \mathfrak{F} , where \mathfrak{F} is the WUSF of the random rooted weighted graph (G, o, w) (see [2, Proposition 7.1]) we get that

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E} \log w(\mathfrak{F}_n)}{|V(G_n)|} = \frac{1}{2} \mathbb{E} \sum_{e \sim o} \mathbb{P}(e \in \mathfrak{F}) \log w(e).$$

Using equation (14), this provides us another formula for the limit $\lim_{n \rightarrow \infty} |V(G_n)|^{-1} \log Z(G_n, w_n)$. Namely,

$$\lim_{n \rightarrow \infty} \frac{\log Z(G_n, w_n)}{|V(G_n)|} = \frac{1}{2} \mathbb{E} \sum_{e \sim o} (\mathbb{P}(e \in \mathfrak{F}) \log w(e) + \bar{h}(G, e, w)).$$

QUESTION 7.4. We have seen that if (G, o) is an infinite random rooted graph which is the limit of finite connected graphs, then

$$\mathbb{E} \left(\log \deg(o) - \sum_{k=1}^{\infty} \frac{1}{k} p_k(G, o) \right) = \frac{1}{2} \mathbb{E} \sum_{e \sim o} \bar{h}(G, e).$$

Is this true for any infinite unimodular random rooted graph?

APPENDIX: MEASURABILITY OF THE POLAR DECOMPOSITION

The key idea is that we can realize every operator on a single fixed Hilbert-space. This can be done by using the canonical representatives defined by Aldous and Lyons in a slightly different setting. See Section 2 of [2]. Now we give the details. Let us call a RGO (G, o, T) half-canonical if the following hold. If G is finite then $V(G) = [0, |V(G)| - 1] \subset \mathbb{N}$, if G is infinite then $V(G) = \mathbb{N}$, the root o is equal to 0, moreover $B_r(G, o) = [0, |B_r(G, o)| - 1]$ for all r . If G is infinite then T is an operator on the Hilbert space $H = \ell^2(\mathbb{N} \times K)$. If G is finite we will still consider T as an operator on H by setting $T(v, k)$ to be 0 for all $(v, k) \in (\mathbb{N} \setminus V(G)) \times K$. Let \mathcal{HC} be the set of half-canonical RGOs. We endow \mathcal{HC} with a metric as follows. Given two elements $(G_1, 0, T_1)$ and $(G_2, 0, T_2)$ of \mathcal{HC} , their distance is defined as the infimum of $\varepsilon > 0$ such that for $r = \lfloor \varepsilon^{-1} \rfloor$ we have that G_1 and G_2 are the same restricted to the vertices $[0, r]$, moreover

$$|\langle T_1(v, k), (v', k') \rangle - \langle T_2(v, k), (v', k') \rangle| < \varepsilon$$

for every $v, v' \in [0, r]$ and $k, k' \in K$. Then the obvious map $g : \mathcal{HC} \rightarrow \mathcal{RG}\mathcal{O}$ is continuous. The next lemma shows that we can go the other direction too.

LEMMA A.5. *There is a measurable map from $f : \mathcal{RG}\mathcal{O} \rightarrow \mathcal{HC}$ such that for any $(G, o, T) \in \mathcal{RG}\mathcal{O}$ we have that (G, o, T) and $f(G, o, T)$ are isomorphic as RGOs. In other words $g \circ f = id$.*

PROOF. The construction given in Section 2 of [2] can be adopted to this situation. \square

Let $B(H)$ be the set of bounded linear operators on H . We endow $B(H)$ with a measurable structure by considering the coarsest σ -algebra such that all the $B(H) \rightarrow \mathbb{R}$ maps $T \mapsto \langle Te, f \rangle$ are measurable for $e, f \in \mathbb{N} \times K$. We also endow H with the measurable structure coming from the norm $\|\cdot\|_2$.

LEMMA A.6. *Let T be a $B(H)$ -valued measurable map, x be an H -valued measurable map defined on the same measurable space. Then Tx is an H -valued measurable map.*

PROOF. Let e_1, e_2, \dots be an enumeration of $\mathbb{N} \times K$. Then Tx is the pointwise limit of

$$y_n = \sum_{i=1}^n \left(\sum_{j=1}^n \langle x, e_j \rangle \langle Te_j, e_i \rangle \right) e_i.$$

Since y_n is measurable, Tx is measurable too. \square

This also has the following consequence.

LEMMA A.7. *Let S and T be $B(H)$ -valued measurable maps. Then ST is a $B(H)$ -valued measurable map.*

PROOF. Let $e, f \in \mathbb{N} \times K$. Then $STe = S(Te)$, here Te is an H -valued measurable map, so $S(Te)$ is an H -valued measurable map. So $\langle STe, f \rangle$ is measurable. \square

Given a $B(H)$ -valued measurable map T , let T^+ be its generalized inverse. Note that T^+ is not necessary bounded. We need the following theorem.

THEOREM A.8 ([16]). *For any $x \in H$ the map T^+x is an H -valued measurable map.*

From this we obtain the following lemma.

LEMMA A.9. *Let T be a $B(H)$ -valued measurable map, and let $T = UP$ be its unique polar decomposition, i.e. $P = \sqrt{A^*A}$ and $U = TP^+$. Then U is a $B(H)$ -valued measurable map.*

PROOF. From Lemma A.7 $(A^*A)^n$ is a measurable map for all n . Approximating the square root function by polynomials we get that P is a $B(H)$ -valued measurable map. The statement follows by combining Theorem A.8 and the argument of the proof of Lemma A.7. \square

Then it is not difficult to prove the following.

LEMMA A.10. *For an RGO (G, o, T) let $T = UP$ be the polar decomposition of T , then the map $(G, o, T) \mapsto (G, o, U)$ is measurable.*

Acknowledgements. The author is grateful to Miklós Abért for the useful discussions throughout the writing of this paper, to Russel Lyons and the anonymous referee for their valuable comments. The author was partially supported by the ERC Consolidator Grant 648017.

REFERENCES

- [1] ABÉRT, M. and WEISS, B. *In preparation.*
- [2] ALDOUS, D. and LYONS, R. (2007). Processes on unimodular random networks. *Electronic Journal of Probability* **12** 1454–1508.
- [3] AUSTIN, T. (2016). Additivity properties of sofic entropy and measures on model spaces. In *Forum of Mathematics, Sigma* **4**. Cambridge University Press.
- [4] AUSTIN, T. and PODDER, M. (2018). Gibbs measures over locally tree-like graphs and percolative entropy over infinite regular trees. *Journal of Statistical Physics* **170** 932–951.
- [5] BENJAMINI, I., LYONS, R., PERES, Y. and SCHRAMM, O. (2001). Special invited paper: uniform spanning forests. *Annals of probability* 1–65.
- [6] BENJAMINI, I. and SCHRAMM, O. (2000). Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.* **6** 1–13.
- [7] BORGS, C., CHAYES, J., KAHN, J. and LOVÁSZ, L. (2013). Left and right convergence of graphs with bounded degree. *Random Structures & Algorithms* **42** 1–28.
- [8] BOWEN, L. (2010). Measure conjugacy invariants for actions of countable sofic groups. *Journal of the American Mathematical Society* **23** 217–245.
- [9] BURTON, R. and PEMANTLE, R. (1993). Local characteristics, entropy and limit theorems for spanning trees and domino tilings via transfer-impedances. *The Annals of Probability* 1329–1371.

- [10] HÄGGSTRÖM, O. (1998). Uniform and minimal essential spanning forests on trees. *Random Structures & Algorithms* **12** 27–50.
- [11] HEICKLEN, D. and LYONS, R. (2003). Change intolerance in spanning forests. *Journal of Theoretical Probability* **16** 47–58.
- [12] LYONS, R. (1998). A bird’s-eye view of uniform spanning trees and forests. *Microsurveys in discrete probability* **41** 135–162.
- [13] LYONS, R. (2003). Determinantal probability measures. *Publications Mathématiques de l’IHÉS* **98** 167–212.
- [14] LYONS, R. (2005). Asymptotic enumeration of spanning trees. *Combinatorics, Probability and Computing* **14** 491–522.
- [15] LYONS, R. and THOM, A. (2016). Invariant coupling of determinantal measures on sofic groups. *Ergodic Theory and Dynamical Systems* **36** 574–607.
- [16] NASHED, M. Z. and SALEHI, H. (1973). Measurability of generalized inverses of random linear operators. *SIAM Journal on Applied Mathematics* **25** 681–692.
- [17] PEMANTLE, R. (1991). Choosing a spanning tree for the integer lattice uniformly. *The Annals of Probability* 1559–1574.
- [18] SEWARD, B. (2016). Weak containment and Rokhlin entropy. *arXiv preprint arXiv:1602.06680*.