

Entropy inequalities and exponential decay of correlations for unique Gibbs measures on trees. *

Andrei Alpeev †

April 30, 2019

Abstract

In a recent paper by A. Backhausz, B. Gerencsér and V. Harangi, it was shown that factors of independent identically distributed random processes (IID) on trees obey certain geometry-driven inequalities. In particular, the mutual information shared between two vertices decays exponentially, and there is an explicit bound for this decay. In this note we show that all of these inequalities could be verbatim translated to the setting of factors of processes driven by unique Gibbs measures. As a consequence, we show that correlations decay exponentially for unique Gibbs measures on trees.

keywords: Gibbs measure, regular tree, exponential decay of correlations, Bethe lattice, random coverings.

1 Introduction

The paper [BGH] provided a unified approach to entropy inequalities for factors of IID's on trees. It was shown that all known entropy inequalities in this setting follow by certain combinatorial constructions from the “general edge-vertex“ entropy inequality. The purpose of this note is to show that this general inequality holds in a broader setting of factors of Gibbs processes whose distribution is a unique Gibbs measure, thus transferring all the entropy inequalities.

Let T_d be a d -regular tree with vertex set $V(T_d)$. Let A be a finite set (*alphabet*). Assume that to each vertex v of the tree a function (*potential*) $\psi_v : A^{V(T)} \rightarrow \mathbb{R}$ is assigned. We would like to require that this potential is “local” and “symmetric”. Let's state these requirements formally. Potential $(\psi_v)_{v \in V(T_d)}$ is local if for each $v \in V(T_d)$ there is a positive integer D and function $\psi'_v : A^{\text{Ball}(v,D)} \rightarrow \mathbb{R}$ such that $\psi_v = \psi'_v \circ \text{pr}_D$, here $\text{pr}_{\text{Ball}(v,D)}$ stands

*This research is supported by “Native towns”, a social investment program of PJSC “Gazprom Neft”

†Chebyshev Laboratory, St. Petersburg State University, 14th Line, 29b, Saint Petersburg, 199178 Russia, alpeevandrey@gmail.com

for the natural projection $A^{V(T_d)} \rightarrow A^{\text{Ball}(v,D)}$. Note that the action of the automorphism group $\text{Aut}(T_d)$ on T_d is lifted naturally to the action on $A^{V(T_d)}$:

$$(\gamma\omega)(v) = \omega(\gamma^{-1}v),$$

for each $v \in V(T_d)$ and $\gamma \in \text{Aut}(T_d)$. By saying that the potential is symmetric we mean that for each $v \in V(T_d)$ and $\gamma \in \text{Aut}(T_d)$ holds $\psi_{\gamma v}(\gamma\omega) = \psi_v(\omega)$. We note that due to the latter requirement, the locality of the potential is uniform: there is a positive integer R such that for each $v \in V(T_d)$ there is a function $\psi_v^R : A^{\text{Ball}(v,R)} \rightarrow \mathbb{R}$ satisfying $\psi_v = \psi_v^R \circ \text{pr}_{\text{Ball}(v,R)}$. This means that we can define ψ_v on A^W for any $W \supset \text{Ball}(v, R)$. We denote $\partial^R(W)$ the R -neighborhood of W .

A Borel probability measure μ on $A^{V(T_d)}$ is called a *Gibbs measure* if for every $\Lambda \Subset V(T_d)$, for any $s \in A^\Lambda$ and $t \in A^{\partial^{2R}\Lambda}$ we have the following for the conditional probability

$$\Pr(\text{pr}_\Lambda(\omega) = s \mid \text{pr}_{\partial^{2R}\Lambda}(\omega) = t) = \frac{\exp\left(-\sum_{v \in \Lambda \cap \partial^R \Lambda} \psi_v(s \sqcup t)\right)}{Z_{\Lambda,t}},$$

where ω is distributed according to μ , and $Z_{\Lambda,t}$ stands for the normalizing coefficient

$$\sum_{s' \in A^\Lambda} \exp\left(-\sum_{v \in \Lambda \cap \partial^R \Lambda} \psi_v(s' \sqcup t)\right).$$

A Gibbs process is simply a process whose distribution is a Gibbs measure.

As we will show in this note, all the inequalities from [BGH] could be transferred to the setting of factors of unique Gibbs measure processes. We will not re-write all of these, but the most striking one:

Theorem 1. *Let T_d be a d -regular tree, $d > 2$. Let A be a finite set. Let $(\psi)_v$ be an $\text{Aut}(T_d)$ -invariant potential such that there is a unique Gibbs measure for it. Let (X_v) be the process distributed according to this measure. Then the following holds for any two vertices u and v of T_d with distance l between them:*

$$\frac{I(X_u, X_v)}{H(X_v)} \leq \begin{cases} \frac{2}{d(d-1)^l} & \text{if } k = 2l + 1 \text{ is odd,} \\ \frac{1}{(d-1)^l} & \text{if } k = 2l \text{ is even.} \end{cases}$$

The interesting thing is that we do not impose any restrictions on the model apart from it being symmetric with respect to the tree automorphisms and having unique Gibbs measure.

In [BGH] the idea was proposed to show that certain processes (including Gibbs processes) are not factors of IID based on their failure to satisfy the entropy inequalities. Our results show that this approach will fail to distinguish factors of IID processes from factors of unique Gibbs measure processes.

Models of statistical physics on regular trees (Bethe lattices) have been extensively studied, see e.g. [Ba]. Gibbs processes on infinite graphs and their approximations by big finite graphs have been studied in [DM10a], [DM10b].

In my paper [A15], it was shown that for unique Gibbs measures, finite models of infinite Gibbs structures admitting unique Gibbs measure provide a way to compute the sofic entropy. In works [AuP17] and [A17] the random ordering technique was used to assess the asymptotic behaviour of certain classes of Gibbs processes on sequences of finite graphs.

One particular property of interest is the exponential decay of correlations. By a symmetric observable we mean a collection of functions $(\alpha_v)_{v \in V(T)}$ satisfying the same requirements as potential (it should be symmetric and local), but taking values in a finite set B (we will call it a real observable if it takes values in \mathbb{R}). In fact, previous theorem holds for any symmetric observable. From that we may deduce that there is an exponential correlation decay:

Corollary 1. *In the setting of the previous theorem, for any symmetric real observable (α_u) there is a constant C such that*

$$\text{Cov}_\nu(\alpha_u, \alpha_v) \leq C(d-1)^{-k/4}.$$

Proof. Let B be the set of values of α_v for some (=any) $v \in V(T)$. Set B is finite. There are two natural projections $\text{pr}_1, \text{pr}_2 : B \times B \rightarrow B$. Denote η the distribution of $\alpha_u()$ for any $u \in V(T)$. For each $u, v \in V(T)$ denote $\xi_{u,v}$ the distribution of (α_u, α_v) . Let M' be the set of probability measures θ on $B \times B$ such that $\text{pr}_1(\theta) = \text{pr}_2(\theta) = \eta$. Note that M' is a finite-dimensional convex set. We note that \mathcal{H} is a smooth function in a neighborhood of $\eta \otimes \eta$. We claim that there is a constant C' such that

$$I(\alpha_u, \alpha_v) \geq C' \|\xi_{u,v} - \eta \otimes \eta\|^2,$$

where $\|\cdot\|$ stands for the total variation distance. This follows from the fact that $\eta \otimes \eta$ is the unique maximal point of function $\xi \mapsto 2\mathcal{H}(\eta) - \mathcal{H}(\xi)$ defined on M' , and that the second derivative of this function is positive in any direction. Next we note that $\text{Cov}(\alpha_u, \alpha_v)$ is an affine function of the distribution of (α_u, α_v) , so we get that there is a constant C such that

$$\text{Cov}(\alpha_u, \alpha_v) \leq C(I(\alpha_u, \alpha_v))^{1/2}.$$

This implies the desired. □

In the paper [BSV15], a stronger decorrelation bound is obtained for factors of IID process. Various results of this flavour were obtained in [BS14], [BGHV], [CHV], [GH17]. It seems that some of these could be transferred to the setting of factors of Gibbs processes.

Our proof mostly follows the steps of [BGH]. We refer the reader to the forementioned paper for the details on how to derive Theorem 1 and other entropy inequalities from the “general edge-vertex inequality”, Theorem 2, presented in Section 3. So it is enough to prove this inequality. Again, the basic argument structure is the same. The main difference lies in the proof of our Lemma 6, which corresponds to Lemma 5.4 of [BGH]. In the factor of IID case of [BGH], a standard measure concentration inequality yielded the desired right away. In

our case, lack of independence in the base process forced us to opt for a second moment argument coupled with some analysis of unique Gibbs measures. Note that the argument takes a kind of a bootstrap form: a weak decorrelation property is used to prove a quantitative one.

2 Preliminaries

For a metrizable compact space X , we denote $M(X)$ the space of Borel probability measures endowed with the weak* topology. For two random variables X and Y , their covariance is denoted $\text{Cov}(X, Y)$:

$$\text{Cov}(X, Y) = \mathbb{E}XY - \mathbb{E}X \mathbb{E}Y = \mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y)).$$

The sign “ \Subset ” stands for “a finite subset”. For a graph G we will denote $V(G)$ its vertex set and $E(G)$ its edge set. Denote $\text{Aut}(G)$ the automorphism group of graph G . For any vertex v and a positive integer R , denote $\text{Ball}(v, R)$ the ball of radius R centered at v .

Let H be a locally-finite simple graph. Let us fix a finite set A (an *alphabet*).

Let W be any subset of $V(H)$. Denote $\text{pr}_W : A^{V(H)} \rightarrow A^W$ the natural projection map. Denote $\mathcal{B}(W)$ the σ -subalgebra on $A^{V(H)}$ that is the preimage of the Borel σ -algebra on A^W under the projection map pr_W .

Fix a positive integer D (a *memory parameter*). A *potential* is a collection $(\psi)_{v \in V(H)}$ of functions $A^{\text{Ball}(v, D)} \rightarrow \mathbb{R}$. We will also say that ψ is a D -*potential*, reflecting that its memory parameter is D . Abusing notation a little, we may assume that ψ_v is defined on A^W for any subset W of $V(H)$ containing $\text{Ball}(v, D)$ (including the whole space $A^{V(H)}$). For a positive integer R and a subset W of $V(H)$, we denote $\partial^R W$ the R -neighborhood of set W , that is the set of all the vertices outside of W at a distance not bigger than R . The potential defines the set of *Gibbs measures*. If graph H is finite, then the Gibbs measure ν is given by the formula:

$$\nu(\{\omega\}) = \frac{\exp\left(-\sum_{v \in V(H)} \psi_v(\omega)\right)}{Z},$$

where Z stands for the normalizing coefficient

$$\sum_{\omega \in A^{V(H)}} \exp\left(-\sum_{v \in V(H)} \psi_v(\omega)\right).$$

For $s \in A^{\Lambda_1}$ and $t \in A^{\Lambda_2}$ with $\Lambda_1 \cap \Lambda_2 = \emptyset$, denote $s \sqcup t \in A^{\Lambda_1 \cup \Lambda_2}$ their glueing. If H is infinite, then there could be multiple Gibbs measures. A Borel probability measure ν on $A^{V(H)}$ is a Gibbs measure for potential $(\psi)_{v \in V(H)}$ if for every $\Lambda \Subset V(H)$, for any $s \in A^\Lambda$ and $t \in A^{\partial^{2D}\Lambda}$, we have the following for the conditional probability:

$$\Pr(\text{pr}_\Lambda(\omega) = s \mid \text{pr}_{\partial^{2D}\Lambda}(\omega) = t) = \frac{\exp\left(-\sum_{v \in \Lambda \cap \partial^D \Lambda} \psi_v(s \sqcup t)\right)}{Z_{\Lambda, t}},$$

where $\omega \in A^{V(H)}$ is distributed according to ν , and $Z_{\Lambda,t}$ stands for the normalizing coefficient

$$\sum_{s' \in A^\Lambda} \exp \left(- \sum_{v \in \Lambda \cap \partial^R \Lambda} \psi_v(s' \sqcup t) \right).$$

We will say that pair (H, ψ) forms a *Gibbs structure*.

3 Coverings and the edge-vertex inequality for Gibbs measures

Let G be a finite simple connected graph. The universal covering \widehat{G} of graph G is the tree obtained in the following way. Let v be any vertex from G . The vertex set of \widehat{G} is the set of all non-backtracking paths on G starting from v (the trivial one included). Two paths are connected by an edge in \widehat{G} iff one of these paths is another extended by one edge. We define the natural projection map $\varphi : V(\widehat{G}) \rightarrow V(G)$ which sends a vertex from \widehat{G} to the end of the corresponding non-backtracking path. Of course, this map could be extended to the set of edges $E(\widehat{G})$ as well. Note that the construction described is unique up to an automorphism.

Consider the subgroup Γ_φ of $\text{Aut}(\widehat{G})$ consisting of all φ -preserving automorphisms of graph \widehat{G} :

$$\Gamma_\varphi = \{ \gamma \in \text{Aut}(\widehat{G}) \mid \varphi(\gamma v) = \varphi(v) \text{ for any } v \in V(\widehat{G}) \}.$$

We will usually write Γ instead of Γ_φ .

Let A be a finite alphabet. Note that Γ acts on $A^{\widehat{G}}$:

$$(\gamma\omega)(v) = \omega(\gamma^{-1}v),$$

for any $\gamma \in \Gamma$, $v \in V(\widehat{G})$ and $\omega \in A^{V(\widehat{G})}$.

Let $\psi = (\psi)_{v \in V(\widehat{G})}$ be a Γ -invariant potential on \widehat{G} . That is, for any $v \in V(\widehat{G})$, any $\gamma \in \Gamma$ and $\omega \in A^{V(\widehat{G})}$ holds:

$$\psi_{\gamma v}(\gamma\omega) = \psi_v(\omega).$$

If ν is the unique measure for (\widehat{G}, ψ) , then ν is automatically Γ -invariant; this follows from the fact that $\gamma\nu$ is a Gibbs measure for any $\gamma \in \Gamma$.

Let ν be a Γ -invariant measure on $A^{V(\widehat{G})}$. Let B be another finite alphabet. A measurable map τ from $A^{V(\widehat{G})}$ to $B^{V(\widehat{G})}$ is Γ -equivariant if the following holds:

$$\tau(\gamma\omega) = \gamma(\tau(\omega)),$$

for any $\gamma \in \Gamma$ and $\omega \in A^{V(\widehat{G})}$.

Let ν be the unique Gibbs measure for (\widehat{G}, ψ) , and τ be a Γ -equivariant measurable map $A^{V(\widehat{G})} \rightarrow B^{V(\widehat{G})}$. Let μ be $\tau(\nu)$. We note that for a fixed edge

$e \in E(G)$ and for any two edges $e', e'' \in E(\widehat{G})$ such that $\phi(e') = \phi(e'') = e$, the marginal distributions $\text{pr}_{e'}(\mu)$ and $\text{pr}_{e''}(\mu)$ on $B \times B$ are the same (up to the natural re-labeling). So it makes sense to define the distribution μ_e on $B^{\{u,v\}} = B \times B$, where $e = \{u, v\}$.

Remind that the Shannon entropy of a measure η on a finite set Q is defined as

$$\mathcal{H}(\eta) = - \sum_{q \in Q} \eta(\{q\}) \log \eta(\{q\}),$$

with the convention $0 \log 0 = 0$. The Shannon entropy of a random variable is defined as the Shannon entropy of its distribution.

The main theorem of this paper is the following:

Theorem 2. *In the notation above, the following inequality holds:*

$$\sum_{v \in V(G)} (\deg(v) - 1) \mathcal{H}(\mu_v) \leq \sum_{e \in E(G)} \mathcal{H}(\mu_e).$$

The proof will ensue in the next section as a consequence of two lemmata concerning colorings of finite random coverings. Basically, the first one shows that the difference between the right-hand and the left-hand sides of the edge-vertex inequality is involved in the estimate of the expected number of some “good” colorings of random covers of graph G . On the other hand, the second one shows that this expected number is at least $1 - o(1)$ (a “good” coloring exist almost surely).

Note that Theorem 2 is exactly the “general edge-vertex inequality” from [BGH], the only change is that we consider factors of unique Gibbs measure processes instead of factors of IID’s. IID processes form a particular case of unique Gibbs measure processes: it is easy to construct a potential whose unique Gibbs measure would be any given product-measure (the “interactionless potential”). Theorem 1 and other entropy inequalities follow by constructing specially tailored factors, see [BGH, Section 3] for details. These constructions work in our case without any noticeable changes.

A Γ -cellular map is a Γ -equivariant *continuous* map τ from $A^{V(\widehat{G})}$ to $B^{V(\widehat{G})}$. The continuity assumption may sound a bit fancy. In fact, it is equivalent to say that for any $v \in V(\widehat{G})$ there is a finite set W of $V(\widehat{G})$ such that $(\tau(\omega))(v)$ depends only on $\text{pr}_W(\omega)$. This equivalence is known as the Curtis-Hedlund-Lyndon theorem. There is a well-known trick to approximate a measurable equivariant map by a cellular map. Namely, if τ' is a measurable equivariant map, then for any $\varepsilon > 0$ there is a cellular map τ such that for any $v \in V(\widehat{G})$ holds

$$\nu\left(\left\{\omega \in A^{V(\widehat{G})} : (\tau(\omega))(v) = (\tau'(\omega))(v)\right\}\right) > 1 - \varepsilon.$$

So it is enough to consider only factors obtained by cellular maps to prove Theorem 2.

4 Random coverings and proof of Theorem 2

Let G be a finite simple connected graph. Let \widehat{G} be its universal covering together with a natural covering map $\varphi : \widehat{G} \rightarrow G$. An N -fold covering of G is a graph \widehat{G}_N whose vertex set is $\{1, \dots, N\} \times V(G)$ together with a graph morphism $q_N : \widehat{G}_N \rightarrow G$ such that

1. for each $v \in V(G)$ and i from $\{1, \dots, N\}$ holds $q_N(i, v) = v$;
2. there are exactly N preimages of each edge from G ;
3. for each edge $\{u, v\}$, the preimage $q_N^{-1}(\{u, v\})$ is a perfect matching between the sets $q_N^{-1}(u)$ and $q_N^{-1}(v)$;
4. if vertices u' and v' from \widehat{G}_N are connected, then $q_N(u')$ and $q_N(v')$ are connected.

We note that, for each connected component of graph \widehat{G}_N , there is a covering map $q : \widehat{G} \rightarrow \widehat{G}_N$ onto this component such that $\varphi = q_N \circ q$

A uniform random N -fold covering of G is an N -fold covering such that the perfect matchings $q_N^{-1}(e)$ for each $e \in E(G)$ are taken uniformly at random. An important property of the random N -fold covering is that for big N it looks like the universal covering in a neighborhood of almost every of its vertices. To be more precise, consider an N -fold covering \widehat{G}_N . Take a vertex u from $V(\widehat{G}_N)$. Let $q : \widehat{G} \rightarrow \widehat{G}_N$ be a covering map to the connected component of \widehat{G}_N containing u . We will say that vertex u is R -nice, for a positive integer R , if q is a bijection from $\text{Ball}(u', R)$ to $\text{Ball}(u, R)$, where u' is any preimage of u under map q . We note that the latter definition is independent of the choice of covering map q . We will say that an edge is R -nice if both of its end are R -nice. For each vertex v in graph G , denote $L_v^{\widehat{G}_N}$ the preimage set of this vertex under the natural covering map to G . We establish similar notation $L_e^{\widehat{G}_N}$ for an edge $e \in E(G)$.

For a positive ε and positive integer R , we will say that an N -fold covering \widehat{G}_N of graph G is (R, ε) -nice if for each vertex $v \in V(G)$ the portion of R -nice vertices in $L_v^{\widehat{G}_N}$ is bigger than $1 - \varepsilon$, and the same holds for each edge. The following lemma is discussed in [BGH, Section 5.a]

Lemma 1. *Let R be a positive integer and let $\varepsilon > 0$. Then for any $\varepsilon > 0$ there is N' such that for every $N > N'$ the random uniform N -fold covering \widehat{G}_N is (R, ε) -nice with probability at least $1 - \varepsilon$.*

Sketch of proof. Let v be a vertex from graph G and p be a finite non-backtracking path on G starting at v . Let $u \in V(\widehat{G}_N)$ be any vertex with $v = q_N(u)$, and let \hat{p} be the lift of path p to the path on \widehat{G}_N starting at u , and whose projection under q_N is p . It is sufficient to prove that, with probability close to 1, \hat{p} has no recurring vertices provided that N is big enough. This can be done easily by induction and using the “lazy evaluation” trick. Namely, instead of constructing at once the random perfect matchings involved in the definition of random

covering, one should only pick randomly the edge needed to continue lifting the path considered. \square

Let \widehat{G}_N be an N -fold covering of G , let B be a finite alphabet. Let c be a map from $V(\widehat{G}_N)$ to B . We associate with it the “empirical distribution”. Namely, for each vertex v we pick uniformly its preimage v' under the covering map q_N . The distribution of $c(v')$, a measure on B , is our empirical distribution and will be denoted μ_v^c . In other words,

$$\mu_v^c = \frac{1}{N} \sum_{v' \in L_v^{\widehat{G}_N}} \delta_{c(v')}.$$

In the same way, for each edge $e = \{u, v\}$, we pick uniformly at random its preimage $e' = \{u', v'\}$ ($q_N(u') = u$ and $q_N(v') = v$) under the covering map. The distribution of $(c(u'), c(v'))$ is a measure on $B^e = B \times B$ and will be denoted μ_e^c . We may also write:

$$\mu_e^c = \frac{1}{N} \sum_{\{u', v'\} \in L_e^{\widehat{G}_N}} \delta_{(c(u'), c(v'))}.$$

In the next lemma we assume for a moment that $(\mu_v)_{v \in V(G)}$ and $(\mu_e)_{e \in E(G)}$ are just two collections of measures on B and $B \times B$ respectively. We will say that these collections are consistent if $\text{pr}_v(\mu_e) = \mu_v$ for every $e \in E(G)$ and v is one of two vertices connected by edge e .

Lemma 2. *Let $(\mu_v)_{v \in V(G)}$ and $(\mu_e)_{e \in E(G)}$ be a consistent collection of measures. For every $\varepsilon > 0$ the following holds:*

$$\begin{aligned} & \mathbb{E}_{\widehat{G}_N} \left| c : V(\widehat{G}_N) \rightarrow B, \|\mu_e^c - \mu_e\| \leq \varepsilon \text{ for every } e \in E(G) \right| \\ &= \exp \left(N \left(\sum_{e \in E(G)} \mathcal{H}(\mu_e) - \sum_{v \in V(G)} (\deg(v) - 1) \mathcal{H}(\mu_v) + o_{\varepsilon, N}(1) \right) \right), \quad (1) \end{aligned}$$

where o -small tends to zero as N tends to infinity and ε tends to zero.

We omit the proof and refer the reader to [BGH, Lemma 5.6]. Note that the statement therein is slightly different, the precise equality is required. Nonetheless, the careful inspection of the proof reveals that it works in our case as well.

Theorem 2 will follow from the previous lemma if only we can prove that the expectation mentioned in that lemma is bigger than $1 - o(1)$. This is exactly what we get in the next one.

Lemma 3. *For every $\varepsilon > 0$ there is N' such that for every $N > N'$, a random N -fold covering \widehat{G}_N with probability more than $1 - \varepsilon$ has a coloring $c : V(\widehat{G}_N) \rightarrow B$ such that $\|\mu_e^c - \mu_e\| \leq \varepsilon$ for every $e \in E(G)$.*

We will prove Lemma 3 in Section 6 after laying below the necessary groundwork on Gibbs measures.

5 Asymptotic independence for unique Gibbs measures

Distant parts of processes governed by unique Gibbs measures are almost independent. It turns out, that the latter is true in some sense even if we warp the Gibbs structure outside of some (big enough) ball in the graph. The purpose of this section is to explain these points.

We refer the reader to the monographs [G11] and [RS15] for the preliminaries on Gibbs measures.

Gibbs measures satisfy the so-called *Markov property* (see [G11, p. 157]). Namely, let ν be a Gibbs measure for the potential ψ . For any finite subset Λ of $V(H)$ and for any $\mathcal{B}(\Lambda)$ -measurable \mathcal{L}_2 -function f , holds

$$\mathbb{E}(f|\mathcal{B}(V(H) \setminus \Lambda)) = \mathbb{E}(f|\mathcal{B}(\partial^{2D} \Lambda)).$$

Remark 1. Consider the following expression:

$$\|\mathbb{E}(f|\mathcal{B}(\partial^{2D} \Lambda)) - \mathbb{E}(f)\|_2.$$

There is an explicit parameter: the measure ν . In fact, only the projection $\nu_0 = \text{pr}_{\Lambda \cup \partial^{2D} \Lambda} \nu$ matters in this respect. We may notice also that f is well-defined on $A^{\Lambda \cup \partial^{2D} \Lambda}$ and that ν'_0 is positive on all elements of $A^{\Lambda \cup \partial^{2D} \Lambda}$. The latter means that the function

$$\nu_1 \mapsto \|\mathbb{E}_{\nu_1}(f|\mathcal{B}(\partial^{2D} \Lambda)) - \mathbb{E}(f)\|_2$$

is continuous in a neighborhood of ν_0 .

Let H be a locally-finite connected graph with a D -potential ψ . Let H' be another locally-finite graph with a D -potential ψ' . Let v be a vertex in H , v' be a vertex in H' and $R > D$ be an integer. A (v, v', R) -partial isomorphism is an isomorphism between induced subgraphs $t : \text{Ball}(v, R) \rightarrow \text{Ball}(v', R)$ such that $t(v) = v'$, and

$$\psi'_{t(u)}(t\omega) = \psi_u(\omega),$$

for every $u \in \text{Ball}(v, R - D)$ and $\omega \in A^{\text{Ball}(v, R)}$ (note that t induces a map $A^{\text{Ball}(v, R)} \rightarrow A^{\text{Ball}(v', R)}$). Roughly speaking, this means that the Gibbs structures are the same in the neighborhoods of v and v' .

Suppose that the Gibbs measure is unique for potential ψ on H and v is a vertex from $V(H)$. Let $f : A^{V(H)} \rightarrow \mathbb{R}$ be a bounded measurable function. Then

$$\lim_{R \rightarrow \infty} \mathbb{E}\left(f|\mathcal{B}(V(H) \setminus \text{Ball}(v, R))\right) = \mathbb{E}(f), \quad (2)$$

where the limit is taken in the \mathcal{L}_2 -norm. The latter is a consequence of the martingale convergence theorem together with the fact that the tail subalgebra $\mathcal{T} = \bigcap_{R > 0} \mathcal{B}(V(H) \setminus \text{Ball}(v, R))$ is trivial on the unique Gibbs measure ([G11, Theorem 7.7 (a), p. 118]).

A unique Gibbs measure is stable in a certain sense. To be more precise,

Lemma 4. *Let ν be the unique Gibbs measure for Gibbs structure (H, ψ) , let Λ be a finite subset of $V(H)$ and $v \in \Lambda$. Then for every $\varepsilon > 0$ there is such an integer R that if the Gibbs structure is changed only outside of $\text{Ball}(v, R)$, then any Gibbs measure ν' (it is not necessarily unique anymore) for the new Gibbs structure satisfies*

$$\|\text{pr}_\Lambda(\nu) - \text{pr}_\Lambda(\nu')\| \leq \varepsilon.$$

Proof. Assume the contrary. There is an $\varepsilon > 0$ and a sequence (H^R, ψ^R) of Gibbs structures such that (H^R, ψ^R) is isomorphic to (H, ψ) in the $\text{Ball}(v, R)$, and for any (H^R, ψ^R) there is a Gibbs measure ν^R such that

$$\|\text{pr}_\Lambda(\nu) - \text{pr}_\Lambda(\nu^R)\| > \varepsilon$$

if $\Lambda \subset \text{Ball}(v, R)$. For each $W \Subset V(H)$, the projection $\text{pr}_W(\nu^R)$ is well-defined as soon as $W \subset \text{Ball}(v, R)$. Using the diagonal trick, we may extract a subsequence, and assume that projections $\text{pr}_W(\nu^i)$ converge for all finite subsets W of $V(H)$. For each W consider the limiting measure ν_W . Using the Kolmogorov theorem, we may extend this collection of measures to a measure ν_1 on $A^{V(H)}$. This has a property that $\|\text{pr}_\Lambda(\nu_1) - \text{pr}_\Lambda(\nu)\| > \varepsilon$. We also note that ν_1 is a Gibbs measure for (H, ψ) . This contradicts our assumption that ν is the unique Gibbs measure for (H, ψ) . \square

Lemma 5. *Suppose that graph H with potential ψ has a unique Gibbs measure. Let f_1 be a function defined on A^{W_1} for some finite subset W_1 of $V(H)$, such that $0 \leq f_1 \leq 1$, and let $v \in W_1$. For every $\varepsilon > 0$ there are such positive integers $R' < R''$ with $W_1 \subset (\text{Ball}(v', R'))$ that if (H, ψ) is (v, v', R') -partially isomorphic to (H', ψ') , and ν' is a Gibbs measure (that is not necessarily unique) for (H', ψ') , then the following holds.*

For any function f_2 defined on A^{W_2} , where $W_2 \subset V(H') \setminus \text{Ball}(v', R')$, such that $0 \leq f_2 \leq 1$, we have:

$$|\text{Cov}_{\nu'}(f'_1, f_2)| \leq \varepsilon,$$

where f'_1 is the transfer of f_1 from H to H' via the partial isomorphism map.

Proof. Denote $\mathcal{A} = \mathcal{B}(V(H') \setminus \text{Ball}(v', R'))$ and $\mathcal{C} = \mathcal{B}(\partial^{2D} \text{Ball}(v', R'))$.

$$\begin{aligned} |\text{Cov}(f'_1, f_2)| &= |\mathbb{E}(f'_1 - \mathbb{E}f'_1)(f_2 - \mathbb{E}f_2)| = \\ &= \left| \mathbb{E} \left(\mathbb{E}((f'_1 - \mathbb{E}f'_1)(f_2 - \mathbb{E}f_2) \mid \mathcal{A}) \right) \right| = \\ &= \left| \mathbb{E} \left(\mathbb{E}((f'_1 - \mathbb{E}f'_1) \mid \mathcal{A}) \cdot (f_2 - \mathbb{E}f_2) \right) \right|. \end{aligned}$$

The last equality is due to the fact that if F_1 and F_2 are two \mathcal{L}_2 -functions and F_2 is \mathcal{A} -measurable, then $\mathbb{E}(F_1 \cdot F_2 \mid \mathcal{A}) = \mathbb{E}(F_1 \mid \mathcal{A}) \cdot F_2$ almost everywhere. We carry on using the Cauchy-Bunyakovsky-Schwartz inequality and then the Markov property:

$$\begin{aligned} \left| \mathbb{E} \left(\mathbb{E}((f'_1 - \mathbb{E}f'_1) \mid \mathcal{A}) \cdot (f_2 - \mathbb{E}f_2) \right) \right| &\leq \left\| \mathbb{E}((f'_1 - \mathbb{E}f'_1) \mid \mathcal{A}) \right\|_2 = \\ &= \left\| \mathbb{E}(f'_1 \mid \mathcal{A}) - \mathbb{E}f'_1 \right\|_2 = \left\| \mathbb{E}(f'_1 \mid \mathcal{C}) - \mathbb{E}f'_1 \right\|_2. \end{aligned}$$

If R'' is big enough, then we can transfer the expression from graph H' back to H using Remark 1 and Lemma 4. Then we apply again the Markov property to change the conditioning subalgebra:

$$\|\mathbb{E}(f'_1 | \mathcal{C}) - \mathbb{E}f'_1\|_2 \leq \varepsilon/2 + \|\mathbb{E}(f_1 | \mathcal{C}') - \mathbb{E}f_1\|_2 = \varepsilon/2 + \|\mathbb{E}(f_1 | \mathcal{A}') - \mathbb{E}f_1\|_2,$$

where $\mathcal{C}' = \mathcal{B}(\partial^{2D} \text{Ball}(v, R'))$ and $\mathcal{A}' = \mathcal{B}(V(H') \setminus \text{Ball}(v', R'))$. Now we use the martingale convergence (equation 2) to note that

$$\|\mathbb{E}(f_1 | \mathcal{A}') - \mathbb{E}f_1\|_2 \leq \varepsilon/2,$$

for big enough R' . Altogether, if we take R' and R'' big enough and such that $R' < R''$, then we have

$$|\text{Cov}_{\nu'}(f'_1, f_2)| \leq \varepsilon.$$

□

6 Nice coverings have good colorings

Suppose \widehat{G}_N is an N -fold covering of graph G . We note that in this case there is a natural translation of the potential to this covering. Indeed, for any $u \in V(\widehat{G}_N)$ let q be a covering map from H to the connected component of \widehat{G}_N containing u . Fix any preimage u' of u under map q . Any coloring $\omega \in A^{V(\widehat{G}_N)}$ can be lifted to a periodic coloring $\omega' \in A^{V(\widehat{G})}$. We now define $\psi_u^{\widehat{G}_N}(\omega)$ to be $\psi_{q^{-1}(u)}(\omega')$. This construction looks very natural when D -ball around u is isomorphic to the corresponding D -ball around u' . In the latter case we literally translate the potential from $\text{Ball}(u', D)$ to $\text{Ball}(u, D)$. In a similar fashion, we can transfer the cellular map τ to the map $\tau^{\widehat{G}_N}$ from $A^{V(\widehat{G}_N)}$ to $B^{V(\widehat{G}_N)}$. Namely, we will define $(\tau^{\widehat{G}_N}(\omega))(u)$ to be $(\tau(\omega'))(u')$. The transferred potential will define the Gibbs measure $\nu^{\widehat{G}_N}$ on $A^{\widehat{G}_N}$ (note that it is defined on a finite set and is unique therefore). Now define $\mu^{\widehat{G}_N} = \tau^{\widehat{G}_N}(\nu^{\widehat{G}_N})$. Remind that for a map $c : \widehat{G}_N \rightarrow B$ and any $e \in E(G)$ we defined the empirical distribution μ_e^c .

Lemma 6. *For any $\varepsilon > 0$ there is $\delta > 0$ and natural numbers R and N' , that if an N -fold covering \widehat{G}_N ($N > N'$) of graph G is (R, δ) -nice, then there is a coloring $c : V(\widehat{G}_N) \rightarrow B$ such that $\|\mu_e - \mu_e^c\| \leq \varepsilon$ for every $e \in E(G)$.*

Note that this lemma implies immediately Lemma 3 since, by Lemma 1, a random covering \widehat{G}_N for $N > N'$ is (R, δ) -nice with probability bigger than $1 - \varepsilon$, provided that N' is big enough.

Proof. As in the proof of Lemma 5.4 from [BGH] we will employ a random construction. We will prove that for big enough N' and R , and small enough $\delta > 0$, a $\mu^{\widehat{G}_N}$ -random c will suffice for us with high probability.

Fix an edge $e \in E(G)$ connecting two vertices u and v of graph G .

We note that $\mathbb{E}_c(\mu_e^c)$ is arbitrarily close to μ_e , provided N' , R are large and δ is small enough. This is due to two observations. First, if R is big,

and i is an R -nice edge of \widehat{G}_N , then the distribution of $B \times B$ -valued random variable $(c(u'), c(v'))$ (where u' and v' are the vertices connected by edge i , and these are the respective preimages of vertices u and v) is close to μ_e . Second, the contribution of the non-nice vertices is controlled by δ and could be made arbitrarily small.

Let us show now that actually μ_e^c concentrates near its expected values. Take any two elements b_u, b_v from B . Edge e has a preimage set $L_e^{\widehat{G}_N}$ of edges from \widehat{G}_N . Define $(f_{i,e,b_u,b_v})_{i \in L_e^{\widehat{G}_N}}$ to be the variable that equals 1 whenever $c(u') = b_u$ and $c(v') = b_v$, and 0 otherwise. Denote

$$\overline{f_{e,b_u,b_v}} = \frac{1}{N} \sum_{i \in L_e^{\widehat{G}_N}} f_{i,e,b_u,b_v}.$$

Note that $\overline{f_{e,b_u,b_v}} = \mu_e^c(\{(b_u, b_v)\})$. Here lies the main difference of our proof from that of [BGH]. In that paper, a measure concentration bound was used. Since our process is not IID, we need to find another option. By the Chebyshev inequality, it amounts to prove the following:

Lemma 7. *For any $\varepsilon > 0$ there are large enough N' and R , and small enough $\delta > 0$, such that the following holds. For any (R, δ) -nice N -fold covering \widehat{G}_N with $N > N'$, we have:*

$$\text{Var}(\overline{f_{e,b_u,b_v}}) < \varepsilon.$$

The rest is devoted to the proof of the lemma above.

Let's explain the variant of the second-moment argument needed. Let $(Y_i)_{i=1 \dots N}$ be a collection of random variables on the same probability space with zero expectation each (we can always obtain the latter by subtracting respective expectation). Assume also that $|Y_i| \leq 1$ almost surely. Let

$$\overline{Y} = \frac{1}{N} \sum_{1 \leq i \leq N} Y_i.$$

We want to show that for big N , the second moment of \overline{Y} is close to 0. Of course we will need some additional assumptions. Let's expand:

$$\mathbb{E}\overline{Y}^2 = \frac{1}{N^2} \sum_i \mathbb{E}Y_i^2 + \frac{1}{N^2} \sum_{i \neq j} \mathbb{E}Y_i Y_j. \quad (3)$$

Note that the first summand vanishes as N gets bigger.

If Y_i 's are pairwise independent, then the second term of expression 3 is zero, and we are done. This is classical Chebyshev's law of large numbers. Let us weaken the independence assumption a little. Assume that each Y_i "interacts" with the uniformly bounded number of Y_j 's. That is there is a constant K (uniform in N) such that for each i there are at most K of j 's with Y_i and Y_j being pairwise dependent. In that case the second summand of 3 is bounded by K/N , and we obtain the desired.

For our exposition we need a further weakening. Roughly speaking, we say that each Y_i interacts in a “significant way” only to a uniformly bounded number of Y_j ’s. We assume that all pairs can interact, but for each $\varepsilon' > 0$ there is a number $K_{\varepsilon'}$ (uniform in N) such that for each i there are at most $K_{\varepsilon'}$ such j ’s that $|\mathbb{E}Y_i Y_j| > \varepsilon'$. So there are “far” pairs (i, j) with covariance less than ε' and the “close” ones with covariance not less than ε' . This assumption implies that the second term of the second moment is bounded by $\varepsilon' + K_{\varepsilon'}/N$.

We need to take into account now that some elements i are “bad” = “not nice” (we don’t have the bound on interaction for them as described above), but their number is limited by δN . In our expression for the second moment we will get:

$$\begin{aligned} \mathbb{E}\bar{Y}^2 &= \frac{1}{N^2} \sum_i \mathbb{E}Y_i^2 + \frac{1}{N^2} \sum_{i \neq j} \mathbb{E}Y_i Y_j = \\ &= \frac{1}{N^2} \sum_i \mathbb{E}Y_i^2 + \frac{1}{N^2} \sum_{\substack{i \neq j \\ \text{nice}}} \mathbb{E}Y_i Y_j + \frac{1}{N^2} \sum_{\substack{i \neq j \\ i \text{ or } j \text{ bad}}} \mathbb{E}Y_i Y_j \leq \\ &= \frac{1}{N} + \varepsilon' + \frac{K_{\varepsilon'}}{N} + 2\delta. \quad (4) \end{aligned}$$

We return to the proof of Lemma 7. By Lemma 5, we can take R so big that for any two R -nice edges $i, j \in L_e^{\hat{G}^N}$ that are $2R$ apart (in the usual graph distance on \hat{G}^N), we will have

$$\text{Cov}(f_{i,e,b_u,b_v}, f_{j,e,b_u,b_v}) \leq \varepsilon/4.$$

So for any R -nice edge i from $L_e^{\hat{G}^N}$ there are no more than $2R(\deg G)^{2R+2}$ R -nice edges j from $L_e^{\hat{G}^N}$ such that

$$\text{Cov}(f_{i,e,b_u,b_v}, f_{j,e,b_u,b_v}) > \varepsilon/4.$$

So in 4 we have $\varepsilon' = \varepsilon/2$ and $K_{\varepsilon'} = 2R(\deg G)^{2R+2}$. Now if we take δ smaller than $\varepsilon/2$, then for big enough N , the variance of $\overline{f_{e,b_u,b_v}}$ will be smaller than ε . This finishes the proof of Lemma 7 and thus of Lemmata 6 and 3. \square

References

- [A15] A. Alpeev, *The entropy of Gibbs measures on sofic groups* Zap. Nauchn. Sem. POMI, 436 (2015): 34–48.
- [A17] A. Alpeev, *A random ordering formula for Gibbs measures for the sofic and Rokhlin entropy of Gibbs measures*, arXiv preprint arXiv:1705.08559 (2017).

- [AuP17] T. Austin and M. Podder, *Gibbs measures over locally tree-like graphs and percolative entropy over infinite regular trees*, arXiv preprint arXiv:1705.03589 (2017).
- [BS14] A. Backhausz and B. Szegedy, *On large girth regular graphs and random processes on trees*, arXiv preprint arXiv:1406.4420 (2014).
- [BGHV] A. Backhausz, B. Gerencsér, V. Harangi, and M. Vizer, *Correlation bound for distant parts of factor of iid processes*, *Combin. Probab. Comput.*, published online (2017).
- [BSV15] A. Backhausz, B. Szegedy, and B. Virág, *Ramanujan graphings and correlation decay in local algorithms*, *Random Structures Algorithms*, 47(3):424–435, (2015).
- [BGH] A. Backhausz, B. Gerencsér and V. Harangi, *Entropy inequalities for factors of IID*, to appear in *Grops, Geometry and Dynamics*, arXiv preprint arXiv:1706.04937 (2018).
- [Ba] R. J. Baxter *Exactly solved models in statistical mechanics*, Academic Press, 1982.
- [CHV] E. Csóka, V. Harangi, and B. Virág *Entropy and expansion*, arXiv preprint arXiv:1811.09560 (2018).
- [DM10a] A. Dembo and A. Montanari, *Gibbs measures and phase transitions on sparse random graphs*, *Brazilian Journal of Probability and Statistics* (2010): 137-211.
- [DM10b] A. Dembo and A. Montanari, *Ising models on locally tree-like graphs*, *The Annals of Applied Probability* 20.2 (2010): 565-592.
- [EW11] M. Einsiedler and T. Ward. *Ergodic theory with a view towards number theory*. Graduate texts in mathematics, 259. Springer, London, 2011.
- [G11] H.-O. Georgii, *Gibbs measures and phase transitions*. Vol. 9. Walter de Gruyter, 2011.
- [GH17] B. Gerencsér and V. Harangi, *Mutual information decay for factors of iid*, *Ergodic Theory and Dynamical Systems*, to appear, arXiv:1703.04387 (2017).
- [RS15] F. Rassoul-Agha and T. Seppäläinen, *A course on Large Deviations with an Introduction to Gibbs Measures*, Vol. 162. American Mathematical Soc., 2015.