# ENTROPY INEQUALITIES FOR FACTORS OF IID

ÁGNES BACKHAUSZ, BALÁZS GERENCSÉR, AND VIKTOR HARANGI

ABSTRACT. This paper is concerned with certain invariant random processes (called factors of IID) on infinite trees. Given such a process, one can assign entropies to different finite subgraphs of the tree. There are linear inequalities between these entropies that hold for any factor of IID process (e.g. "edge versus vertex" or "star versus edge"). These inequalities turned out to be very useful: they have several applications already, the most recent one is the Backhausz–Szegedy result on the eigenvectors of random regular graphs.

We present new entropy inequalities in this paper. In fact, our approach provides a general "recipe" for how to find and prove such inequalities. Our key tool is a generalization of the edge-vertex inequality for a broader class of factor processes with fewer symmetries.

## 1. INTRODUCTION

1.1. **Entropy inequalities for processes on $T_d$.** For an integer $d \geq 3$ let $T_d$ denote the $d$-regular tree: the (infinite) connected graph with no cycles and with each vertex having exactly $d$ neighbors.

The main focus of this paper is the class of *factor of IID processes*. Loosely speaking, independent and identically distributed (say uniform $[0, 1]$) random labels are assigned to the vertices of $T_d$, then each vertex gets another label (a *state* chosen from a finite *state space M*) that depends on the labeled rooted graph as seen from that vertex, all vertices "using the same rule". This way we get a probability distribution on $M^{V(T_d)}$ (called a factor of IID) that is invariant under the *automorphism group* $\mathrm{Aut}(T_d)$ of $T_d$. A formal definition will be given in Section 1.2 below.

One of the reasons why factor of IID processes have attracted a growing attention in recent years is that they give rise to randomized local algorithms that can be carried out on arbitrary regular graphs with "large essential girth", e.g. random regular graphs. See [9, 10, 13, 14] how factors of IID/local algorithms can be used to obtain large independent sets for large-girth graphs. Factors of IID are also studied by ergodic theory under the name of *factors of Bernoulli shifts*, see Section 2.5 for details.

The starting point of our investigations is the following *edge-vertex entropy inequality* that holds for any factor of IID process on $T_d$:

$$(1) \qquad \frac{d}{2} H(\text{\textbf{:}}) \geq (d-1) H(\bullet).$$

Here • represents a vertex, and $H(\bullet)$ is the (Shannon) entropy of the (random) state of a vertex. Similarly, $\updownarrow$ represents an edge, and $H(\updownarrow)$ stands for the entropy of the joint distribution of the states of two neighbors. (Note that the state space $M$ is assumed to be finite here.) This inequality can be found implicitly in Lewis Bowen's work from 2009 [8]. Rahman and Virág proved it in a special setting [17]. A full and concise proof was given by Backhausz and Szegedy in [2]; see also [16]. The counting argument behind this inequality actually goes back to a result of Bollobás on the independence ratio of random regular graphs [6].

A *star-edge entropy inequality* was also proved in [2]:

$$(2) \qquad\qquad H\left(\overset{\cdot}{\underset{\cdot}{\Leftarrow}} d\right) \geq \frac{d}{2} H\left(\updownarrow\right),$$

where $H(\overset{\cdot}{\underset{\cdot}{\Leftarrow}} d)$ denotes the entropy of the joint distribution of the states of a vertex and its $d$ neighbors. (Note that because of the $\mathrm{Aut}(T_d)$-invariance the distribution of every vertex/edge/star is the same.)

The above inequalities played a central role in a couple of intriguing results recently: the Rahman–Virág result [17] about the maximal size of a factor of IID independent set on $T_d$ and the Backhausz–Szegedy result [3] on the "local statistics" of eigenvectors of random regular graphs.

The goal of this paper is to obtain further inequalities between the entropies corresponding to different subgraphs of $T_d$. The ultimate goal would be to somehow describe the class of (linear) entropy inequalities that hold for any factor of IID process. We make progress towards this goal in this paper by developing a general method that can be used to find and prove such inequalities. See Section 1.3 for some examples of the new inequalities that this method produces. These examples include an upper bound for the (normalized) mutual information of two vertices at distance $k$. Another inequality we obtain is $H(\overset{\cdot}{\underset{\cdot}{\Leftarrow}} d) \geq (d-1)H(\bullet)$, where $\overset{\cdot}{\underset{\cdot}{\Leftarrow}} d$ represents the $d$ neighbors of any given vertex in $T_d$. This inequality can be used to improve earlier results about *tree-indexed Markov chains*, see Section 4.3 for details.

1.2. **General edge-vertex entropy inequalities.** Our key tool is a generalization of the edge-vertex inequality (1) for processes with weaker invariance properties. For a given finite connected simple graph $G$ (that is not a tree itself) the universal cover is an infinite "periodic" tree $T$. Let $\Gamma$ be a subgroup of the automorphism group $\mathrm{Aut}(T)$. By a $\Gamma$-*invariant process* over (the vertex set $V(T)$ of) $T$ we mean a probability distribution on $M^{V(T)}$ that is invariant under the natural $\Gamma$-action. Although this makes sense for any measurable space $M$, in this paper the *state space* $M$ will always be a finite set (with the discrete $\sigma$-algebra).

Now we define factors of IID in this more general setting. A measurable function $F \colon [0,1]^{V(T)} \to M^{V(T)}$ is said to be a $\Gamma$-*factor* if it is $\Gamma$-equivariant, that is, it commutes with the natural $\Gamma$-actions. Given an IID process $Z = (Z_v)_{v \in V(T)}$ on $[0,1]^{V(T)}$, applying $F$ yields a factor of IID process $X = F(Z)$, which can be viewed as a collection $X = (X_v)_{v \in V(T)}$ of $M$-valued random variables. It follows immediately that the distribution of $X$ is indeed $\Gamma$-invariant.

In the special case when the degree of each vertex of $G$ is the same (that is, when $G$ is $d$-*regular* for some $d$) the universal cover is the $d$-regular tree $T_d$. If we simply say *factor of IID process on $T_d$* (without specifying the group $\Gamma$), we usually refer to the case when $\Gamma$ is the full automorphism group $\mathrm{Aut}(T_d)$. The edge-vertex inequality (1) holds for any

$\mathrm{Aut}(T_d)$-factor of IID process. The next theorem and its corollary provide generalizations of (1) for certain subgroups $\Gamma$ of $\mathrm{Aut}(T_d)$.

**Theorem 1.** *Suppose that $G$ is a finite connected (simple) graph, $T$ is the universal cover of $G$, $\varphi \colon T \to G$ is an arbitrary fixed covering map. By $\Gamma_\varphi \leq \mathrm{Aut}(T)$ we denote the group of* covering transformations, *that is, the automorphisms $\gamma \in \mathrm{Aut}(T)$ for which $\varphi \circ \gamma = \varphi$. Let $M$ be a finite state space and $X$ a $\Gamma_\varphi$-factor of IID process on $M^{V(T)}$. Given a vertex $v$ of the base graph $G$ let $\mu_v^X$ denote the distribution of $X_{\hat{v}}$ for any lift $\hat{v}$ of $v$. Similarly, for an edge $e \in E(G)$ let $\mu_e^X$ be the joint distribution of $(X_{\hat{u}}, X_{\hat{v}})$ for any lift $\hat{e} = (\hat{u}, \hat{v})$ of $e$. Note that these distributions are well defined because of the $\Gamma_\varphi$-invariance of the process. Then the Shannon entropies of these distributions satisfy the following inequality:*

$$(3) \qquad \sum_{e \in E(G)} H(\mu_e^X) \geq \sum_{v \in V(G)} (\deg v - 1) H(\mu_v^X),$$

*where $\deg v$ is the degree (i.e. number of neighbors) of the vertex $v$ in $G$.*

Compare this with the trivial upper bound $\sum_{e \in E(G)} H(\mu_e^X) \leq \sum_{v \in V(G)} \deg(v) H(\mu_v^X)$, where we have equality if and only if the states of two neighbors are independent. Thus the above theorem can be considered as a quantitative result as to "how independent" neighboring states are in a factor of IID process.

We state the special case when $G$ is $d$-regular in a separate corollary.

**Corollary 2.** *Let $\varphi \colon T_d \to G$ be a covering map for a finite $d$-regular connected (simple) graph $G$ with $d \geq 3$. Using the notations of Theorem 1, for any $\Gamma_\varphi$-factor of IID process on $M^{V(T_d)}$ it holds that*

$$(4) \qquad \sum_{e \in E(G)} H(\mu_e^X) \geq (d-1) \sum_{v \in V(G)} H(\mu_v^X).$$

This essentially says that (1) holds for $\Gamma_\varphi$-factors if $H(\bullet)$ and $H(\text{⁝})$ are replaced by the average of the entropies of different "types" of vertices/edges. (Note that the number of edges is equal to $d/2$ times the number of vertices.) This means that the original edge-vertex entropy inequality (1) for $\mathrm{Aut}(T_d)$-factors follows from (4) for any $d$-regular $G$. Indeed, given an $\mathrm{Aut}(T_d)$-factor, it is also a $\Gamma_\varphi$-factor with the extra property that each vertex/edge has the same distribution.

Another special case of Corollary 2 is a result of Lewis Bowen saying that the so-called *f-invariant* is non-negative for factors of Bernoulli shifts, see Section 2.5.

We will prove Theorem 1 in Section 5 by considering random finite lifts of the base graph $G$ and counting the (expected) number of $M$-colorings on these lifts with the property that the "local statistics" of the coloring is close to that of the process $X$.

1.3. **New inequalities.** As we have mentioned, if we apply Corollary 2 to an $\mathrm{Aut}(T_d)$-factor, then we simply get the original version (1). Hence it appears, falsely, that these more general inequalitites cannot be used to obtain new results in the most-studied special case of $\mathrm{Aut}(T_d)$-factors.

The point is that starting from an $\mathrm{Aut}(T_d)$-factor of IID process $Y$ on $T_d$, there are many ways to turn this into a $\Gamma_\varphi$-factor $X$ because one can use the extra structure on $T_d$ given by a covering $\varphi \colon T_d \to G$. Then applying Corollary 2 to this new process $X$ yields an inequality for the original process $Y$. We demonstrate this on the following simple example. Let $G = K_{d+1}$ be the complete graph on $d+1$ vertices which is clearly $d$-regular. Let $o$ denote a distinguished vertex of $G$. Given a $T_d \to G$ covering map $\varphi$, every vertex of
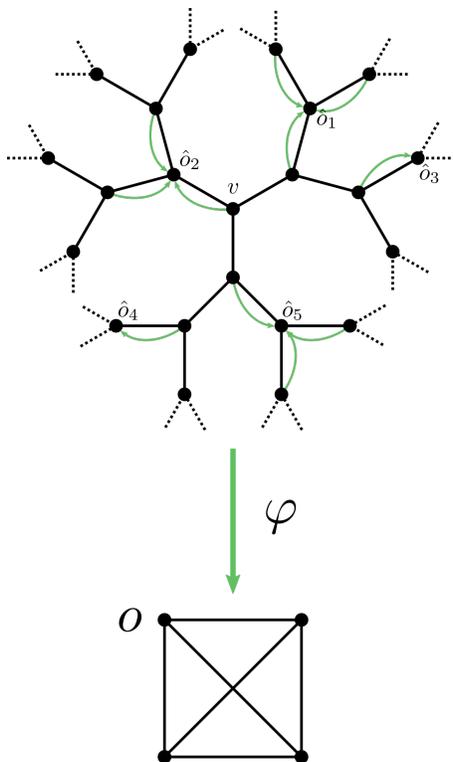
FIGURE 1. The case $d = 3$: $\hat{o}_i$ are the lifts of $o \in V(K_4)$; $X_v$ is defined as $Y_{\hat{o}_2}$.

$T_d$ is either a lift of $o$, or has a unique neighbor that is a lift of $o$ (see Figure 1). Suppose that $Y$ is an $\mathrm{Aut}(T_d)$-factor of IID on $T_d$, and set

$$X_v := Y_{v'}, \text{ where } v' \text{ is the unique vertex such that } \varphi(v') = o \text{ and } \mathrm{dist}(v, v') \leq 1.$$

It is easy to see that $X = (X_v)_{v \in V(T_d)}$ is a $\Gamma_\varphi$-factor of IID and hence Corollary 2 can be applied to $X$. Given two neighboring vertices $u$ and $v$ in $T_d$, the corresponding $u'$ and $v'$ either coincide (if $\varphi(u) = o$ or $\varphi(v) = o$), or they have distance 3. It follows that

$$H(\mu_e^X) = \begin{cases} H(\bullet) & \text{if the edge } e \in E(G) \text{ is incident to } o, \\ H(\bullet\!\!-\!\circ\!-\!\circ\!-\!\bullet) & \text{otherwise,} \end{cases}$$

where $\bullet\!\!-\!\circ\!-\!\circ\!-\!\bullet$ represents two vertices of distance 3, and the notations $H(\bullet)$ and $H(\bullet\!\!-\!\circ\!-\!\circ\!-\!\bullet)$ refer to entropies corresponding to the ($\mathrm{Aut}(T_d)$-factor of IID) process $Y$. Substituting these and $H(\mu_v^X) = H(\bullet)$ into (4) we obtain, after cancellations, the following inequality for the process $Y$:

$$H(\bullet\!\!-\!\circ\!-\!\circ\!-\!\bullet) \geq \left( 2 - \frac{2}{d(d-1)} \right) H(\bullet).$$

This actually means that the normalized mutual information $I(Y_u; Y_v)/H(Y_v)$ is at most $\frac{2}{d(d-1)}$ for any vertices $u$ and $v$ of distance 3 in $T_d$. The above argument can be generalized to obtain the following bounds for the normalized mutual information for arbitrary distance $\mathrm{dist}(u, v) = k$. This result was already proved in our earlier paper [12] for the case when the degree $d$ is even. The current proof works for arbitrary $d \geq 3$.

**Theorem 3.** [12, Theorem 1] *Let $d \geq 3$ be an integer. For any $u, v \in V(T_d)$ at distance $k$ and for any $\mathrm{Aut}(T_d)$-factor of IID process $Y$ on $T_d$ we have*

(5)
$$\frac{I(Y_u; Y_v)}{H(Y_v)} \leq \begin{cases} \frac{2}{d(d-1)^l} & \text{if } k = 2l + 1 \text{ is odd,} \\ \frac{1}{(d-1)^l} & \text{if } k = 2l \text{ is even.} \end{cases}$$

Our general method is described in Section 3, it provides countless new entropy inequalities. We list a few examples in the rest of the introduction.

Let us fix an $\mathrm{Aut}(T_d)$-factor of IID process $Y$. Then for a finite set $V \subset V(T_d)$ the entropy of the joint distribution of $Y_v$, $v \in V$, will be denoted by $H(V)$. Because of the $\mathrm{Aut}(T_d)$-invariance of the process this joint distribution, and hence $H(V)$, depends only on the "isomorphism type" of $V$ in $T_d$.

For instance, if $V$ consists of the four vertices of a path of length three, then we do not need to specify where this path is in $T_d$ and we can simply write $H(\bullet\text{-}\bullet\text{-}\bullet)$ for $H(V)$. The next theorem compares $H(\bullet\text{-}\bullet\text{-}\bullet)$ to $H(\mathbf{:})$.

**Theorem 4.** *The following* path-edge inequality *holds for any $\mathrm{Aut}(T_d)$-factor of IID process on $T_d$:*
$$H(\bullet\text{-}\bullet\text{-}\bullet) \geq \frac{2d-3}{d-1} H(\mathbf{:}).$$

Another new inequality we obtain is

(6)
$$H(\mathord{\Lsh} d) \geq (d-1) H(\bullet).$$

The following two theorems generalize this inequality in different ways.

**Theorem 5.** *Let $S_k$ denote the set of vertices at distance $k$ from a fixed vertex of $T_d$. Then for any $\mathrm{Aut}(T_d)$-factor of IID process it holds that*

(7)
$$H(S_k) \geq (d-1)^k H(\bullet).$$

**Theorem 6.** *Let $\mathord{\Lsh} i$ denote the set of $i$ neighbors of a fixed vertex. Then for any $\mathrm{Aut}(T_d)$-factor of IID process and for any $1 \leq i < d$ it holds that*
$$(d-i) H(\mathord{\Lsh} i{+}1) \geq (d-i-1) H(\mathord{\Lsh} i) + (d-1) H(\bullet),$$
*and hence by induction for any $1 \leq i \leq d$:*
$$H(\mathord{\Lsh} i) \geq \frac{id - 2i + 1}{d-1} H(\bullet), \quad \text{in particular, } H(\mathord{\Lsh} d) \geq (d-1) H(\bullet).$$

We will see in Section 4 that each of these inequalities is sharp in the sense that there are $\mathrm{Aut}(T_d)$-factors of IID processes for which the two sides of the inequality are asymptotically equal. We will also examine how strong our new inequalities are: it turns out that (6) and (7) are stronger than (1) and (2) for Markov chains indexed by $T_d$.

**Outline of the paper.** The rest of the paper is structured as follows. In Section 2 we go through basic definitions and elaborate on the strength of Theorem 1 for different base graphs. In Section 3 we describe our general method for deriving new entropy inequalities from our general edge-vertex inequalities. In Section 4 we show that these new inequalities are sharp, and we compare them to previously-known ones. Finally, the proof of Theorem 1 is given in Section 5.

## 2. Preliminaries

2.1. **Factors of IID.** Suppose that a group $\Gamma$ acts on a countable set $S$. Then $\Gamma$ also acts on the space $M^S$ for a set $M$: for any function $f\colon S \to M$ and for any $\gamma \in \Gamma$ let

$$(8) \qquad\qquad (\gamma \cdot f)(s) := f(\gamma^{-1} \cdot s) \quad \forall s \in S.$$

First we define the notion of factor maps.

**Definition 2.1.** Let $M_1, M_2$ be measurable spaces and $S_1, S_2$ countable sets with a group $\Gamma$ acting on both. A measurable mapping $F\colon M_1^{S_1} \to M_2^{S_2}$ is said to be a $\Gamma$-*factor* if it is $\Gamma$-equivariant, that is, it commutes with the $\Gamma$-actions.

By an *invariant process* on $M^S$ we mean an $M^S$-valued random variable (or a collection of $M$-valued random variables) whose (joint) distribution is invariant under the $\Gamma$-action. For example, if $Z_s$, $s \in S_1$, are independent and identically distributed $M_1$-valued random variables, then we say that $Z = (Z_s)_{s \in S_1}$ is an IID process on $M_1^{S_1}$. Given a $\Gamma$-factor $F\colon M_1^{S_1} \to M_2^{S_2}$, we say that $X := F(Z)$ is a $\Gamma$-*factor of the IID process* $Z$. It can be regarded as a collection of $M_2$-valued random variables: $X = (X_s)_{s \in S_2}$.

The results of this paper are concerned with factor of IID processes on infinite trees $T$: $S_1$ and $S_2$ are the vertex set $V(T)$ and $\Gamma$ is a subgroup of the automorphism group $\mathrm{Aut}(T)$. The most important special case is $T = T_d$ and $\Gamma = \mathrm{Aut}(T_d)$. When we say $\Gamma$-factor of IID process, we should also specify which IID process we have in mind (that is, specify $M_1$ and a probability distribution on it). By default we will work with the uniform distribution on $[0, 1]$. In fact, as far as the class of $\mathrm{Aut}(T_d)$-factors is concerned, it does not really matter which IID process we consider. For example, for the uniform distribution on $\{0, 1\}$ we get the same class of factors as for the uniform distribution on $[0, 1]$. This follows from the fact that these two IID processes are $\mathrm{Aut}(T_d)$-factors of each other [5].

The other important special case is when $T$ is the universal cover of a finite connected simple graph $G$ and $\Gamma = \Gamma_\varphi$ is the group of covering transformations for a covering $\varphi\colon T \to G$. In this case it holds that for any $\hat{v}_1, \hat{v}_2 \in V(T)$ with $\varphi(\hat{v}_1) = \varphi(\hat{v}_2)$ there exists a unique $\gamma \in \Gamma_\varphi$ such that $\gamma(\hat{v}_1) = \hat{v}_2$. It follows that if we choose a fixed pre-image $\bar{v} \in \varphi^{-1}(v)$ for every vertex $v \in V(G)$ of the base graph, then a $\Gamma_\varphi$-factor $F\colon [0,1]^{V(T)} \to M^{V(T)}$ is determined by the functions $f_{\bar{v}} := \pi_{\bar{v}} \circ F\colon [0,1]^{V(T)} \to M$, where $\pi_{\bar{v}}$ denotes the coordinate projection $M^{V(T)} \to M$ corresponding to the vertex $\bar{v}$. Conversely, any collection of measurable functions $f_{\bar{v}}\colon [0,1]^{V(T)} \to M$, $v \in V(G)$, gives rise to a $\Gamma_\varphi$-factor mapping. (Note that an $\mathrm{Aut}(T_d)$-factor $F$ is determined by a single function $f_o := \pi_o \circ F\colon [0,1]^{V(T)} \to M$, but in that case $f_o$ needs to be invariant under all automorphisms of $T_d$ fixing the vertex $o \in V(T_d)$. See [1, Section 2.1] for details.)

2.2. **Finite-radius factors.** Let $X$ be a $\Gamma$-factor of the IID process $Z$. We say that $X$ is a *finite-radius factor* (or a *block factor*) if there exists a positive integer $R$ such that for any vertex $v$ the value of $X_v$ depends only on the values $Z_u$ for vertices $u$ in the $R$-neighborhood around $v$.

Can a factor of IID process be approximated by finite-radius factors? In many cases the answer is positive. This means that it suffices to prove certain statements for finite-radius factors. For instance, in the proof of Theorem 1 we will need the fact that an arbitrary $\Gamma_\varphi$-factor of IID process is the weak limit of finite-radius $\Gamma_\varphi$-factors. As we have seen, a $\Gamma_\varphi$-factor $F$ is determined by finitely many measurable $[0,1]^{V(T)} \to M$ maps. The pre-image of an element $m \in M$ under such a map is a measurable set in the product space $[0,1]^{V(T)}$, and, as such, it can be approximated by a finite union of measurable cylinder
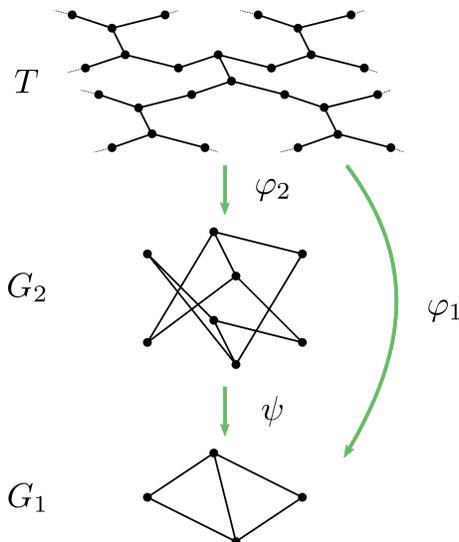
FIGURE 2. Coverings $T \to G_2 \to G_1$

sets. Since $M$ is finite in our case, it follows that any measurable $[0,1]^{V(T)} \to M$ map can be approximated by maps for which all the pre-images are finite unions of cylinder sets, and consequently any $\Gamma_\varphi$-factor can be approximated by finite-radius factors.

2.3. **Finite coverings.** Theorem 1 provides an inequality for any finite base graph $G$. Next we elaborate on how these inequalities are related to each other.

Suppose that $G_1$ and $G_2$ are finite connected (simple) graphs such that there is a covering map $\psi\colon G_2 \to G_1$. Then the $G_2$-version of Theorem 1 is stronger than the $G_1$-version. Indeed, let $T$ denote their universal cover. Given a covering map $\varphi_2\colon T \to G_2$, setting $\varphi_1 := \psi \circ \varphi_2$ yields a $T \to G_1$ covering map, see Figure 2. Clearly $\Gamma_{\varphi_2} \le \Gamma_{\varphi_1}$. It follows that any $\Gamma_{\varphi_1}$-factor of IID process $X$ on $T$ is also $\Gamma_{\varphi_2}$-factor with the extra property that $\mu_v^X$ ($\mu_e^X$) depends only on the $\psi$-image of $v \in V(G_2)$ ($e \in E(G_2)$). Therefore it is easy to see that if we take the $G_2$-version of the general edge-vertex inequality (3) and apply it to a $\Gamma_{\varphi_1}$-factor, we simply get back the $G_1$-version of (3).

This means that one can get stronger and stronger versions of (3) by repeatedly lifting the finite base graph $G$.

2.4. **Multiple edges and loops.** A graph is called *simple* if it does not contain loops or multiple edges. For the sake of simplicity we stated (and we will prove) Theorem 1 for the case when the base graph $G$ is simple. What can be said for base graphs that are not simple?

If $G$ has multiple edges (but no loops), then essentially the same result holds. The only difference is in the definition of a *covering map* $T \to G$. In the case of simple graphs, one can simply say that a covering map is a mapping $V(T) \to V(G)$ such that the neighbors of a vertex $v$ are mapped bijectively to the neighbors of the image of $v$. When we have multiple edges, we also need to define the image of an edge: a covering map is a mapping $V(T) \to V(G)$ and a mapping $E(T) \to E(G)$ such that edges incident to a vertex $v$ are mapped bijectively to edges incident to the image of $v$. Once we know Theorem 1 for simple base graphs, it easily follows that it also holds when the base graph $G$ has multiple edges: simply take a finite simple graph $G_2$ that covers $G$; then the $G_2$-version of (3) implies the

$G$-version. (The proof of Theorem 1 presented in Section 5 would actually work for base graphs with multiple edges.)

As for loops the situation is a bit more complicated. In fact, one should distinguish between two kinds of loops. Loosely speaking: a *full-loop* can be travelled in two directions (contributing to the degree of the vertex by 2 and adding a free factor $\mathbb{Z}$ to the fundamental group) while for a *half-loop* there is just one way of "going around" (contributing to the degree by only 1 and adding a free factor $\mathbb{Z}_2$ to the fundamental group). For our purposes the difference between them is how they behave under coverings. In short, an edge "double-covers" a half-loop while two parallel edges are needed to double-cover a full-loop. We should define covering maps rigorously for graphs containing full-loops, half-loops, multiple edges. Then this could lead to a version of (3) for arbitrary base graphs. The reason why we do not go into the details here is that, again, one can always take a finite simple lift of an arbitrary base graph and get a stronger version of the inequality.

If $G$ has parallel edges (multiple edges between two vertices or more than one loops at one vertex), then we may choose not to "distinguish" some of those parallel edges but this would again lead to weaker inequalities. Note that in this terminology the original edge-vertex inequality (1) would correspond to the case when the base graph $G$ consists of one vertex and $d$ undistinguished half-loops, which is the weakest version of (4) in the $d$-regular case.

2.5. **Connections to dynamical systems.** These processes can be viewed in the context of ergodic theory. An invariant process (as defined in Section 2.1) gives rise to a dynamical system over $\Gamma$: the group $\Gamma$ acts by measure-preserving transformations on the measurable space $M^S$ equipped with a probability measure (the distribution of the invariant process). An IID process simply corresponds to a (generalized) Bernoulli shift. Therefore factor of IID processes are factors of Bernoulli shifts.

In fact, the general edge-vertex inequality (3) is related to a result of Lewis Bowen saying that the so-called $f$-*invariant* (for actions of the free group $F_r$) is non-negative for factors of the Bernoulli shift [8, Corollary 1.8]. This is essentially equivalent to Corollary 2 in the special case when the base graph $G$ consists of one vertex and $r = d/2$ distinguished full-loops. See [12, Section 2.3] for details.

## 3. NEW INEQUALITIES FOR $\mathrm{Aut}(T_d)$-FACTORS

In the introduction we already demonstrated on a simple example how Corollary 2 can be used to get new entropy inequalities for $\mathrm{Aut}(T_d)$-factors. In this section we describe our general method and present further examples.

Suppose that $Y$ is an $\mathrm{Aut}(T_d)$-factor of IID process on $M^{V(T_d)}$. Using the extra structure that a covering $\varphi\colon T_d \to G$ gives, $Y$ can be turned into a $\Gamma_\varphi$-factor in many ways. For any $v \in V(G)$ let us fix a non-backtracking walk starting at $v$. Then for any lift $\hat{v} \in V(T_d)$ of $v$ this walk can be lifted to get a path starting at $\hat{v}$. Let the endpoint of this path be assigned to $\hat{v}$. This assignment yields a mapping $f\colon V(T_d) \to V(T_d)$. It is easy to see that $f$ is $\Gamma_\varphi$-equivariant, and consequently $X_u := Y_{f(u)}$ defines a process $X$ that is a $\Gamma_\varphi$-factor of IID, and hence Corollary 2 can be applied to $X$. (The example in the introduction is the special case when $G = K_{d+1}$, and for the distinguished vertex $o \in V(G)$ we choose the walk $o$ of length 0, and for any other vertex $v$ we choose the walk $v \to o$ of length 1.)

The general construction (where one can choose a finite collection of walks for each vertex) is described by the following lemma.

**Lemma 3.1.** *Let $G$ be a finite connected $d$-regular (simple) graph and $\varphi \colon T_d \to G$ a covering map. Suppose that we have an $\mathrm{Aut}(T_d)$-factor of IID process $Y$ on $M^{V(T_d)}$. For any $v \in V(G)$ let us choose a finite collection of (non-backtracking) walks on $G$ (each starting at $v$): $W_{v,i}$, $1 \le i \le k_v$.*

*For any lift $\hat{v} \in V(T_d)$ of $v$ we lift each $W_{v,i}$ starting at $\hat{v}$. Then we consider the endpoints of these $k_v$ paths and $X_{\hat{v}}$ is defined to be the $k_v$-tuple of the $Y$-labels of these endpoints. It can be seen easily that the obtained process $X$ is a $\Gamma_\varphi$-factor of the IID process. (Note that the state space for $X$ is $M' = M \cup (M \times M) \cup (M \times M \times M) \cup \ldots$)*

If we apply Corollary 2 to this process $X$, then we will get an inequality between the entropies of various finite subsets of $V(T_d)$ for the original $\mathrm{Aut}(T_d)$-factor of IID process $Y$. This works for any choice of a finite $d$-regular base graph $G$ and walks $W_{v,i}$. In the remainder of this section we will show a few specific examples.

To keep our notations simple, in this section we will write $\mu_v$ and $\mu_e$ for $\mu_v^X$ and $\mu_e^X$. Also, $H(\bullet)$ or $H(\mathbf{\updownarrow})$, and more generally $H(V)$ for some $V \subset V(T_d)$, will always refer to the entropy corresponding to the original $\mathrm{Aut}(T_d)$-factor process $Y$.

**Two-vertex base graph, Theorem 4 and 6.** As we discussed in Section 2.4 the general edge-vertex inequality is true even when the base graph $G$ has multiple edges. So let $G$ be the graph with two vertices ($u$ and $v$) and $d$ multiple edges $e_1, \ldots, e_d$ between them. Given a positive integer $i \le d-1$ the following $i$ walks (of length 1) are associated to $u$: $u \xrightarrow{e_1} v; \ldots; u \xrightarrow{e_i} v$; while only the zero-length walk $v$ is associated to $v$. Then

$$H(\mu_u) = H(\langle\!\!\!\cdot\, i); \quad H(\mu_v) = H(\bullet); \quad H(\mu_{e_j}) = \begin{cases} H(\langle\!\!\!\cdot\, i) & \text{if } j \le i, \\ H(\langle\!\!\!\cdot\, i+1) & \text{if } j > i. \end{cases}$$

Substituting these into (4) we get the first inequality in Theorem 6. The second inequality follows easily by induction.

Next we consider the same base graph with different associated walks. Two walks starting at $u$, namely, $u$ and $u \xrightarrow{e_1} v$; and two walks starting at $v$, namely, $v$ and $v \xrightarrow{e_1} u$. It is easy to see that $H(\mu_u) = H(\mu_v) = H(\mu_{e_1}) = H(\mathbf{\updownarrow})$, while for $j \ge 2$ we have $H(\mu_{e_j}) = H(\bullet\!\!\bullet\!\!\bullet\!\!\bullet)$, and consequently Theorem 4 follows from (4).

**Sphere versus vertex, Theorem 5.** For a set $V \subset V(T_d)$ and a non-negative integer $k$ let $B_k(V) := \{u \; : \; \mathrm{dist}(u, V) \le k\}$. The $k$-ball $B_k(\{o\})$ around some root $o$ will be denoted by $B_k$, while $S_k := B_k \setminus B_{k-1} = \{u \; : \; \mathrm{dist}(o, u) = k\}$ is the sphere of radius $k$. Our goal is to get an inequality between $H(S_k)$ and $H(\bullet)$.

We will need the following auxiliary graph to define our base graph: let $T_{d,k}$ denote a finite tree that is isomorphic to the subgraph of $T_d$ induced by the $k$-ball $B_k$. The vertex set of $T_{d,k}$ can be partitioned into levels $0, 1, \ldots, k$ (based on the distance to the root), level $i > 0$ consisting of $d(d-1)^{i-1}$ vertices. All vertices have degree $d$ except vertices at level $k$ having degree 1. Any vertex at level $0 < i < k$ is connected to one vertex at level $i-1$ and $d-1$ vertices at level $i+1$.

Now we take $d$ copies of $T_{d,k}$ and "glue" them along their level-$k$ vertices. This way we get a $d$-regular base graph $G$ (essentially $d$ balls of radius $k$ with a shared boundary). See Figure 3 for the case $d = k = 3$. The level-$k$ vertices (that is, vertices on the shared boundary that we will denote by $B$) only get the zero-length walks. Any other vertex $v$ belongs to exactly one copy of $T_{d,k}$. If we only use edges in this copy, then there is a unique
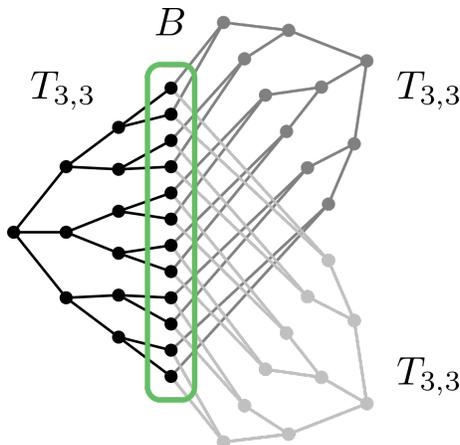
FIGURE 3. Three copies of $T_{3,3}$ glued together along their boundary $B$

path from $v$ to each vertex in $B$; let us associate these $|B|$ paths to $v$. Then we have

$$H(\mu_v) = \begin{cases} H(\bullet) & \text{if } v \in B, \\ H(S_k) & \text{if } v \notin B; \end{cases} \text{ and } H(\mu_e) = H(S_k) \text{ for any } e \in E(G).$$

Using (4) we get that

$$d\,|E(T_{d,k})|\,H(S_k) \geq \underbrace{(d-1)|B|}_{d(d-1)^k}\,H(\bullet) + d\underbrace{(d-1)\,(|V(T_{d,k})| - |B|)}_{|E(T_{d,k})|-1}\,H(S_k),$$

and Theorem 5 follows.

**Blow-ups.** By a *blow-up* of an entropy inequality we mean the inequality we get if we replace each $H(V)$ with $H(B_k(V))$ for a fixed positive integer $k$. It is not hard to show that if a linear entropy inequality is true for all $\mathrm{Aut}(T_d)$-factors of IID, then the blow-ups of this inequality are also true for all $\mathrm{Aut}(T_d)$-factors of IID.

For example, the blow-ups of the original edge-vertex inequality are:

(9) $$\frac{d}{2}H\left(B_k(\updownarrow)\right) \geq (d-1)H\left(B_k(\bullet)\right).$$

These blow-ups are closely related to Bowen's definition of the $f$-invariant [7, 8]; in particular, (9) follows from these papers.

There is a very short proof for (9) using our general method: one can take any base graph $G$ and for each vertex take all non-backtracking random walks of length at most $k$. It is easy to see that every $H(\mu_v)$ equals $H\left(B_k(\bullet)\right)$ and every $H(\mu_e)$ equals $H\left(B_k(\updownarrow)\right)$, and hence we get (9). Moreover, if an inequality is attainable by our method, then so are its blow-ups: one needs to replace each associated walk in $G$ with all walks obtained by concatenating this walk and any walk of length at most $k$.

We also mention that in [3] the blow-ups of the star-edge inequality (2) were proved for a broader class of invariant processes that were called *typical processes*. These blow-up inequalities played a central role in the proof of the main result of that paper. (Loosely speaking, a process is typical if it arises as a limit of labelings of random $d$-regular graphs. Their significance lies in the fact that many questions about random regular graphs can be studied through typical processes. It would be very interesting to know whether our new inequalities are also true for this broader class.)
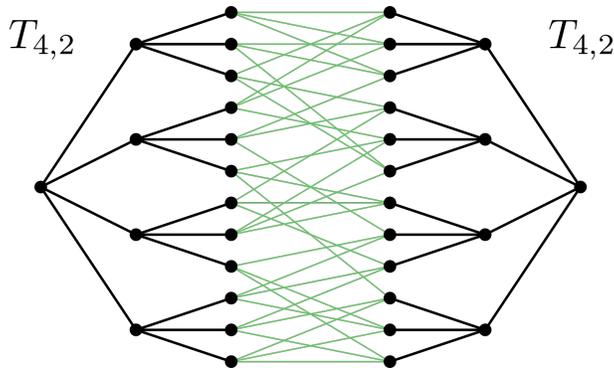
FIGURE 4. Two disjoint copies of $T_{4,2}$ with additional edges between the boundaries

**Mutual information decay, Theorem 3.** As we pointed out in the introduction, Theorem 3 was already proved in our earlier paper [12]. The proof therein works only for even $d$, however. Next we show how this inequality follows easily from Corollary 2 for arbitrary $d \geq 3$.

We need to define the base graph $G$ slightly differently for odd and even $k$. For an odd distance $k = 2l + 1$ let us take two copies of $T_{d,l}$ and add edges between their boundaries (that is, their level-$l$ parts) in such a way that the obtained graph $G$ is $d$-regular. Figure 4 shows the base graph for the case when $d = 4; k = 5; l = 2$.

As for the case when $k = 2l$ is even, one needs to connect the boundaries of a $T_{d,l}$ and a $T_{d,l-1}$. Their boundaries are not of the same size, though, so we need to take $d - 1$ copies of $T_{d,l-1}$ and one copy of $T_{d,l}$. Then we can add edges connecting the boundary vertices of $T_{d,l}$ to the boundary vertices of the copies of $T_{d,l-1}$ in such a way that the obtained graph $G$ is $d$-regular.

In both cases we have one walk associated to each vertex of $G$: the unique path going to the root inside that copy. For all $v \in V(G)$ and for all original edges $e$ (going inside a copy) we have $H(\mu_v) = H(\mu_e) = H(\bullet)$. As for additional edges $e$ (going between the boundaries of different copies), $\mu_e$ is the joint distribution of $(Y_u, Y_v)$ for vertices $u, v$ at distance $k$. Substituting these into (4) leads to Theorem 3. The calculations are straightforward, we include the odd case $k = 2l + 1$ here. Let $B$ denote the boundary of $T_{d,l}$; then

$$2 \left| E(T_{d,l}) \right| H(\bullet) + (d - 1)|B| H(Y_u, Y_v) \geq 2(d - 1) \left| V(T_{d,l}) \right| H(\bullet).$$

Then for the mutual information $I(Y_u; Y_v) := 2H(\bullet) - H(Y_u, Y_v)$ we have

$$\underbrace{(d - 1)|B|}_{d(d-1)^l} \frac{I(Y_u; Y_v)}{H(\bullet)} \leq 2 \left| E(T_{d,l}) \right| + 2(d - 1)|B| - 2(d - 1) \left| V(T_{d,l}) \right| = 2.$$

## 4. Sharpness, comparisons, applications

4.1. **Sharpness.** All the inequalities stated in this paper for $\mathrm{Aut}(T_d)$-factors (Theorem 3–6) are sharp in the following sense. Given a linear entropy inequality it is natural to normalize it by dividing both sides by the entropy of a vertex. We claim that there exist $\mathrm{Aut}(T_d)$-factor of IID processes for which the two sides of the inequality are arbitrarily close to each other (after normalization). In fact, for each inequality the same examples can be used to demonstrate the sharpness. These examples were already presented in our earlier paper [12] to show that our upper bound for the normalized mutual information is sharp. For the sake of completeness we briefly recall these examples.

The idea is very simple: given IID labels at the vertices, let the factor process "list" all the labels within some large distance $R$ at any given vertex. One needs to be careful since listing the labels should be done in an $\mathrm{Aut}(T_d)$-invariant way. One possibility is to use the following lemma.

**Lemma 4.1.** [12, Lemma 5.2] *For any positive integer $L$ there exists a factor of IID coloring of the vertices of $T_d$ such that finitely many colors are used and vertices of the same color have distance greater than $L$.*

Let us fix $R$ and pick a very large $L$. Let $C = (C_w)_{w \in V(T_d)}$ be a factor of IID coloring provided by the lemma above. Given a positive integer $N$ let $Z_w$, $w \in V(T_d)$ be IID uniform labels from $\{1, 2, \ldots, N\}$. We set

$$Y_v = \{(C_w, Z_w) \mid w \in B_R(v)\}.$$

Then $Y_v$ can be viewed as the list of variables $(C_w, Z_w)$, $w \in B_R(v)$, ordered by $C_w$ (which are all different if $L$ is large enough). This is now an $\mathrm{Aut}(T_d)$-invariant description. Furthermore, conditioned on the coloring process $C$ the entropy corresponding to a finite subset $V \subset V(T_d)$ is $|B_R(V)| \log(N)$ provided that $L$ is large enough. On the other hand, the contribution of the coloring to the entropies does not depend on $N$, so it gets negligible as $N$ goes to infinity. One can easily check that if we replace $H(V)$ by $|B_R(V)|$ in any of our inequalities, then the two sides will be asymptotically equal as $R \to \infty$, and sharpness follows.

### 4.2. Hierarchy of entropy inequalities.
We say that an entropy inequality $A$ is *stronger* than an inequality $B$ ($A \Rightarrow B$ in notation) if the following is true: whenever an $\mathrm{Aut}(T_d)$-invariant process $Y$ (not necessarily factor of IID) satisfies $A$, then $Y$ also satisfies $B$. There is a nested hierarchy between the blow-ups of the edge-vertex and star-edge inequalities:

$$\cdots \Rightarrow \frac{d}{2} H\left(B_{k+1}(\updownarrow)\right) \geq (d-1) H\left(B_{k+1}(\bullet)\right) \Rightarrow H\left(B_k(\text{\Lightning}_d)\right) \geq \frac{d}{2} H\left(B_k(\updownarrow)\right)$$

$$\Rightarrow \frac{d}{2} H\left(B_k(\updownarrow)\right) \geq (d-1) H\left(B_k(\bullet)\right) \Rightarrow \cdots$$

In particular, the star-edge inequality (2) is stronger than the edge-vertex inequality (1), and, in turn, the blow-up (9) (for $k = 1$) of the edge-vertex inequality implies the star-edge inequality.

This can be seen using conditional entropies; we only include a sketch of the argument. For finite sets $U, W \subset V(T_d)$ let $H(W|U)$ denote the conditional entropy $H(Y_w, w \in W \mid Y_u, u \in U)$. We will only use this in the special case when $U \subset W$, where we have $H(W|U) = H(W) - H(U)$.

To see that (2) is stronger than (1): for any invariant process $Y$ satisfying (2) we have

$$\frac{d}{2} H(\updownarrow) \overset{(2)}{\leq} H(\text{\Lightning}_d) = H(\updownarrow) + H(\text{\Lightning}_d \mid \updownarrow) \leq H(\updownarrow) + (d-1) H(\updownarrow \mid \bullet) = d H(\updownarrow) - (d-1) H(\bullet),$$

and (1) follows.

A similar argument shows that for a process satisfying (9) (for $k = 1$) we have

$$\frac{2(d-1)}{d} \underbrace{H(B_1(\bullet))}_{H(\text{\Lightning}_d)} \overset{(9)}{\leq} H(B_1(\updownarrow)) \leq H(\text{\Lightning}_d) + H(\text{\Lightning}_d \mid \updownarrow) = 2 H(\text{\Lightning}_d) - H(\updownarrow),$$

and (2) follows.

Similar arguments were known by Bowen in the dynamical system context, see [7, Proposition 5.1].

4.3. **Tree-indexed Markov chains.** We have already seen that all our new entropy inequalitites are sharp but the question remains: how strong are they compared to previously-known ones? Next we compare them for a specific class of processes.

An intriguing open problem about factor of IID processes is to determine the parameter regime where the *Ising model* on $T_d$ can be obtained as a factor of IID process. More generally, given a Markov chain indexed by $T_d$ with some transition matrix, decide whether the corresponding invariant process is a factor of IID or not. (See [15, 2] and references therein.)

Here we focus on obtaining constraints for a Markov chain to be factor of IID. Two approaches have been used to show that a tree-indexed Markov chain cannot be factor of IID. The correlation bound given in [4] implies that the spectral radius of the transition matrix is at most $1/\sqrt{d-1}$ in the factor of IID case. The edge-vertex entropy inequality yields another constraint. For the Ising model the former gives a slightly better result. There are examples, however, where the latter performs significantly better [2, Theorem 5].

One might think that the entropy approach can be improved by considering the stronger blow-up inequalities described above. However, for Markov chains all these blow-ups are equivalent to the edge-vertex inequality. This is due to the fact that for any connected subset $V \subset V(T_d)$ we have

$$H(V) = H(\bullet) + (|V| - 1) \underbrace{H(\updownarrow \mid \bullet)}_{H(\updownarrow) - H(\bullet)} = (|V| - 1)H(\updownarrow) - (|V| - 2)H(\bullet)$$

because of the Markov property. It follows that all known inequalities involving entropies of connected sets are equivalent to the edge-vertex inequality for tree-indexed Markov chains. In particular, $H(\prec d) \geq (d-1)H(\bullet)$, which follows by combining (1) and (2), is also equivalent to (1) for these processes.

We claim that our new entropy inequalities (7), proved in Theorem 5 for $\mathrm{Aut}(T_d)$-factors of IID, are stronger than (1) for tree-indexed Markov chains.

**Proposition 4.2.** *For tree-indexed Markov chains the inequality $H(S_k) \geq (d-1)^k H(\bullet)$ is stronger than the edge-vertex inequality* (1) *and its blow-ups* (9) *for any given $k$.*

*Proof.* The inequality $H(S_k) \geq (d-1)^k H(\bullet)$ is clearly stronger than $H(B_k) \geq (d-1)^k H(\bullet)$. The latter, however, is equivalent to (1) and (9) for tree-indexed Markov chains. □

Therefore whenever the entropy approach performs better than the correlation bound, using Theorem 5 for any $k \geq 1$ instead of (1) will give an even better result.

As for which $k$ we get the strongest inequality (for Markov chains), we do not have a complete answer. We can prove that for $k = 2$ the theorem is stronger than for $k = 1$, but we do not know if larger $k$ always provides stronger inequality in Theorem 5.

## 5. Proof of the general edge-vertex inequality

To prove the original edge-vertex inequality (1) one needs to count colorings with given "local statistics" on random $d$-regular graphs [2, 16]. In order to obtain Theorem 1 we will generalize this argument for random lifts of a finite base graph $G$.

Let us fix a finite connected simple graph $G$ and a covering map $\varphi \colon T \to G$ for the universal covering tree $T$. By $\Gamma = \Gamma_\varphi$ we denote the group of covering transformations of $T$. We will consider finite lifts $\hat{G}$ of $G$ and colorings of the vertices of $\hat{G}$.

**Definition 5.1.** Let $\hat{G}$ be an $N$-fold lift of $G$. That is, we have a (deterministic) graph $\hat{G}$ and a covering $\hat{G} \to G$ such that every vertex/edge has exactly $N$ *lifts* (i.e. pre-images under the covering map). Suppose that $c \colon V(\hat{G}) \to M$ is a (deterministic) coloring for some finite set $M$ of colors.

By the *local statistics* of the coloring $c$ we mean the following distributions: given a vertex $v$ (or an edge $e$) of $G$, let $\mu_v^c$ (or $\mu_e^c$) be the "empirical distribution" of the colors of the $N$ lifts of $v$ (or $e$). More precisely, for $v \in V(G)$ let $\mu_v^c$ be the distribution of $c(\hat{v})$, where $\hat{v} \in V(\hat{G})$ is chosen uniformly at random among the lifts of $v$. Similarly, for $e = (u, v) \in E(G)$ let $\mu_e^c$ denote the joint distribution of $(c(\hat{u}), c(\hat{v}))$, where $\hat{e} = (\hat{u}, \hat{v}) \in E(\hat{G})$ is chosen uniformly at random among the lifts of $e$.

Note that $\mu_e^c$ is a probability distribution on $M \times M$ with the two marginals being $\mu_u^c$ and $\mu_v^c$. Also, all the probabilities occuring in these distributions are multiples of $1/N$.

From this point on $\varepsilon = \varepsilon(N)$ will denote a positive quantity that slowly converges to $0$ as $N \to \infty$. To be more specific, let $\varepsilon = C/\log N$, where $C$ does not depend on $N$, but it might depend on the base graph $G$, the size of the state space $M$, and the radius $R$ of the factor process. Note that $C$ might be different at each occurence of $\varepsilon$. The proof will have the following ingredients. (Some of the notions used here will be defined later.)

a) It holds with high probability that the random $N$-fold lift of a finite graph $G$ has large *essential girth*, that is, the number of short cycles is small compared to the number of vertices.

b) Given any finite-radius $\Gamma$-factor of IID process $X$ with finite state space $M$ and a finite covering $\hat{G} \to G$ the following holds: there exists a deterministic $M$-coloring $c$ of $\hat{G}$ such that the local statistics $\mu_v^c$ and $\mu_e^c$ are $\varepsilon$-close to $\mu_v^X$ and $\mu_e^X$ provided that the essential girth of $\hat{G}$ is large enough.

c) Finally, we determine the expected number of $M$-colorings with given local statistics on a random $N$-fold lift of $G$.

The general edge-vertex inequality (3) will follow easily by combining the above ingredients.

**a) Random lifts.** Given a finite simple base graph $G$ and a positive integer $N$, a *random $N$-fold lift of $G$*, denoted by $\hat{G}_N$, is the following random graph: for each $v \in V(G)$ we take $N$ vertices $L_v := \{\hat{v}_1, \ldots, \hat{v}_N\}$, and for each $e = (u, v) \in E(G)$ we take a uniform random perfect matching between $L_u$ and $L_v$ (independently for every edge $e$). Figure 5 shows such a random lift for a base graph with four vertices and five edges.

The above definition works for base graphs without loops. In this paper we do not need to use the notion of random lift for base graphs with loops. Let us note nevertheless that random $d$-regular graphs can be considered as random lifts of the graph with one vertex and $d$ *half-loops*.

It is well known that a random $N$-fold lift has few short cycles. More precisely, [11, Lemma 2.1] shows that for any fixed positive integer $l$ the expected number of $l$-cycles in a random $N$-fold lift stays bounded as $N \to \infty$. Using Markov's inequality this immediately implies that with high probability the number of cycles of length at most $l$ is small compared to the number of vertices, which, in turn, implies that the random lift is locally a tree around most vertices. The exact statement we will use is the following.
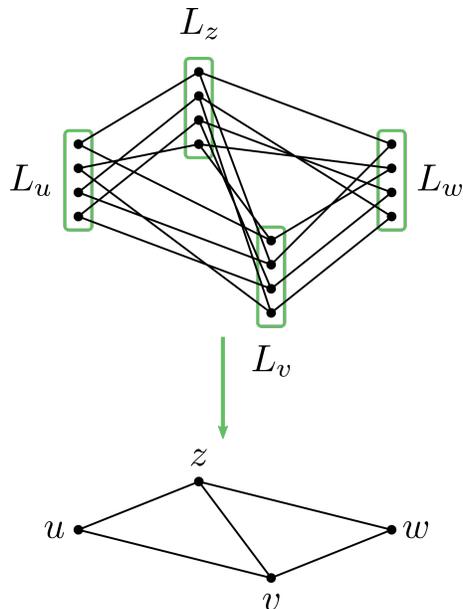
FIGURE 5. The 4-fold random lift of a finite simple graph.

**Lemma 5.2.** *Given any $G$ and any positive integer $R$ the random $N$-fold lift of $G$ has the following property with probability $1 - o(1)$ as $N$ goes to infinity: the $R$-neighborhoods of all but at most $\varepsilon N$ edges are trees.*

**b) Projecting finite-radius factors onto large-girth graphs.** The content of this section can be found in [16, Section 2.1] for the $\mathrm{Aut}(T_d)$-invariant case. The following is a straightforward adaptation for our setting.

Suppose that we have a finite-radius $\Gamma$-factor of IID process with radius $R$ and let $F\colon [0,1]^{V(T)} \to M^{V(T)}$ be the corresponding $\Gamma$-factor mapping. (See Section 2.1 and 2.2 for definitions.) Next we explain how one can "project" such a process onto finite lifts of $G$.

Let $\hat{G}$ be a fixed (deterministic) lift of $G$. We call a vertex/edge of $\hat{G}$ $R$-*nice* if its $R$-neighborhood is a tree. By the *type* of a vertex $\hat{v} \in V(\hat{G})$ we mean its image $v \in V(G)$ under the covering map. Similarly, we can talk about the type of a vertex of the universal cover $T$.

Given an $R$-nice vertex $\hat{v} \in V(\hat{G})$ and an arbitrary vertex $\bar{v} \in V(T)$ with the same type $v \in V(G)$, their $R$-neighborhoods are clearly isomorphic. Moreover, there is a unique isomorphism between these neighborhoods that preserves the vertex types. In what follows we will use this unique isomorphism to identify these neighborhoods.

Now suppose that $[0,1]$ labels are assigned to the vertices of $\hat{G}$. We will refer to these labels as *input labels*. Depending on these input labels we assign a state (i.e. an element from $M$) to each vertex $\hat{v} \in V(\hat{G})$, that is, we define a $[0,1]^{V(\hat{G})} \times V(\hat{G}) \to M$ mapping. We pick an arbitrary fixed state $m_0 \in M$. If $\hat{v}$ is not $R$-nice, we assign $m_0$ to $\hat{v}$. If $\hat{v}$ is $R$-nice, then we can "pretend" that we are at a vertex $\bar{v}$ of the universal cover $T$: we copy the input labels onto the $R$-neighborhood of $\bar{v}$ and apply the function $f_{\bar{v}} := \pi_{\bar{v}} \circ F\colon [0,1]^{V(T)} \to M$; the value of $f_{\bar{v}}$ gets assigned to $\hat{v}$. (Recall that $\pi_{\bar{v}}$ denotes the coordinate projection $M^{V(T)} \to M$ corresponding to the vertex $\bar{v}$.)

For any $\Gamma$-factor process $X$ with finite radius $R$ and for any finite cover $\hat{G}$ of $G$ we described a mapping $[0,1]^{V(\hat{G})} \times V(\hat{G}) \to M$. If we choose the input labels randomly (IID and uniform $[0,1]$), then we get a random function $c \colon V(\hat{G}) \to M$. We will think of $c$ as a random $M$-coloring of the vertices of $\hat{G}$ that depends deterministically on the IID input labels. It is easy to see that this random coloring has the following properties.

- The distribution of the random color of an $R$-nice vertex of type $v$ is $\mu_v^X$. Similarly, for an $R$-nice edge $\hat{e}$ the joint distribution of the colors on the endpoints of $\hat{e}$ is $\mu_e^X$ for the corresponding $e \in E(G)$. (See Theorem 1 for the definition of $\mu_v^X$ and $\mu_e^X$.)
- The color of a vertex depends only on the input labels in its $R$-neighborhood. That is, if we change the labels outside its $R$-neighborhood, its color remains the same.

From now on we will assume that all but at most $\varepsilon N$ edges of $\hat{G}$ are $R$-nice. Definition 5.1 defines the local statistics $\mu_v^c$ and $\mu_e^c$ of a deterministic coloring $c \colon V(\hat{G}) \to M$. Here we have a random coloring $c$, therefore $\mu_v^c$ and $\mu_e^c$ are random measures depending on the input labels. Taking expectation (with respect to the input labels) we get the measures $\mathbb{E}\mu_v^c$ and $\mathbb{E}\mu_e^c$. We claim that $\mathbb{E}\mu_e^c$ is $\varepsilon$-close to $\mu_e^X$ in total variation distance for each $e \in E(G)$. This follows from the fact that the color pair of an $R$-nice lift of $e$ has distribution $\mu_e^X$ and that at most $\varepsilon N$ edges are not $R$-nice among the $N$ lifts of $e$.

Our goal is to show the existence of a deterministic coloring $c \colon V(\hat{G}) \to M$ with the property that $\mu_e^c$ is $\varepsilon$-close to $\mu_e^X$ for each $e \in E(G)$. At this point we have a random coloring for which this is true in expectation. We will use the following form of the Azuma–Hoeffding inequality to show that the local statistics of our random coloring are concentrated around their expectations.

**Lemma 5.3.** *Let $(\Omega^n, \nu^n)$ be a product probability space. For a Lipschitz continuous function $f \colon \Omega^n \to \mathbb{R}$ with Lipschitz constant $K$ (w.r.t. the Hamming distance on $\Omega^n$) we have*

$$(10) \qquad \nu^n \left( \{ \omega \in \Omega^n \,:\, |f(\omega) - \mathbb{E}f| > \lambda \} \right) \leq 2 \exp \left( \frac{-\lambda^2}{2K^2 n} \right).$$

We use this in the following setting: $\Omega = [0,1]$, $\nu$ is the uniform measure on $[0,1]$, and $n = |V(\hat{G})| = N|V(G)|$. We will apply (10) to different functions $f$. Next we describe these functions.

Our random coloring $c$ depends on the configuration $\omega \in \Omega^n \cong [0,1]^{V(\hat{G})}$ of the input labels. For a given edge $e = (v_1, v_2) \in E(G)$ and a given pair of colors $m_1, m_2 \in M$ let $f(\omega) := N\mu_e^c(\{(m_1, m_2)\})$, that is, $f$ is the number of lifts of $e = (v_1, v_2)$ with the first endpoint having color $m_1$ and the second endpoint having color $m_2$. Using the fact that the random color of a vertex depends only on the input labels in its $R$-neighborhood, it is easy to see that $f$ is Lipschitz continuous with $K = 2d_{\max}^{R+1}$, where $d_{\max}$ is the maximum degree of the base graph $G$.

Using (10) with $\lambda = \varepsilon N$ we get that the probability that $\mu_e^c(\{(m_1, m_2)\})$ is not $\varepsilon$-close to $\mathbb{E}\mu_e^c(\{(m_1, m_2)\})$ is very small: at most $2\exp(-\varepsilon^2 N)$. Recall that $\varepsilon$ can denote any quantity $C/\log N$ where $C$ might depend on $G, M, R$ but not on $N$.

Using union bound for all $e$ and all pairs $(m_1, m_2)$ we get that for large enough $N$ it holds with positive probability that $\mu_e^c$ is $\varepsilon$-close to $\mathbb{E}\mu_e^c$ for each $e \in E(G)$. We have already seen that $\mathbb{E}\mu_e^c$ is $\varepsilon$-close to $\mu_e^X$, thus we have proved the following.

**Lemma 5.4.** *Suppose that all but at most $\varepsilon N$ edges of $\hat{G}$ are $R$-nice and that $N$ is large enough. Then there exists a deterministic coloring $c \colon V(\hat{G}) \to M$ such that $\mu_e^c$ is $\varepsilon$-close (say in total variation distance) to $\mu_e^X$ for each edge $e \in E(G)$.*

**c) The expected number of good colorings.** Next we determine the expected number of colorings with prescribed local statistics on random lifts of a base graph. These local statistics need to be consistent in the following sense.

**Definition 5.5.** For a finite simple graph $G$ and a finite color set $M$ by a *consistent collection of distributions* we mean the following: a probability distribution $\mu_v$ on $M$ for each $v \in V(G)$ and a probability distribution $\mu_e$ on $M \times M$ for each $e \in E(G)$ such that the marginals of $\mu_e$ for $e = (u, v)$ are $\mu_u$ and $\mu_v$.

**Lemma 5.6.** *Let $\mu_v$, $v \in V(G)$, and $\mu_e$, $e \in E(G)$, be a consistent collection of distributions as in the definition above. Recall that $\hat{G}_N$ denotes the random $N$-fold lift of $G$. Then the following formula holds for the expectation (w.r.t. $\hat{G}_N$) of the number of colorings on $\hat{G}_N$ for which the edge-statistics coincide with $\mu_e$:*

$$(11) \quad \mathbb{E}_{\hat{G}_N} \left| \left\{ c \colon V(\hat{G}_N) \to M : \mu_e^c = \mu_e \; \forall e \in E(G) \right\} \right|$$

$$= \exp\left( N \left( \sum_{e \in E(G)} H(\mu_e) - \sum_{v \in V(G)} (\deg v - 1) H(\mu_v) + o(1) \right) \right) \quad as \; N \to \infty$$

*provided that the probabilities occuring in the discrete distributions $\mu_e$ are rational numbers and $N$ is a common multiple of all the denominators (otherwise the number of such colorings is clearly 0).*

To prove the above lemma we will adapt the arguments in [2, Section 4] for our more general setting.

Given a discrete distribution $\mu$ on $M$ (set of colors) the multinomial coefficients describe the number of $M$-colorings of a finite set with color distribution $\mu$. Using the Stirling formula it is easy to derive an asymptotic formula as the number of elements $N$ goes to infinity: there are

$$\exp\left( N \left( H(\mu) + o(1) \right) \right)$$

ways to choose the colors of $N$ elements in a way that the number of elements with color $m \in M$ is $N\mu(\{m\})$ (provided that these numbers are integers).

We will also need the following statement which is a slight variant of [2, Lemma 4.1].

**Claim.** *Let $L_u$ and $L_v$ be disjoint sets of size $N$. Fix $M$-colorings of $L_u$ and $L_v$ with color distributions $\mu_u$ and $\mu_v$, respectively. Let $\mu_e$ be any distribution on $M \times M$ with marginals $\mu_u$ and $\mu_v$ and with the property that all probabilities occuring in $\mu_e$ are multiples of $1/N$. Then the probability that a uniform random perfect matching between $L_u$ and $L_v$ has color distribution $\mu_e$ is*

$$(12) \qquad \exp\left( N \left( H(\mu_e) - H(\mu_u) - H(\mu_v) + o(1) \right) \right).$$

*(The color distribution of a matching is the distribution of the pair of colors on the endpoints of the edges.)*

Before proving this claim we show how Lemma 5.6 follows. First we take disjoint sets $L_v$ of size $N$ for each $v \in V(G)$. Then we color each $L_v$ with statistics $\mu_v$. This can be done in

$$(13) \qquad \exp\left( N \left( \sum_{v \in V(G)} H(\mu_v) + o(1) \right) \right)$$

different ways. Let us fix such a coloring $c\colon \cup_{v\in V(G)} L_v \to M$. To get a random lift of $G$ we need to choose a uniform random perfect matching between $L_u$ and $L_v$ independently for each edge $e = (u,v)$. The probability that this perfect matching has statistics $\mu_e$ (for any fixed coloring $c$) is given by the formula (12). These probabilities are independent and consequently the probability that a fixed coloring $c$ is "good" for a random lift is the product of (12) with $e$ running through $E(G)$. To get the expected number of good colorings for a random lift we need to multiply this product by (13), and Lemma 5.6 follows.

Finally we prove the claim.

*Proof of Claim.* By a *colored perfect matching* between $L_u$ and $L_v$ we mean a coloring of the vertices in $L_u \cup L_v$ and a perfect matching between $L_u$ and $L_v$. There are two different ways to count the number of colored perfect matchings with color distribution $\mu_e$:

$$(\#\text{all perfect matchings}) \cdot \exp\left(N\left(H(\mu_e) + o(1)\right)\right)$$
$$= \underbrace{(\#\text{vertex colorings})}_{\exp(N(H(\mu_u)+H(\mu_v)+o(1)))} \cdot (\#\text{good perfect matchings for any given vertex-coloring}).$$

The claim immediately follows from this equality.  □

**Putting the ingredients together.** As we explained in Section 2.2, an arbitrary $\Gamma$-factor of IID process $X$ is the weak limit of finite-radius factors. Since the entropies $H(\mu_v^X)$ and $H(\mu_e^X)$ are continuous under weak convergence, it suffices to prove Theorem 1 for finite-radius factors. So let us assume that $X$ is a $\Gamma$-factor of IID process with some finite radius $R$.

On a random $N$-fold lift of $G$ let us consider the colorings $c$ with the property that $\mu_e^c$ is $\varepsilon$-close to $\mu_e^X$ for all $e \in E(G)$. We claim that the expected number of such colorings on a random lift is, on the one hand, at least $1 - o(1)$, and, on the other hand, asymptotically equal to

$$(14) \qquad \exp\left(N\left(\sum_{e\in E(G)} H(\mu_e^X) - \sum_{v\in V(G)} (\deg v - 1)H(\mu_v^X) + o(1)\right)\right) \quad \text{as } N \to \infty.$$

Combining Lemma 5.2 and Lemma 5.4 implies that at least one such coloring exists for a random $N$-fold lift of $G$ with probability $1 - o(1)$. Therefore the expected number of such colorings is indeed at least $1 - o(1)$.

To get (14) we need to apply Lemma 5.6 for all collections of distributions $\mu_v$ and $\mu_e$ with the property that they are $\varepsilon$-close to $\mu_v^X$ and $\mu_e^X$, respectively, and that all the probabilities occuring are multiples of $1/N$. It is easy to see that the total number of such collections is polynomial in $N$. We need to take the sum of (11) for all these collections. We can replace the entropies $H(\mu_v)$ and $H(\mu_e)$ with $H(\mu_v^X)$ and $H(\mu_e^X)$ at the expense of an $o(1)$ difference as $N \to \infty$. We get (14) with an extra factor that is polynomial in N but that can be also incorporated in the $N \cdot o(1)$ term in the exponent.

Therefore (14) is at least $1 - o(1)$ as $N \to \infty$ meaning that the term

$$\sum_{e\in E(G)} H(\mu_e^X) - \sum_{v\in V(G)} (\deg v - 1)H(\mu_v^X)$$

in the exponent cannot be negative, and this is exactly what we wanted to prove.

## References

[1] Ágnes Backhausz, Balázs Gerencsér, Viktor Harangi, and Máté Vizer. Correlation bound for distant parts of factor of iid processes. *Combin. Probab. Comput.*, to appear.

[2] Ágnes Backhausz and Balázs Szegedy. On large girth regular graphs and random processes on trees. arXiv:1406.4420, 2014.

[3] Ágnes Backhausz and Balázs Szegedy. On the almost eigenvectors of random regular graphs. arXiv:1607.04785, 2016.

[4] Ágnes Backhausz, Balázs Szegedy, and Bálint Virág. Ramanujan graphings and correlation decay in local algorithms. *Random Structures Algorithms*, 47(3):424–435, 2015.

[5] Karen Ball. Factors of independent and identically distributed processes with non-amenable group actions. *Ergodic Theory Dyn. Syst.*, 25(3):711–730, 2005.

[6] B. Bollobás. The independence ratio of regular graphs. *Proc. Amer. Math. Soc.*, 83(2):433–436, 1981.

[7] Lewis Bowen. A measure-conjugacy invariant for free group actions. *Ann. Math. (2)*, 171(2):1387–1400, 2010.

[8] Lewis Bowen. The ergodic theory of free group actions: entropy and the $f$-invariant. *Groups Geom. Dyn.*, 4(3):419–432, 2010.

[9] Endre Csóka. Independent sets and cuts in large-girth regular graphs. arXiv:1602.02747, 2016.

[10] Endre Csóka, Balázs Gerencsér, Viktor Harangi, and Bálint Virág. Invariant Gaussian processes and independent sets on regular graphs of large girth. *Random Structures Algorithms*, 47(2):284–303, 2015.

[11] Jean-Philippe Fortin and Samantha Rudinsky. Asymptotic eigenvalue distribution of random lifts. *The Waterloo Mathematics Review*, pages 20–28, 2013.

[12] Balázs Gerencsér and Viktor Harangi. Mutual information decay for factors of iid. arXiv:1703.04387, 2017.

[13] Viktor Harangi and Bálint Virág. Independence ratio and random eigenvectors in transitive graphs. *Ann. Probab.*, 43(5):2810–2840, 2015.

[14] Carlos Hoppen and Nicholas Wormald. Local algorithms, regular graphs of large girth, and random regular graphs. *To appear in Combinatorica.* arXiv:1308.0266.

[15] Russell Lyons. Factors of IID on trees. *Combin. Probab. Comput.*, 26(2):285–300, 2017.

[16] Mustazee Rahman. Factor of IID percolation on trees. *SIAM J. Discrete Math.*, 30(4):2217–2242, 2016.

[17] Mustazee Rahman and Bálint Virág. Local algorithms for independent sets are half-optimal. *Annals of Prob.*, to appear.

ELTE Eötvös Loránd University, Budapest, Hungary; Department of Probability Theory and Statistics; H-1117 Budapest, Pázmány Péter sétány 1/c; and MTA Alfréd Rényi Institute of Mathematics H-1053 Budapest, Reáltanoda utca 13-15

*E-mail address*: `agnes@math.elte.hu`

MTA Alfréd Rényi Institute of Mathematics H-1053 Budapest, Reáltanoda utca 13-15; and ELTE Eötvös Loránd University, Budapest, Hungary; Department of Probability Theory and Statistics H-1117 Budapest, Pázmány Péter sétány 1/c

*E-mail address*: `gerencser.balazs@renyi.mta.hu`

MTA Alfréd Rényi Institute of Mathematics H-1053 Budapest, Reáltanoda utca 13-15

*E-mail address*: `harangi@renyi.hu`