Ergod. Th. & Dynam. Sys., page 1 of 38

doi:10.1017/etds.2022.27

The relative f-invariant and non-uniform random sofic approximations

CHRISTOPHER SHRIVER

Department of Mathematics, University of California, Los Angeles, Los Angeles 90095, CA, USA
(e-mail: christopher.shriver@math.utexas.edu)

(*Received* 29 *May* 2021 *and accepted in revised form* 1 *March* 2022)

Abstract. The f-invariant is an isomorphism invariant of free-group measure-preserving actions introduced by Lewis Bowen, who first used it to show that two finite-entropy Bernoulli shifts over a finitely generated free group can be isomorphic only if their base measures have the same Shannon entropy. Bowen also showed that the f-invariant is a variant of sofic entropy; in particular, it is the exponential growth rate of the expected number of good models over a uniform random homomorphism. In this paper we present an analogous formula for the relative f-invariant and use it to prove a formula for the exponential growth rate of the expected number of good models over a random sofic approximation which is a type of stochastic block model.

Key words: sofic entropy, random regular graphs, stochastic block model
2020 Mathematics Subject Classification: 37A35 (Primary); 28D20 (Secondary)

1. *Introduction and main results*

Let $G = \langle S \rangle$ denote the rank-r free group with generating set $S = \{s_1, \ldots, s_r\}$ and identity e, and let (X, μ, T) be a measure-preserving G-system, that is, T is a homomorphism from G to the automorphism group of the standard probability space (X, μ). We will not need to make explicit use of the σ-algebra on X, so we leave it unnamed.

An *observable* on X is a measurable map with domain X. In this paper the codomain will be a finite set endowed with the discrete sigma algebra; in this case we call the map a *finite observable* and the codomain an *alphabet*.

Any observable $\alpha : X \to \mathtt{A}$ induces a map $\alpha^G : X \to \mathtt{A}^G$ by setting

$$(\alpha^G(x))_g = \alpha(T_g x) \quad \text{for all } g \in G.$$

We call the $\mathtt{A}$-coloring $\alpha^G(x)$ of G the *itinerary* of x, since it records the observations that will be made over the entire orbit of x under the action of G. We also similarly define

the map $\alpha^H : X \to \mathtt{A}^H$ for any subset H of G. We abbreviate $\alpha^n := \alpha^{\mathrm{B}(e,n)}$, where $\mathrm{B}(e, n)$ is the closed ball of radius n centered at the identity in G, which is endowed with the word-length metric. If $\beta : X \to \mathtt{B}$ is a second finite observable, we denote by $\alpha\beta : X \to \mathtt{A} \times \mathtt{B}$ the map $\alpha\beta(x) = (\alpha(x), \beta(x))$.

The (Shannon) entropy of a finite observable $\alpha : X \to \mathtt{A}$ is defined by

$$\mathrm{H}_\mu(\alpha) = -\sum_{a \in \mathtt{A}} \alpha_* \mu(a) \log \alpha_* \mu(a),$$

where $\alpha_*\mu \in \mathrm{Prob}(\mathtt{A})$ is the pushforward measure, with the convention $0 \log 0 = 0$. The entropy of α can be interpreted as the expected amount of information revealed by observing α, assuming its distribution $\alpha_*\mu$ is known.

An early application of Shannon's entropy to ergodic theory was its use by Kolmogorov and Sinai to show that there exist non-isomorphic Bernoulli shifts over $\mathbb{Z}$. A Bernoulli shift over $\mathbb{Z}$ is a system of the form $(\mathtt{A}^{\mathbb{Z}}, \mu^{\mathbb{Z}}, S)$ for some alphabet $\mathtt{A}$ and $\mu \in \mathrm{Prob}(\mathtt{A})$; S is the shift action of $\mathbb{Z}$. They did this by defining an *entropy rate* for $\mathbb{Z}$-systems, which can be interpreted as the average information per unit time revealed by observing the system. For a Bernoulli shift $(\mathtt{A}^{\mathbb{Z}}, \mu^{\mathbb{Z}}, S)$, the entropy rate is simply the 'base entropy' $\mathrm{H}_\mu(\alpha)$, where $\alpha : \mathtt{A}^n \to \mathtt{A}$ is the 'time zero' observable.

Isomorphism invariance of the Kolmogorov–Sinai entropy rate is typically proven using the fact that entropy rate is non-increasing under factor maps (which are surjective homomorphisms of measure-preserving systems). This fact can be interpreted as stating that a system cannot simulate another system that is 'more random'.

The entropy rate was soon generalized to systems acted on by an arbitrary amenable group (such as $\mathbb{Z}^d$). Extending beyond amenable groups proved more difficult, and in fact it was found to be impossible for such an extension to preserve all desirable properties of the Kolmogorov–Sinai entropy rate. In particular, an entropy rate for non-amenable group actions which assigns Bernoulli shifts their base entropy cannot be non-increasing under factor maps [13, Appendix C].

The first invariant to distinguish between Bernoulli shifts over free groups is Lewis Bowen's f-invariant. Following [2], this can be defined by

$$F_\mu(T, \alpha) = (1 - 2r)\mathrm{H}_\mu(\alpha) + \sum_{i=1}^{r} \mathrm{H}_\mu(\alpha^{\{e, s_i\}}),$$

$$f_\mu(T, \alpha) = \inf_n F_\mu(T, \alpha^n) = \lim_{n \to \infty} F_\mu(T, \alpha^n).$$

The main theorem of [3] is that $f_\mu(T, \alpha)$ depends on the observable α only through the σ-algebra it generates. In particular, the common value of $f_\mu(T, \alpha)$ among all α which generate the σ-algebra of the measurable space X (assuming such α exist) is a measure-conjugacy invariant of the system (X, μ, T). In the same paper, Bowen showed that the f-invariant of a Bernoulli shift is the Shannon entropy of the base measure; in particular, Bernoulli shifts with different base entropies are non-isomorphic.

In [2], Bowen gave an alternate formula for the f-invariant, which we now introduce.

For any homomorphism $\sigma : G \to \mathrm{Sym}(n)$ we have a G-system $([n], \mathrm{Unif}(n), \sigma)$, and we can consider a labeling $\mathbf{x} \in \mathtt{A}^n$ as an $\mathtt{A}$-valued observable on this system. We denote

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

the law of its itinerary by $P_{\mathbf{x}}^{\sigma} = \mathbf{x}_*^G \operatorname{Unif}(n)$ and call this the *empirical distribution* of $\mathbf{x}$. We say that $\mathbf{x}$ is a good model for α over σ if it is difficult to distinguish the G-systems (X, μ, T) and $([n], \operatorname{Unif}(n), \sigma)$ via their respective observables α and $\mathbf{x}$. To make this precise, we denote

$$\Omega(\sigma, \mathcal{O}) := \{\mathbf{x} \in \mathtt{A}^n : P_{\mathbf{x}}^{\sigma} \in \mathcal{O}\},$$

which is a set of good models for α over σ if $\mathcal{O}$ is a weak*-open neighborhood of $\alpha_*^G \mu \in \operatorname{Prob}(\mathtt{A}^G)$; the particular set $\mathcal{O}$ quantifies how good the models are. The alphabet $\mathtt{A}$ is given the discrete topology and $\mathtt{A}^G$ the product topology, so 'weak*-close' means marginals on some finite sets are close in total variation norm.

For each $n \in \mathbb{N}$, let $\mathtt{s}_n = \operatorname{Unif}(\operatorname{Hom}(G, \operatorname{Sym}(n)))$. Bowen showed in [2] that the f-invariant is given by

$$f_\mu(T, \alpha) = \inf_{\mathcal{O} \ni \alpha_*^G \mu} \limsup_{n \to \infty} \frac{1}{n} \log \mathop{\mathbb{E}}_{\sigma \sim \mathtt{s}_n} |\Omega(\sigma, \mathcal{O})|.$$

To make an analogy with statistical physics, we can think of $\alpha_*^G \mu$ as a macroscopic statistical distribution of the state of a system; then the f-invariant is the exponential growth rate of the expected number of 'microstates' that are consistent with these statistics. What we here call good models are often called microstates for this reason.

More generally, given any random or deterministic sofic approximation $\Sigma = \{\mathtt{s}_n\}_{n=1}^{\infty}$, we can define the sofic entropy relative to Σ by

$$\mathrm{h}_{\Sigma,\mu}(T, \alpha) = \inf_{\mathcal{O} \ni \alpha_*^G \mu} \limsup_{n \to \infty} \frac{1}{n} \log \mathop{\mathbb{E}}_{\sigma \sim \mathtt{s}_n} |\Omega(\sigma, \mathcal{O})|.$$

Here each $\mathtt{s}_n$ is a probability measure on the set of functions $G \to \operatorname{Sym}(n)$ which is supported on functions which are approximately free homomorphisms.

This paper is motivated by a desire to better understand the dependence of sofic entropy on the sofic approximation Σ. For any choice of Σ, the sofic entropy agrees with Kolmogorov–Sinai entropy if the acting group is amenable [6] and with the Shannon entropy of the base if the system is a Bernoulli shift [4]. For some systems, the sofic entropy can be finite relative to some sofic approximations and $-\infty$ relative to others. It is unknown whether two deterministic sofic approximations can yield different finite entropy values for the same system.

In this paper, we express the entropy relative to a type of stochastic block model in terms of the relative f-invariant, which we now introduce.

If α, β are two finite observables with codomains $\mathtt{A}$, $\mathtt{B}$, the conditional entropy is

$$\mathrm{H}_\mu(\alpha|\beta) = \mathrm{H}_\mu(\alpha\beta) - \mathrm{H}_\mu(\beta).$$

This can be interpreted as the expected amount of information revealed by observing α if both the value of β and the joint distribution of α and β are known. The relative f-invariant

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

is defined by

$$\begin{aligned} F_\mu(T, \alpha|\beta) &= F_\mu(T, \alpha\beta) - F_\mu(T, \beta) \\ &= (1-2r)\mathrm{H}_\mu(\alpha|\beta) + \sum_{i=1}^{r} \mathrm{H}_\mu(\alpha^{\{e,s_i\}} \mid \beta^{\{e,s_i\}}), \\ f_\mu(T, \alpha|\beta) &= \inf_{k_1\in\mathbb{N}} \sup_{k_2\in\mathbb{N}} F_\mu(T, \alpha^{k_1} \mid \beta^{k_2}). \end{aligned}$$

Both the infimum and supremum can be replaced by limits; this follows from Lemma 3.2 below. It follows from Corollary 3.5 that we could also directly define

$$f_\mu(T, \alpha|\beta) = f_\mu(T, \alpha\beta) - f_\mu(T, \beta),$$

as long as $f_\mu(T, \beta) > -\infty$.

We now define the relevant type of stochastic block model. If H is a finite subset of G, we denote by $d^H(\mu, \nu)$ the total variation distance between the marginals of μ and ν on $\mathtt{A}^H$. Our convention for the total variation distance between measures $\mu, \nu \in \mathrm{Prob}(\mathtt{A})$ is

$$\|\mu - \nu\|_{\mathrm{TV}} = \frac{1}{2}\sum_{a\in\mathtt{A}} |\mu\{a\} - \nu\{a\}|.$$

For each $k \in \mathbb{N}$ we define a pseudometric on $\mathrm{Prob}(\mathtt{A}^G)$ by

$$d_k^*(\mu, \nu) = \sum_{i\in[r]} d^{\mathrm{B}(e,k)\cup\mathrm{B}(s_i,k)}(\mu, \nu).$$

Note that $\{d_k^*\}_{k\in\mathbb{N}}$ together generate the weak* topology on $\mathrm{Prob}(\mathtt{A}^G)$. These generalize the function d_σ^* from [2], which corresponds to the case $k = 0$. For $\mathcal{O} = \{\nu \in \mathrm{Prob}(\mathtt{A}^G) : d_k^*(\alpha_*^G\mu, \nu) < \varepsilon\}$ we write

$$\Omega(\sigma, \mathcal{O}) =: \Omega_k^*(\sigma, \alpha, \varepsilon) \subseteq \mathtt{A}^n.$$

Our stochastic block model is now defined as follows: given $\mathbf{y}_0 \in \mathtt{B}^n$, $\sigma_0 \in \mathrm{Hom}(G, \mathrm{Sym}(n))$, and $k \in \mathbb{N}$, let

$$\mathtt{SBM}(\sigma_0, \mathbf{y}_0, k) := \mathrm{Unif}(\{\sigma \in \mathrm{Hom}(G, \mathrm{Sym}(n)) : d_k^*(P_{\mathbf{y}_0}^\sigma, P_{\mathbf{y}_0}^{\sigma_0}) = 0\}).$$

The labeling $\mathbf{y}_0$ partitions the elements of $[n]$ into $|\mathtt{B}|$ communities, and we can think of the random homomorphism σ as a random choice of directed edges between and within the communities. Certain statistics of these random edge choices are determined by the reference homomorphism σ_0; note that for $k > 0$ these statistics are more precise than those specified by a standard stochastic block model. In §2 we define weights, which are the objects used to record the relevant statistics.

1.1. *Main results.* Our main theorems show that the relative f-invariant can be interpreted as the growth rate of the expected number of ways to extend a planted good model for β to a good model for $\alpha\beta$, over a stochastic block model which has statistics determined by β and its planted model.

We first prove that if $\beta_*^G\mu$ is Markov then we can use a stochastic block model which only takes into account 'one-step statistics'.

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

THEOREM A. *Let $\alpha : X \to \mathtt{A}$ and $\beta : X \to \mathtt{B}$ be finite observables, and for each n let $\mathbf{y}_n \in \mathtt{B}^n$ and $\sigma_n \in \mathrm{Hom}(G, \mathrm{Sym}(n))$ be such that*

$$\lim_{n\to\infty} d_0^*(P_{\mathbf{y}_n}^{\sigma_n}, \beta_*^G\mu) = 0.$$

Suppose that $\beta_*^G\mu$ is a Markov measure. With $\mathtt{s}_n = \mathtt{SBM}(\sigma_n, \mathbf{y}_n, 0)$, we have

$$f_\mu(T, \alpha \mid \beta) = \inf_{O \ni (\alpha\beta)_*^G\mu} \limsup_{n\to\infty} \frac{1}{n} \log \mathop{\mathbb{E}}_{\sigma\sim\mathtt{s}_n} |\{\mathbf{x} \in \mathtt{A}^n : (\mathbf{x}, \mathbf{y}_n) \in \Omega(\sigma, O)\}|.$$

PROPOSITION A. *The assumptions of Theorem A are non-vacuous; that is, for any finite observable $\beta : X \to \mathtt{B}$ there exist sequences $\{\mathbf{y}_n \in \mathtt{B}^n\}_{n=1}^\infty$ and $\{\sigma_n \in \mathrm{Hom}(G, \mathrm{Sym}(n))\}_{n=1}^\infty$ such that $\lim_{n\to\infty} d_0^*(P_{\mathbf{y}_n}^{\sigma_n}, \beta_*^G\mu) = 0$.*

This follows from the fact that free group actions are 'sofic', which is proven for example in [10, 14, 15]. A more elementary proof is given in §4 below.

If $\beta_*^G\mu$ is not Markov, then the same formula holds with a more precise type of stochastic block model.

THEOREM B. *Let $\alpha : X \to \mathtt{A}$ and $\beta : X \to \mathtt{B}$ be finite observables. Let m_n approach infinity as n goes to infinity while satisfying $m_n = o(\log\log n)$. For each n, let $\mathbf{y}_n \in \mathtt{B}^n$ and $\sigma_n \in \mathrm{Hom}(G, \mathrm{Sym}(n))$ be such that*

$$d_{m_n}^*(P_{\mathbf{y}_n}^{\sigma_n}, \beta_*^G\mu) = O\left(\frac{1}{\log n}\right).$$

Suppose that $f_\mu(T, \beta) > -\infty$. With $\mathtt{s}_n = \mathtt{SBM}(\sigma_n, \mathbf{y}_n, m_n)$,

$$f_\mu(T, \alpha \mid \beta) = \inf_{O \ni (\alpha\beta)_*^G\mu} \limsup_{n\to\infty} \frac{1}{n} \log \mathop{\mathbb{E}}_{\sigma\sim\mathtt{s}_n} |\{\mathbf{x} \in \mathtt{A}^n : (\mathbf{x}, \mathbf{y}_n) \in \Omega(\sigma, O)\}|.$$

PROPOSITION B. *The assumptions of Theorem B are non-vacuous; that is, for any finite observable $\beta : X \to \mathtt{B}$ and any sequence $\{m_n \in \mathbb{N}\}_{n=1}^\infty$ approaching infinity while satisfying $m_n = o(\log\log n)$, there exist sequences $\{\mathbf{y}_n \in \mathtt{B}^n\}_{n=1}^\infty$ and $\{\sigma_n \in \mathrm{Hom}(G, \mathrm{Sym}(n))\}_{n=1}^\infty$ such that $\lim_{n\to\infty} d_{m_n}^*(P_{\mathbf{y}_n}^{\sigma_n}, \beta_*^G\mu) = O(1/\log n)$.*

Using Theorem B, we prove the following formula for the growth rate of the expected number of good models over a stochastic block model. This can be compared to the variational principle in [12], and has a similar proof.

THEOREM C. *Let $\mathtt{s}_n, \alpha, \beta$ be as in the statement of Theorem B. Then*

$$\inf_{O \ni \alpha_*^G\mu} \limsup_{n\to\infty} \frac{1}{n} \log \mathop{\mathbb{E}}_{\sigma\sim\mathtt{s}_n} |\Omega(\sigma, O)| = \sup_{\lambda \in \mathrm{J}(\alpha_*^G\mu,\, \beta_*^G\mu)} f_\lambda(S, \mathtt{a} \mid \mathtt{b}).$$

Here $\mathrm{J}(\alpha_*^G\mu,\ \beta_*^G\mu)$ is the set of joinings of the G-systems $(\mathtt{A}^G, \alpha_*^G\mu, S)$ and $(\mathtt{B}^G, \beta_*^G\mu, S)$, that is, shift-invariant probability measures on $(\mathtt{A} \times \mathtt{B})^G$ whose $\mathtt{A}^G, \mathtt{B}^G$ marginals are $\alpha_*^G\mu,\ \beta_*^G\mu$, respectively. S denotes the shift action of G. We use $\mathtt{a}, \mathtt{b}$ to denote the maps

$$\mathtt{a} : (\mathtt{A} \times \mathtt{B})^G \to \mathtt{A}$$
$$((a_g, b_g))_{g\in G} \mapsto a_e$$

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

and

$$\mathtt{b}\colon (\mathtt{A}\times\mathtt{B})^G \to \mathtt{B}$$
$$((a_g, b_g))_{g\in G} \mapsto b_e,$$

which observe the A (respectively, B) label at the identity.

Note that the supremum is always greater than or equal to $f_\mu(T,\alpha)$, with equality attained by the product joining; this means that the expected number of good models for α over a block model with built-in good models for any β is at least the expected number of good models over a uniformly random homomorphism. It is possible for the supremum to be strictly larger, however. For example, suppose $f_\mu(T,\alpha) < 0$ and $\alpha = \beta$, and let λ be the diagonal joining. Then

$$f_\lambda(S, \mathtt{a} \mid \mathtt{b}) = 0 > f_\mu(T,\alpha).$$

1.2. *Related work.* The expressions appearing on the right-hand sides of Theorems A and B are very closely related to Ben Hayes' definition of 'relative sofic entropy in the presence' [11, Definition 2.5]. Some differences are that we consider *expected* numbers of good models over *random* sofic approximations, and that Hayes takes a supremum inside the logarithm over which good model is to be extended, while we fix a sequence $\{\mathbf{y}_n\}$ of planted good models. Hayes also does not restrict to shift systems as we do here.

In [8], the free energy (that is, the limit of normalized log partition functions) over a type of stochastic block model is shown to satisfy a variational principle; see Propositions 3.6 and 3.7 of that paper.

1.3. *Random sofic approximations.* As noted above, the f-invariant is closely related to another invariant of measure-preserving systems called sofic entropy, which was introduced by Bowen in [4].

A homomorphism $\sigma \in \mathrm{Hom}(G, \mathrm{Sym}(n))$ is called (D,δ)-sofic for some finite $D \subset G$ and $\delta > 0$ if

$$|\{j \in [n] : \sigma(\gamma)j \neq j \text{ for all } \gamma \in D \setminus \{e\}\}| > (1-\delta)n.$$

A sequence of homomorphisms $\Sigma = (\sigma_n \in \mathrm{Hom}(G, \mathrm{Sym}(n)))_{n\in\mathbb{N}}$ is called a sofic approximation if, for every (D,δ), the homomorphism σ_n is (D,δ)-sofic for all large enough n.

The sofic entropy relative to Σ is the exponential growth rate of the number of good models over σ_n. Specifically, for any finite observable α on X we have

$$\mathrm{h}_{\Sigma,\mu}(T,\alpha) = \inf_{\mathcal{O}\ni\alpha^G_*\mu} \limsup_{n\to\infty} \frac{1}{n}\log|\Omega(\sigma_n,\mathcal{O})|.$$

This is an isomorphism invariant of the system (X,μ,T) if α is any *generating* observable, that is if the σ-algebra of the measurable space X is the coarsest one which is shift-invariant and α-measurable.

By analogy with this expression, we might call the sequences of random homomorphisms appearing in expressions above 'random sofic approximations'. The following proposition provides further justification for this terminology.

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

PROPOSITION 1.1. *If* $(\mathtt{s}_n)$ *is any of the sequences appearing in Theorems A, B, and C, then for any* (D, δ) *there exists* $\varepsilon > 0$ *such that*

$$\mathop{\mathbb{P}}_{\sigma\sim\mathtt{s}_n}(\sigma \text{ is } (D,\delta)\text{-sofic}) \geq 1 - n^{-\varepsilon n}$$

for all large enough n.

In particular, if $\sigma_1 \sim \mathtt{s}_1, \sigma_2 \sim \mathtt{s}_2$ etc. are independent then (σ_n) is a sofic approximation with probability 1.

1.4. *Organization.* In §2 we define weights and discuss some of their useful properties. In §3 we prove a few basic results about the functions f and F. Some of the results of these two sections are used in §4 to show that the assumptions of the main theorems are not vacuous. In §5 we show how the function F is related to the number of homomorphism-labeling pairs $(\sigma, \mathbf{y})$ that realize a given weight, which is the main ingredient of the proofs of Theorems A and B given in the next two sections. In §8 we show how to deduce Theorem C from Theorem B. Section 9 contains a proof of Proposition 1.1. The final section contains a proof of Lemma 2.3, which asserts that a weight can be approximated by a denominator-n weight with a specified marginal.

2. *Weights*

If $\alpha : X \to \mathtt{A}$ is a finite observable, for $a, a' \in \mathtt{A}$ and $i \in [r]$ let

$$W_\alpha(a, a'; i) = \alpha_*^{\{e,s_i\}}\mu(a, a') = \mu\{x \in X : \alpha(x) = a,\ \alpha(T_{s_i}x) = a'\}$$

and also denote

$$W_\alpha(a) = \alpha_*\mu(a).$$

For $\mathbf{x} \in \mathtt{A}^n$ and $\sigma \in \mathrm{Hom}(G, \mathrm{Sym}(n))$ let

$$W_{\sigma,\mathbf{x}}(a, a'; i) = P_{\mathbf{x}}^{\sigma,\{e,s_i\}}(a, a')$$

and $W_{\sigma,\mathbf{x}}(a) = P_{\mathbf{x}}^{\sigma,\{e\}}(a)$. This could equivalently be defined as a special case of the previous construction, with σ specifying an action on $X = [n]$ with an observable $\mathbf{x} : [n] \to \mathtt{A}$.

More abstractly, any $W \in (\mathrm{Prob}(\mathtt{A}^2))^r$ is called an $\mathtt{A}$-*weight* if

$$\sum_{a'\in\mathtt{A}} W(a, a'; i) = \sum_{a'\in\mathtt{A}} W(a', a; j)$$

for all $i, j \in [r]$ and $a \in \mathtt{A}$. For each $a \in \mathtt{A}$ we denote this common value by $W(a)$. Note that the objects W_α and $W_{\sigma,\mathbf{x}}$ defined above satisfy this condition.

We say that W has denominator n if $n \cdot W(a, a'; i) \in \mathbb{N}$ for all a, a', i.

The measures $W(\cdot, \cdot; i)$ for $i \in [r]$ are called the *edge measures* of W, and $W(\cdot)$ is called the *vertex measure*.

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

For any alphabet $\mathtt{A}$, we use the metric on $\mathtt{A}$-weights defined by

$$\begin{aligned} d(W_1, W_2) &:= \sum_{i\in[r]} \|W_1(\cdot,\cdot;i) - W_2(\cdot,\cdot;i)\|_{\mathrm{TV}} \\ &= \frac{1}{2}\sum_{i\in[r]}\sum_{a,a'\in\mathtt{A}} |W_1(a,a';i) - W_2(a,a';i)|. \end{aligned}$$

We can use weights to count good models up to equivalence under the pseudometrics d_k^* using the following proposition.

PROPOSITION 2.1. *If* $\sigma \in \mathrm{Hom}(G, \mathrm{Sym}(n))$ *and* $\mathbf{x} \in \mathtt{A}^n$, *then for any observable* $\alpha : X \to \mathtt{A}$,

$$d(W_{\sigma,\mathbf{x}^k}, W_{\alpha^k}) = d_k^*(P_{\mathbf{x}}^{\sigma}, \alpha_*^G\mu).$$

Note this implies also that

$$d_k^*(P_{\mathbf{x}}^{\sigma}, \alpha_*^G\mu) = d_0^*(P_{\mathbf{x}^k}^{\sigma}, (\alpha^k)_*^G\mu).$$

Proof. By definition of the distance between weights,

$$\begin{aligned} d(W_{\sigma,\mathbf{x}^k}, W_{\alpha^k}) &= \frac{1}{2}\sum_{i\in[r]}\sum_{\mathbf{a},\mathbf{a}'\in\mathtt{A}^{\mathrm{B}(e,k)}} |W_{\sigma,\mathbf{x}^k}(\mathbf{a},\mathbf{a}';i) - W_{\alpha^k}(\mathbf{a},\mathbf{a}';i)| \\ &= \frac{1}{2}\sum_{i\in[r]}\sum_{\mathbf{a},\mathbf{a}'\in\mathtt{A}^{\mathrm{B}(e,k)}} \left| \frac{1}{n}\left| \left\{ j\in[n] : \begin{matrix} (\mathbf{x}^k)_j = \mathbf{a} \\ (\mathbf{x}^k)_{\sigma(s_i)j} = \mathbf{a}' \end{matrix} \right\} \right| \right. \\ &\qquad \left. - \mu\left\{ x\in X : \begin{matrix} \alpha^k(x) = \mathbf{a} \\ \alpha^k(T_{s_i}x) = \mathbf{a}' \end{matrix} \right\} \right|. \end{aligned}$$

For many 'incompatible' pairs $\mathbf{a}, \mathbf{a}'$, both terms will be zero: suppose $g \in \mathrm{B}(e,k) \cap \mathrm{B}(s_i, k)$, so that $gs_i^{-1} \in \mathrm{B}(e,k)$. If the second term in the absolute value is non-zero, then for some $x \in X$ we have $\alpha^k(x) = \mathbf{a}$ and $\alpha^k(T_{s_i}x) = \mathbf{a}'$, and therefore

$$\mathbf{a}'_{gs_i^{-1}} = (\alpha^k(T_{s_i}x))_{gs_i^{-1}} = \alpha(T_{gs_i^{-1}}T_{s_i}x) = \alpha(T_g x) = (\alpha^k(x))_g = \mathbf{a}_g.$$

The same argument shows that $\mathbf{a}'_{gs_i^{-1}} = \mathbf{a}_g$ for all $g \in \mathrm{B}(e,k) \cap \mathrm{B}(s_i,k)$ whenever the first term is non-zero. Therefore we can restrict the sum to pairs $\mathbf{a}, \mathbf{a}'$ with $\mathbf{a}'_{gs_i^{-1}} = \mathbf{a}_g$ for all $g \in \mathrm{B}(e,k) \cap \mathrm{B}(s_i,k)$. Equivalently, we can sum over all $\mathbf{A} \in \mathtt{A}^{\mathrm{B}(e,k)\cup\mathrm{B}(s_i,k)}$ to get

$$\begin{aligned} d(W_{\sigma,\mathbf{x}^k}, W_{\alpha^k}) &= \frac{1}{2}\sum_{i\in[r]}\sum_{\mathbf{A}\in\mathtt{A}^{\mathrm{B}(e,k)\cup\mathrm{B}(s_i,k)}} \left| \frac{1}{n}|\{j\in[n] : (\mathbf{x}^{\mathrm{B}(e,k)\cup\mathrm{B}(s_i,k)})_j = \mathbf{A}\}| \right. \\ &\qquad \left. - \mu\{x\in X : \alpha^{\mathrm{B}(e,k)\cup\mathrm{B}(s_i,k)}(x) = \mathbf{A}\} \right| \\ &= \sum_{i\in[r]} d^{\mathrm{B}(e,k)\cup\mathrm{B}(s_i,k)}(P_{\mathbf{x}}^{\sigma}, \alpha_*^G\mu). \end{aligned}$$

□

It will be useful to consider the pushforward map induced by a map between alphabets: if $\pi : \mathtt{A} \to \mathtt{B}$ is a measurable map and W is an $\mathtt{A}$-weight, then πW is the $\mathtt{B}$-weight given by

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

$$\pi W(b, b'; i) = \sum_{a \in \pi^{-1}\{b\}} \sum_{a' \in \pi^{-1}\{b'\}} W(a, a'; i).$$

Note that this implies that the vertex measure of W is

$$\pi W(b) = \sum_{a \in \pi^{-1}\{b\}} W(a).$$

For example, let $\pi_{\mathtt{B}} : \mathtt{A} \times \mathtt{B} \to \mathtt{B}$ be the projection map. If W is an $\mathtt{A} \times \mathtt{B}$-weight then $\pi_{\mathtt{B}} W$ is given by

$$\pi_{\mathtt{B}} W(b_1) = \sum_{a \in \mathtt{A}} W((a, b_1)) \quad \pi_{\mathtt{B}} W(b_1, b_2; i) = \sum_{a_1, a_2 \in \mathtt{A}} W((a_1, b_1), (a_2, b_2); i).$$

We call this the $\mathtt{B}$-marginal of W.

All weights in the present paper will be over alphabets of the form $\mathtt{A}^{\mathrm{B}(e,k)} \times \mathtt{B}^{\mathrm{B}(e,k')}$. We use this fact to introduce some simplified notation for projections.

- π_A denotes projection onto the entire $\mathtt{A}$ factor $\mathtt{A}^{\mathrm{B}(e,k)}$; π_B is used similarly.
- For $m < k$ and $m' < k'$, $\pi_{m,m'}$ denotes projection onto $\mathtt{A}^{\mathrm{B}(e,m)} \times \mathtt{B}^{\mathrm{B}(e,m')}$.
- π_m denotes the projection $\mathtt{A}^{\mathrm{B}(e,k)} \to \mathtt{A}^{\mathrm{B}(e,m)}$, except that if $m = 0$ we write π_e.

We define $F(W)$ for an abstract weight W by

$$F(W) = (1 - 2r)\mathrm{H}(W(\cdot)) + \sum_{i \in [r]} \mathrm{H}(W(\cdot, \cdot; i))$$

where H is the Shannon entropy. Note that this is consistent with the above definitions in that, for example,

$$F(W_\alpha) = F_\mu(T, \alpha).$$

We can revisit the definition of our version of the stochastic block model using weights. Let $H \subset G$ and let W be a denominator-n $\mathtt{B}^{\mathrm{B}(e,k)}$-weight. Suppose there exist $\mathbf{y} \in \mathtt{B}^n$ and $\sigma \in \mathrm{Hom}(G, \mathrm{Sym}(n))$ such that $W = W_{\sigma, \mathbf{y}^k}$. Then

$$\mathtt{SBM}(\sigma, \mathbf{y}, k) = \mathrm{Unif}(\{\sigma' \in \mathrm{Hom}(G, \mathrm{Sym}(n)) : W_{\sigma', \mathbf{y}^k} = W\}),$$

so we can also denote this distribution by $\mathtt{SBM}(\mathbf{y}, W)$. Specifying the distribution by a weight rather than a specific homomorphism will occasionally be more convenient.

2.1. *Constructing weights and good models.* We borrow the first result of this type from [2]; it allows us to find a denominator-n approximation to a given weight.

LEMMA 2.2. (Lemma 2.3 of [2]) *There is a constant C such that for any $\mathtt{A}$-weight W there is a denominator-n $\mathtt{A}$-weight within distance $C|\mathtt{A}|^2 r/n$ of W.*

The following lemma allows us not only to construct a denominator-n approximation to a given weight, but also to specify a marginal of this approximation:

LEMMA 2.3. *Let W be an $\mathtt{A} \times \mathtt{B}$-weight. If $W_{\mathtt{B}}$ is a $\mathtt{B}$-weight of denominator n with $d(W_{\mathtt{B}}, \pi_{\mathtt{B}} W) < \delta$ then there is an $\mathtt{A} \times \mathtt{B}$-weight $W_{\mathtt{AB}}$ with denominator n such that $\pi_{\mathtt{B}} W_{\mathtt{AB}} = W_{\mathtt{B}}$ and $d(W_{\mathtt{AB}}, W) < 265r(|\mathtt{A} \times \mathtt{B}|^2/n + \delta)$.*

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

The construction is fairly involved, so it is postponed to §10. The constant 265 is not intended to be optimal.

The definition of a weight $W_{\sigma,\mathbf{x}^k}$ in terms of a homomorphism σ and a labeling $\mathbf{x}$ is straightforward. However, we will also need to know whether a given weight can be realized in this way. The next two results address this inverse problem.

PROPOSITION 2.4. *If W is a denominator-n $\mathtt{A}$-weight, then there exist $\mathbf{x} \in \mathtt{A}^n$ and $\sigma \in \mathrm{Hom}(G, \mathrm{Sym}(n))$ such that $W = W_{\sigma,\mathbf{x}}$.*

Proof. This is implied by Proposition 2.1 of [2]. □

Unfortunately, this does not imply that for every denominator-n $\mathtt{A}^{\mathrm{B}(e,k)}$-weight W there is some $\sigma \in \mathrm{Hom}(G, \mathrm{Sym}(n))$ and $\mathbf{x} \in \mathtt{A}^n$ such that $W = W_{\sigma,\mathbf{x}^k}$; instead it provides $\mathbf{X} \in (\mathtt{A}^{\mathrm{B}(e,k)})^n$ such that $W = W_{\sigma,\mathbf{X}}$.

However, if we already know that W is close to a weight of the form W_{α^k} for some observable α, then the following proposition shows that W is also close to a weight of the form $W_{\sigma,\mathbf{x}^k}$.

PROPOSITION 2.5. *Let $\alpha : X \to \mathtt{A}$, $\sigma \in \mathrm{Hom}(G, \mathrm{Sym}(n))$, and $\mathbf{X} \in (\mathtt{A}^{\mathrm{B}(e,k)})^n$ be such that $d(W_{\sigma,\mathbf{X}}, W_{\alpha^k}) \le \varepsilon$ for some $\varepsilon \ge 0$. Writing $\mathbf{x} = \pi_e \mathbf{X} \in \mathtt{A}^n$, we have*

$$d(W_{\sigma,\mathbf{X}}, W_{\sigma,\mathbf{x}^k}) \le 2r|\mathrm{B}(e,k)|\varepsilon.$$

An immediate consequence is that $\mathbf{X} \in \Omega_0^*(\sigma, \alpha^k, \varepsilon)$ implies $\pi_e \mathbf{X} \in \Omega_k^*(\sigma, \alpha, c\varepsilon)$ where $c = 1 + 2r|\mathrm{B}(e,k)|$; cf. Claim 2 in the proof of Proposition 3.2 of [2].

Proof. Claim 4 in the proof of Proposition 3.2 of [2] implies that

$$|\{j \in [n] : \mathbf{X}(j) \ne \mathbf{x}^k(j)\}| \le n|\mathrm{B}(e,k)|\varepsilon.$$

It follows that for any $i \in [r]$,

$$\begin{aligned}
&|\{j \in [n] : \mathbf{X}^{\{e,s_i\}}(j) \ne (\mathbf{x}^k)^{\{e,s_i\}}(j)\}| \\
&\quad \le |\{j \in [n] : \mathbf{X}(j) \ne \mathbf{x}^k(j)\}| + |\{j \in [n] : \mathbf{X}(\sigma(s_i)j) \ne \mathbf{x}^k(\sigma(s_i)j)\}| \\
&\quad \le 2n|\mathrm{B}(e,k)|\varepsilon,
\end{aligned}$$

so

$$\begin{aligned}
d(W_{\sigma,\mathbf{X}}, W_{\sigma,\mathbf{x}^k}) &= \sum_{i\in[r]} \|(\mathbf{X}^{\{e,s_i\}})_* \,\mathrm{Unif}(n) - ((\mathbf{x}^k)^{\{e,s_i\}})_* \,\mathrm{Unif}(n)\|_{\mathrm{TV}} \\
&\le \sum_{i\in[r]} 2|\mathrm{B}(e,k)|\varepsilon = 2r|\mathrm{B}(e,k)|\varepsilon.
\end{aligned}$$

□

The following corollary of the first part of the proof will be useful later. It says that if the weight $W_{\sigma,\mathbf{X}}$ generated by some $\mathbf{X} \in (\mathtt{A}^{\mathrm{B}(e,k)})^n$ and $\sigma \in \mathrm{Hom}(G, \mathrm{Sym}(n))$ is exactly attainable in some sense, then $\mathbf{X}$ can be exactly recovered from σ and the projection $\pi_e \mathbf{X} \in \mathtt{A}^n$.

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

COROLLARY 2.6. *Suppose that* $\sigma \in \mathrm{Hom}(G, \mathrm{Sym}(n))$ *and* $\mathbf{X} \in (\mathtt{A}^{\mathrm{B}(e,k)})^n$ *are such that either*

(1) $W_{\sigma,\mathbf{X}} = W_{\alpha^k}$ *for some* $\alpha : X \to \mathtt{A}$, *or*

(2) $W_{\sigma,\mathbf{X}} = W_{\sigma_0,\mathbf{x}_0^k}$ *for some* $\sigma_0 \in \mathrm{Hom}(G, \mathrm{Sym}(m))$ *and* $\mathbf{x}_0 \in \mathtt{A}^m$.

Then $(\pi_e\mathbf{X})^k = \mathbf{X}$.

Note that $(\pi_e\mathbf{X})^k$ is the k-neighborhood labeling generated from $\pi_e\mathbf{X}$ using σ, rather than σ_0 or some other homomorphism.

Proof. In the first case, we are in the setting of the previous proposition with $\varepsilon = 0$, so the first inequality of its proof gives the claimed result.

The second case is actually the same; this is only obscured somewhat by the notation. We are in the setting of the previous proposition with the space $X = [m]$ having a G-action specified by σ_0 and a finite observable $\mathbf{x}_0 : [m] \to \mathtt{A}$. □

3. *Properties of F and f*

LEMMA 3.1. (Continuity as weight function) *If* W_1, W_2 *are* $\mathtt{A}$*-weights with* $d(W_1, W_2) \leq \varepsilon \leq 1$ *then*

$$|F(W_1) - F(W_2)| \leq 4r(\mathrm{H}(\varepsilon) + \varepsilon \log_2|\mathtt{A}|),$$

where $\mathrm{H}(p)$ *denotes the entropy of the probability measure* $(p, 1-p) \in \mathrm{Prob}(\{0,1\})$.

Proof. We use Fano's inequality in the following form (equation (2.139) of [9]). Suppose X, Y are $\mathtt{A}$-valued random variables defined on the same probability space and let $p_e = \mathbb{P}(X \neq Y)$ be their probability of disagreement. Then

$$\mathrm{H}(X \mid Y) \leq \mathrm{H}(p_e) + p_e \log|\mathtt{A}|.$$

Using the chain rule and non-negativity of Shannon entropy, we can deduce that

$$|\mathrm{H}(X) - \mathrm{H}(Y)| \leq \mathrm{H}(p_e) + p_e \log|\mathtt{A}|.$$

Let $\mu_1, \mu_2 \in \mathrm{Prob}(\mathtt{A})$ be the respective distributions of X_1, X_2. Because $\|\mu_1 - \mu_2\|_{\mathrm{TV}}$ is the minimum value of $\mathbb{P}(X_1 \neq X_2)$ over all possible couplings, if $\|\mu_1 - \mu_2\|_{\mathrm{TV}} < \varepsilon$ then

$$|\mathrm{H}(\mu_1) - \mathrm{H}(\mu_2)| \leq \mathrm{H}(\varepsilon) + \varepsilon \log|\mathtt{A}|.$$

The assumed bound $d(W_1, W_2) \leq \varepsilon$ implies that each vertex and edge measure of W_1 is within total variation distance ε of its counterpart in W_2, so

$$\begin{aligned}
|F(W_1) - F(W_2)| &\leq |1-2r| \cdot |\mathrm{H}(W_1(\cdot)) - \mathrm{H}(W_2(\cdot))| \\
&\quad + \sum_{i \in [r]} |\mathrm{H}(W_1(\cdot,\cdot;i)) - \mathrm{H}(W_2(\cdot,\cdot;i))| \\
&\leq (2r-1)(\mathrm{H}(\varepsilon) + \varepsilon \log|\mathtt{A}|) \\
&\quad + r \cdot (\mathrm{H}(\varepsilon) + \varepsilon \log|\mathtt{A}|^2) \\
&\leq 4r(\mathrm{H}(\varepsilon) + \varepsilon \log|\mathtt{A}|).
\end{aligned}$$

□

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

Let $\alpha : X \to \mathtt{A}$ and $\beta : X \to \mathtt{B}$ be observables. We say that β is a *coarsening* of α if each part of the partition of X induced by β is a union of parts of the partition induced by α (up to null sets). Equivalently, there is some function $g : \mathtt{A} \to \mathtt{B}$ such that $\beta = g \circ \alpha$ almost surely. In this situation we can also call α a refinement of β.

A useful property of the Shannon entropy $\mathrm{H}_\mu(\alpha)$ is monotonicity under refinement. The function F does not share this property, but it is monotone under the following particular kind of refinement introduced in [3].

We say that β is a *simple splitting* of α if there is some $s \in \{s_1^{\pm 1}, \ldots, s_r^{\pm 1}\}$ and a coarsening $\tilde{\alpha}$ of α such that, up to null sets, the partition induced by β is the coarsest common refinement of the partitions induced by α and $\tilde{\alpha} \circ T_s$.

We say that β is a *splitting* of α if there are observables $\alpha = \beta_0, \beta_1, \ldots, \beta_n = \beta$ such that β_i is a simple splitting of β_{i-1} for $i = 1, 2, \ldots, n$. We will use the following monotonicity properties of the relative version of F.

LEMMA 3.2. (Monotonicity under splitting)

(1) *If α_1 is a splitting of α_2 then $F(\alpha_1|\beta) \leq F(\alpha_2|\beta)$.*
(2) *If β_1 is a splitting of β_2 then $F(\alpha|\beta_1) \geq F(\alpha|\beta_2)$.*

Proof. (1) This is essentially Proposition 5.1 of [3]; conditioning on β makes no difference to the proof.

(2) The proof is based on the proof of part (1), but in place of the chain rule for conditional entropy we use the following bound:

$$\begin{aligned}
\mathrm{H}(\alpha \mid \beta_2) &\leq \mathrm{H}(\alpha, \beta_1 \mid \beta_2) \quad \text{(monotonicity)} \\
&= \mathrm{H}(\beta_1 \mid \beta_2) + \mathrm{H}(\alpha \mid \beta_1, \beta_2) \quad \text{(chain rule)} \\
&\leq \mathrm{H}(\beta_1 \mid \beta_2) + \mathrm{H}(\alpha \mid \beta_1) \quad \text{(monotonicity)}.
\end{aligned}$$

We will also use the following consequence of the previous bound:

$$\begin{aligned}
&\mathrm{H}(\alpha^{\{e,s_i\}} \mid \beta_1^{\{e,s_i\}}) - \mathrm{H}(\alpha^{\{e,s_i\}} \mid \beta_2^{\{e,s_i\}}) \\
&\quad \geq -\mathrm{H}(\beta_1^{\{e,s_i\}} \mid \beta_2^{\{e,s_i\}}) \quad \text{(previous bound)} \\
&\quad \geq -(\mathrm{H}(\beta_1^{\{s_i\}} \mid \beta_2^{\{e,s_i\}}) + \mathrm{H}(\beta_1 \mid \beta_2^{\{e,s_i\}})) \quad \text{(subadditivity)} \\
&\quad = -(\mathrm{H}(\beta_1 \mid \beta_2^{\{e,s_i^{-1}\}}) + \mathrm{H}(\beta_1 \mid \beta_2^{\{e,s_i\}})) \quad (T\text{-invariance of } \mu).
\end{aligned}$$

It suffices to check the case where β_1 is a simple splitting of β_2. Let $t \in \{s_1^{\pm 1}, \ldots, s_r^{\pm 1}\}$ and let $\tilde{\beta}$ be a coarsening of β_2 such that the partition induced by β_1 is the same as the coarsest common refinement of the partitions induced by β_2 and $\tilde{\beta} \circ T_t$ up to null sets. Then, using the two bounds just derived,

$$\begin{aligned}
F(\alpha|\beta_1) - F(\alpha|\beta_2) &= (1 - 2r)(\mathrm{H}(\alpha|\beta_1) - \mathrm{H}(\alpha|\beta_2)) \\
&\quad + \sum_{i \in [r]} (\mathrm{H}(\alpha^{\{e,s_i\}}|\beta_1^{\{e,s_i\}}) - \mathrm{H}(\alpha^{\{e,s_i\}}|\beta_1^{\{e,s_i\}}))
\end{aligned}$$

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

$$\begin{aligned}&\geq (1-2r)(-\mathrm{H}(\beta_1|\beta_2)) - \sum_{i\in[r]}(\mathrm{H}(\beta_1 \mid \beta_2^{\{e,s_i^{-1}\}}) \\ &\quad + \mathrm{H}(\beta_1 \mid \beta_2^{\{e,s_i\}})) \\ &= (2r-1)\mathrm{H}(\beta_1|\beta_2) - \sum_{s\in\{s_1^{\pm 1}\ldots s_r^{\pm 1}\}} \mathrm{H}(\beta_1 \mid \beta_2^{\{e,s\}})\end{aligned}$$

But

$$\mathrm{H}(\beta_1 \mid \beta_2^{\{e,t\}}) \leq \mathrm{H}(\beta_1 \mid \beta_2\tilde{\beta}^{\{t\}}) = 0,$$

so we can remove the t term from the sum to get

$$\begin{aligned}F(\alpha|\beta_1) - F(\alpha|\beta_2) &\geq (2r-1)\mathrm{H}(\beta_1|\beta_2) - \sum_{s\in\{s_1^{\pm 1}\ldots s_r^{\pm 1}\}\setminus\{t\}} \mathrm{H}(\beta_1 \mid \beta_2^{\{e,s\}}) \\ &= \sum_{s\in\{s_1^{\pm 1}\ldots s_r^{\pm 1}\}\setminus\{t\}} (\mathrm{H}(\beta_1|\beta_2) - \mathrm{H}(\beta_1 \mid \beta_2^{\{e,s\}})) \\ &\geq 0.\end{aligned}$$

□

One corollary is the following convenient formula.

COROLLARY 3.3. *Let α, β be finite observables such that $\beta_*^G\mu$ is a Markov measure. Then $F_\mu(T, \alpha^{k_1} \mid \beta^{k_2})$ is independent of k_2. In particular,*

$$f_\mu(T, \alpha \mid \beta) = \inf_k F_\mu(T, \alpha^k \mid \beta).$$

Proof. By the previous proposition, for any $k \leq k_2$ we have

$$F_\mu(T, \alpha^{k_1} \mid \beta^k) \leq F_\mu(T, \alpha^{k_1} \mid \beta^{k_2}).$$

On the other hand, by Theorem 6.1 of [5] $F_\mu(T, \beta^k) = F_\mu(T, \beta^{k_2})$ so

$$F_\mu(T, \alpha^{k_1} \mid \beta^k) = F_\mu(T, \alpha^{k_1}\beta^k) - F_\mu(T, \beta^{k_2}).$$

Applying monotonicity under splitting to the first term on the right gives

$$F_\mu(T, \alpha^{k_1} \mid \beta^k) \geq F_\mu(T, \alpha^{k_1}\beta^{k_2}) - F_\mu(T, \beta^{k_2}) = F_\mu(T, \alpha^{k_1} \mid \beta^{k_2}).$$

This establishes independence of k_2; the formula for f follows. □

PROPOSITION 3.4. *Let α, β be finite observables. Then for any $k \in \mathbb{N}$,*

$$F_\mu(T, \alpha^k \mid \beta) \leq \mathrm{H}_\mu(\alpha \mid \beta).$$

It follows that

$$f_\mu(T, \alpha \mid \beta) \leq \mathrm{H}_\mu(\alpha \mid \beta).$$

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

Proof. By Lemma 3.2, $F_\mu(T, \alpha^k \mid \beta) \leq F_\mu(T, \alpha \mid \beta)$. Using elementary properties of Shannon entropy, we have

$$\begin{aligned} F_\mu(T, \alpha \mid \beta) &= (1-2r)\mathrm{H}_\mu(\alpha \mid \beta) + \sum_{i\in[r]} \mathrm{H}_\mu(\alpha^{\{e,s_i\}} \mid \beta^{\{e,s_i\}}) \\ &\leq (1-2r)\mathrm{H}_\mu(\alpha \mid \beta) + \sum_{i\in[r]} [\mathrm{H}_\mu(\alpha \mid \beta^{\{e,s_i\}}) + \mathrm{H}_\mu(\alpha^{\{s_i\}} \mid \beta^{\{e,s_i\}})] \\ &\leq (1-2r)\mathrm{H}_\mu(\alpha \mid \beta) + \sum_{i\in[r]} [\mathrm{H}_\mu(\alpha \mid \beta) + \mathrm{H}_\mu(\alpha^{\{s_i\}} \mid \beta^{\{s_i\}})]. \end{aligned}$$

By T-invariance of μ we have

$$\mathrm{H}_\mu(\alpha^{\{s_i\}} \mid \beta^{\{s_i\}}) = \mathrm{H}_\mu(\alpha \mid \beta),$$

so the first inequality follows.

For any $k_1, k_2 \in \mathbb{N}$ this gives

$$F_\mu(T, \alpha^{k_1} \mid \beta^{k_2}) \leq \mathrm{H}_\mu(\alpha \mid \beta^{k_2}) \leq \mathrm{H}_\mu(\alpha \mid \beta),$$

so the second inequality follows upon taking the supremum over k_2 then the infimum over k_1. □

We can use this bound to give a proof of the chain rule for the relative f-invariant, a version of which first appeared in [5] (there it is called the Abramov–Rokhlin formula; see also [7]).

COROLLARY 3.5. (Chain rule)

$$f_\mu(T, \alpha\beta) = f_\mu(T, \alpha \mid \beta) + f_\mu(T, \beta).$$

Proof. By definition of the relative version of F and the chain rule for conditional entropy, for each k_1, k_2 we have

$$F_\mu(T, \alpha^{k_1}\beta^{k_2}) = F_\mu(T, \alpha^{k_1} \mid \beta^{k_2}) + F_\mu(T, \beta^{k_2}).$$

By Lemma 3.2 each term is monotone in k_2, so the limits as $k_2 \to \infty$ exist. By Proposition 3.4 all terms are bounded above (recall we only consider finite observables, so in particular all observables have finite entropy), so we can split the limit across the sum on the right to get

$$\lim_{k_2\to\infty} F_\mu(T, \alpha^{k_1}\beta^{k_2}) = \lim_{k_2\to\infty} F_\mu(T, \alpha^{k_1} \mid \beta^{k_2}) + f_\mu(T, \beta).$$

Taking k_1 to infinity gives the result. □

4. *Non-vacuity of main theorems*

4.1. *Theorem A.* Here we prove Proposition A, which asserts the non-vacuity of Theorem A. Given $\beta : X \to \mathtt{B}$, we need to show that there exist $\mathbf{y}_n \in \mathtt{B}^n$ and $\sigma_n \in \mathrm{Hom}(G, \mathrm{Sym}(n))$ such that $\lim_{n\to\infty} d_0^*(P_{\mathbf{y}_n}^{\sigma_n}, \beta_*^G \mu) = 0$.

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

By Lemma 2.2, there is a sequence $\{W_n\}_{n=1}^{\infty}$ of $\mathtt{B}$-weights such that W_n has denominator n for each n and $d(W_n, W_\beta) = o(1)$. By Proposition 2.4, for each n we can pick $\mathbf{y}_n, \sigma_n$ such that $W_{\sigma_n,\mathbf{y}_n} = W_n$. Since $d_0^*(P_{\mathbf{y}_n}^{\sigma_n}, \beta_*^G\mu) = d(W_{\sigma_n,\mathbf{y}_n}, W_\beta)$, these suffice.

4.2. *Theorems B and C.* Here we prove Proposition B, which asserts the non-vacuity of Theorem B (and by extension Theorem C, since the assumptions are the same).

Let m_n approach infinity as n approaches infinity while satisfying $m_n = o(\log\log n)$ and let $\beta : X \to \mathtt{B}$ be a finite observable. We need to show that there exist $\mathbf{y}_n \in \mathtt{B}^n$ and $\sigma_n \in \mathrm{Hom}(G, \mathrm{Sym}(n))$ such that $d_{m_n}^*(P_{\mathbf{y}_n}^{\sigma_n}, \beta_*^G\mu) = O(1/\log n)$.

By Lemma 2.2, there is a sequence $\{W_n\}_{n=1}^{\infty}$ of weights such that W_n is a denominator-n $\mathtt{B}^{\mathrm{B}(e,m_n)}$-weight for each n and $d(W_n, W_{\beta^{m_n}}) = O(|\mathtt{B}^{\mathrm{B}(e,m_n)}|^2/n)$. By Proposition 2.4, for each n we can pick $\mathbf{Y}_n, \sigma_n$ such that $W_{\sigma_n,\mathbf{Y}_n} = W_n$. Let $\mathbf{y}_n = \pi_e\mathbf{Y}_n$. By Proposition 2.5,

$$d_{m_n}^*(P_{\mathbf{y}_n}^{\sigma_n}, \beta_*^G\mu) = d(W_{\sigma_n,\mathbf{y}_n^{m_n}}, W_{\beta^{m_n}}) = O\left(|\mathrm{B}(e, m_n)| \cdot \frac{|\mathtt{B}^{\mathrm{B}(e,m_n)}|^2}{n}\right) = O\left(\frac{1}{\log n}\right).$$

5. *Counting lemmas*

For a $\mathtt{B}$-weight W, let $Z_n(W)$ denote the number of pairs $(\sigma, \mathbf{y}) \in \mathrm{Hom}(G, \mathrm{Sym}(n)) \times \mathtt{B}^n$ such that $W_{\sigma,\mathbf{y}} = W$.

PROPOSITION 5.1. *If W is a $\mathtt{B}$-weight with denominator n then*

$$(3\sqrt{n})^{-r|\mathtt{B}|^2} \le \frac{Z_n(W)}{e^{F(W)n}(n!)^r n^{(1-r)/2}} \le (3\sqrt{n})^{r|\mathtt{B}|^2}.$$

Proof. We write

$$Z_n(W) = \sum_\sigma |\{\mathbf{y} \in \mathtt{B}^n : W_{\sigma,\mathbf{y}} = W\}| = (n!)^r \mathbb{E}_\sigma |\{\mathbf{y} \in \mathtt{B}^n : W_{\sigma,\mathbf{y}} = W\}|.$$

where $\mathbb{E}_\sigma$ denotes the expectation over a uniform choice of $\sigma \in \mathrm{Hom}(G, \mathrm{Sym}(n))$.

Proposition 2.1 of [2] states that

$$\mathbb{E}_\sigma |\{\mathbf{y} \in \mathtt{B}^n : W_{\sigma,\mathbf{y}} = W\}| = \frac{n!^{1-r} \prod_{b\in\mathtt{B}} (nW(b))!^{2r-1}}{\prod_{i=1}^r \prod_{b,b'\in\mathtt{B}} (nW(b, b'; i))!}.$$

Lemma 2.2 of the same paper gives an estimate of this quantity, but for our purposes we need to be more careful about how the estimate depends on the size of the alphabet.

We use the version of Stirling's approximation given by

$$k^{k+1/2}e^{-k} \le k! \le 3 \cdot k^{k+1/2}e^{-k},$$

valid for $k \ge 1$. To estimate the products that appear in the expectation, we will need to omit all factors which equal $0! = 1$ since Stirling's approximation is not valid for these. To do this carefully, let

$$\mathtt{B}' = \{b \in \mathtt{B} : W(b) \ne 0\}$$

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

and for each $i \in [r]$ let

$$\mathtt{B}'_i = \{(b, b') \in \mathtt{B}^2 : W(b, b'; i) \neq 0\}.$$

For the numerator of the above expectation we get

$$\begin{aligned} n!^{1-r} \prod_{b \in \mathtt{B}'} (nW(b))!^{2r-1} &\leq (3n^{n+1/2}\, e^{-n})^{1-r} \prod_{b \in \mathtt{B}'} (3(nW(b))^{nW(b)+1/2} e^{-nW(b)})^{2r-1} \\ &= 3^{1-r+|\mathtt{B}'|(2r-1)}\, n^{rn+1/2-r/2+(2r-1)|\mathtt{B}'|/2} \\ &\quad \times e^{-rn+(2r-1)[n \sum_{b \in \mathtt{B}'} W(b) \log W(b) + 1/2 \sum_{b \in \mathtt{B}'} \log W(b)]} \end{aligned}$$

and a lower bound which is identical except missing the first factor. For the denominator, let $S = \sum_{i \in [r]} |\mathtt{B}'_i|$. We get

$$\begin{aligned} \prod_{i=1}^{r} \prod_{(b,b') \in \mathtt{B}'_i} (nW(b, b'; i))! &\leq \prod_{i=1}^{r} \prod_{(b,b') \in \mathtt{B}'_i} 3(nW(b, b'; i))^{nW(b,b';i)+1/2} e^{-nW(b,b';i)} \\ &= 3^S\, n^{nr+S/2} \\ &\quad \times e^{n \sum_i \sum_{b,b'} W(b,b';i) \log W(b,b';i) + 1/2 \sum_{i,b,b'} \log W(b,b';i) - nr}, \end{aligned}$$

and again we have a lower bound which is identical except missing the first factor 3^S. Therefore the quotient is bounded above by

$$3^{1-r+|\mathtt{B}'|(2r-1)}\, n^{(1-r)/2+(2r-1)|\mathtt{B}'|/2-S/2}\, e^{nF(W)+(2r-1)\frac{1}{2}\sum_b \log W(b) - \frac{1}{2}\sum_{i,b,b'} \log W(b,b';i)}$$

and below by

$$3^{-S}\, n^{(1-r)/2+(2r-1)|\mathtt{B}'|/2-S/2}\, e^{nF(W)+(2r-1)\frac{1}{2}\sum_b \log W(b) - \frac{1}{2}\sum_{i,b,b'} \log W(b,b';i)}.$$

Since W has denominator n, we have

$$0 \geq (2r-1)\frac{1}{2} \sum_{b \in \mathtt{B}'} \log W(b) \geq (2r-1)\frac{1}{2} \sum_{b \in \mathtt{B}'} \log \frac{1}{n} = -\frac{2r-1}{2} |\mathtt{B}'| \log n$$

and

$$0 \leq -\frac{1}{2} \sum_i \sum_{(b,b') \in B'_i} \log W(b, b'; i) \leq -\frac{1}{2} \sum_i \sum_{(b,b') \in \mathtt{B}'_i} \log \frac{1}{n} = \frac{S}{2} \log n.$$

Therefore $Z_n(W)$ satisfies

$$3^{-S} n^{((1-r)-S)/2} e^{F(W)n} (n!)^r \leq Z_n(W) \leq 3^{1-r+|\mathtt{B}'|(2r-1)} n^{((1-r)+(2r-1)|\mathtt{B}'|)/2} e^{F(W)n} (n!)^r.$$

Since $S \leq r|\mathtt{B}|^2$ and $|\mathtt{B}'| \leq |\mathtt{B}|$, we conclude that

$$\begin{aligned} &3^{-r|\mathtt{B}|^2} n^{((1-r)-r|\mathtt{B}|^2)/2} e^{F(W)n} (n!)^r \\ &\quad \leq Z_n(W) \leq 3^{1-r+|\mathtt{B}|(2r-1)} n^{((1-r)+(2r-1)|\mathtt{B}|)/2} e^{F(W)n} (n!)^r, \end{aligned}$$

and the stated inequality follows. □

The following proposition establishes the connection between the relative version of F and expected numbers of good models over stochastic block models.

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

PROPOSITION 5.2. *Given any denominator-n* $(\mathtt{A} \times \mathtt{B}^{\mathrm{B}(e,k)})$*-weight* $W_{\mathtt{AB}}$*, let* $W_{\mathtt{B}}$ *denote the* $\mathtt{B}^{\mathrm{B}(e,k)}$*-weight* $\pi_{\mathtt{B}} W_{\mathtt{AB}}$*. Let* $\mathbf{y} \in \mathtt{B}^n$ *be a fixed labeling with* $p_{\mathbf{y}} = \pi_e W_{\mathtt{B}}(\cdot)$*, and let*

$$\mu = \mathtt{SBM}(\mathbf{y}, W_{\mathtt{B}}) = \mathrm{Unif}(\{\sigma \in \mathrm{Hom}(G, \mathrm{Sym}(n)) : W_{\sigma,\mathbf{y}^k} = W_{\mathtt{B}}\}),$$

assuming $W_{\mathtt{B}}$ *is such that the desired support is non-empty. Then*

$$\mathcal{E} := \mathop{\mathbb{E}}_{\sigma\sim\mu} |\{\mathbf{x} \in \mathtt{A}^n : W_{\sigma,(\mathbf{x},\mathbf{y}^k)} = W_{\mathtt{AB}}\}| = \frac{Z_n(W_{\mathtt{AB}})}{Z_n(W_{\mathtt{B}})}.$$

In particular,

$$\frac{\mathcal{E}}{e^{n(F(W_{\mathtt{AB}})-F(W_{\mathtt{B}}))}} \in ((9n)^{-r|\mathtt{B}|^2(|\mathtt{A}|^2+1)}, (9n)^{r|\mathtt{B}|^2(|\mathtt{A}|^2+1)}).$$

LEMMA 5.3. *Let* $W_{\mathtt{AB}}$ *be an* $\mathtt{A} \times \mathtt{B}^{\mathrm{B}(e,k)}$*-weight of denominator n. Then*

$$|\{(\sigma, \mathbf{x}, \mathbf{y}) : W_{\sigma,(\mathbf{x},\mathbf{y}^k)} = W_{\mathtt{AB}}\}| \in \{0, |\{(\sigma, \mathbf{x}, \mathbf{Y}) : W_{\sigma,(\mathbf{x},\mathbf{Y})} = W_{\mathtt{AB}}\}|\}.$$

Proof. Suppose $|\{(\sigma, \mathbf{x}, \mathbf{y}) : W_{\sigma,(\mathbf{x},\mathbf{y}^k)} = W_{\mathtt{AB}}\}| \neq 0$; we then need to show

$$|\{(\sigma, \mathbf{x}, \mathbf{y}) : W_{\sigma,(\mathbf{x},\mathbf{y}^k)} = W_{\mathtt{AB}}\}| = |\{(\sigma, \mathbf{x}, \mathbf{Y}) : W_{\sigma,(\mathbf{x},\mathbf{Y})} = W_{\mathtt{AB}}\}|.$$

The inequality $\leq$ is clear, since we have an injection $(\sigma, \mathbf{x}, \mathbf{y}) \mapsto (\sigma, \mathbf{x}, \mathbf{y}^k)$.

The converse inequality holds because $(\sigma, \mathbf{x}, \mathbf{Y}) \mapsto (\sigma, \mathbf{x}, \pi_e \mathbf{Y})$ in an injection from the set on the right to the set on the left. This follows from Corollary 2.6. □

Proof of Proposition 5.2. Let

$$\tilde{\mu} = \mathrm{Unif}(\{(\sigma, \tilde{\mathbf{y}}) : W_{\sigma,\tilde{\mathbf{y}}^k} = W_{\mathtt{B}}\}),$$

and let $\tilde{\mu}_2$ be its marginal on the '$\tilde{\mathbf{y}}$'-coordinate. This marginal is supported on $\{\tilde{\mathbf{y}} : p_{\tilde{\mathbf{y}}} = \pi_e W_{\mathtt{B}}(\cdot)\}$. Note that $\tilde{\mu}$ conditioned on any particular $\tilde{\mathbf{y}}$ in the support of $\tilde{\mu}_2$ is $\mathtt{SBM}(\tilde{\mathbf{y}}, W_{\mathtt{B}})$, and that

$$\mathop{\mathbb{E}}_{\sigma\sim\mathtt{SBM}(\tilde{\mathbf{y}},W_{\mathtt{B}})} |\{\mathbf{x} \in \mathtt{A}^n : W_{\sigma,(\mathbf{x},\tilde{\mathbf{y}}^k)} = W_{\mathtt{AB}}\}|$$

is the same for each $\tilde{\mathbf{y}}$ in the support of $\tilde{\mu}_2$, with one choice being $\mathbf{y}$ from the proposition statement. This is because any two choices have the same label frequencies and hence are related by a permutation of $[n]$. With the choice $\tilde{\mathbf{y}} = \mathbf{y}$ the expectation is $\mathcal{E}$, so

$$\begin{aligned}
\mathcal{E} &= \mathop{\mathbb{E}}_{\tilde{\mathbf{y}}\sim\tilde{\mu}_2} \mathcal{E} \\
&= \mathop{\mathbb{E}}_{\tilde{\mathbf{y}}\sim\tilde{\mu}_2} [\mathbb{E}_{\sigma\sim\mathtt{SBM}(\tilde{\mathbf{y}},W_{\mathtt{B}})} |\{\mathbf{x} \in \mathtt{A}^n : W_{\sigma,(\mathbf{x},\tilde{\mathbf{y}}^k)} = W_{\mathtt{AB}}\}|] \\
&= \mathop{\mathbb{E}}_{(\sigma,\tilde{\mathbf{y}})\sim\tilde{\mu}} |\{\mathbf{x} \in \mathtt{A}^n : W_{\sigma,(\mathbf{x},\tilde{\mathbf{y}}^k)} = W_{\mathtt{AB}}\}| \\
&= \frac{\sum_{\sigma,\tilde{\mathbf{y}}} |\{\mathbf{x} \in \mathtt{A}^n : W_{\sigma,(\mathbf{x},\tilde{\mathbf{y}}^k)} = W_{\mathtt{AB}}\}|}{|\{(\sigma, \tilde{\mathbf{y}}) : W_{\sigma,\tilde{\mathbf{y}}^k} = W_{\mathtt{B}}\}|} \\
&= \frac{|\{(\sigma, \mathbf{x}, \tilde{\mathbf{y}}) : W_{\sigma,(\mathbf{x},\tilde{\mathbf{y}}^k)} = W_{\mathtt{AB}}\}|}{|\{(\sigma, \tilde{\mathbf{y}}) : W_{\sigma,\tilde{\mathbf{y}}^k} = W_{\mathtt{B}}\}|}
\end{aligned}$$

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

$$= \frac{|\{(\sigma, \mathbf{x}, \mathbf{Y}) : W_{\sigma,(\mathbf{x},\mathbf{Y})} = W_{\mathtt{AB}}\}|}{|\{(\sigma, \mathbf{Y}) : W_{\sigma,\mathbf{Y}} = W_{\mathtt{B}}\}|} \quad \text{(previous lemma)}$$
$$= \frac{Z_n(W_{\mathtt{AB}})}{Z_n(W_{\mathtt{B}})}.$$

Note that our assumption that the intended support of μ is non-empty allows us to rule out the '0' case in the application of Lemma 5.3.

The rest of the result then follows from our estimates on Z_n in Proposition 5.1. □

6. *Proof of Theorem A*

6.1. *Upper bound.* Note that we will not rely on the Markov assumption for the upper bound.

For each $k \in \mathbb{N}$,

$$\begin{aligned}
&\inf_{O \ni (\alpha\beta)^G_* \mu} \limsup_{n\to\infty} \frac{1}{n} \log \mathop{\mathbb{E}}_{\sigma\sim \mathtt{S}_n} |\{\mathbf{x} \in \mathtt{A}^n : (\mathbf{x}, \mathbf{y}_n) \in \Omega(\sigma, O)\}| \\
&\quad \le \inf_{\varepsilon} \limsup_{n\to\infty} \frac{1}{n} \log \mathop{\mathbb{E}}_{\sigma\sim \mathtt{S}_n} |\{\mathbf{x} \in \mathtt{A}^n : (\mathbf{x}, \mathbf{y}_n) \in \Omega^*_k(\sigma, \alpha\beta, \varepsilon)\}| \\
&\quad = \inf_{\varepsilon} \limsup_{n\to\infty} \frac{1}{n} \log \mathop{\mathbb{E}}_{\sigma\sim \mathtt{S}_n} |\{\mathbf{x} \in \mathtt{A}^n : (\mathbf{x}^k, \mathbf{y}_n^k) \in \Omega^*_0(\sigma, (\alpha\beta)^k, \varepsilon)\}| \\
&\quad \le \inf_{\varepsilon} \limsup_{n\to\infty} \frac{1}{n} \log \mathop{\mathbb{E}}_{\sigma\sim \mathtt{S}_n} |\{\mathbf{X} \in (\mathtt{A}^{\mathtt{B}(e,k)})^n : (\mathbf{X}, \mathbf{y}_n^k) \in \Omega^*_0(\sigma, (\alpha\beta)^k, \varepsilon)\}|.
\end{aligned}$$

Write

$$\begin{aligned}
\mathcal{E}_k(n, \varepsilon) &:= \mathop{\mathbb{E}}_{\sigma\sim \mathtt{S}_n} |\{\mathbf{X} \in (\mathtt{A}^{\mathtt{B}(e,k)})^n : (\mathbf{X}, \mathbf{y}_n^k) \in \Omega^*_0(\sigma, (\alpha\beta)^k, \varepsilon)\}| \\
&= \mathop{\mathbb{E}}_{\sigma\sim \mathtt{S}_n} |\{\mathbf{X} \in (\mathtt{A}^{\mathtt{B}(e,k)})^n : d(W_{\sigma,(\mathbf{X},\mathbf{y}_n^k)}, W_{(\alpha\beta)^k}) < \varepsilon)\}|
\end{aligned}$$

and assume that n is large enough that $m_n \ge k$.

Writing $\mathcal{W}_n(\alpha\beta, k, \varepsilon)$ for the set of all denominator-n weights W with $d(W, W_{(\alpha\beta)^k}) < \varepsilon$,

$$\begin{aligned}
\mathcal{E}_k(n, \varepsilon) &= \mathop{\mathbb{E}}_{\sigma\sim \mathtt{S}_n} \sum_{W\in\mathcal{W}_n(\alpha\beta,k,\varepsilon)} |\{\mathbf{X} \in (\mathtt{A}^{\mathtt{B}(e,k)})^n : W_{\sigma,(\mathbf{X},\mathbf{y}_n^k)} = W\}| \\
&= \sum_{W\in\mathcal{W}_n(\alpha\beta,k,\varepsilon)} \mathop{\mathbb{E}}_{\sigma\sim \mathtt{S}_n} [|\{\mathbf{X} \in (\mathtt{A}^{\mathtt{B}(e,k)})^n : W_{\sigma,(\mathbf{X},\mathbf{y}_n^k)} = W\}| | W_{\sigma,\mathbf{y}_n^k} = \pi_{\mathtt{B}} W] \\
&\quad \cdot \mathop{\mathbb{P}}_{\sigma\sim \mathtt{S}_n} (W_{\sigma,\mathbf{y}_n^k} = \pi_{\mathtt{B}} W)
\end{aligned}$$

since if $W_{\sigma,\mathbf{y}_n^k} \ne \pi_{\mathtt{B}} W$ then $W_{\sigma,(\mathbf{X},\mathbf{y}_n^k)} \ne W$. But $\mathtt{s}_n$ conditioned on $\{W_{\sigma,\mathbf{y}_n^k} = \pi_{\mathtt{B}} W\}$ is $\mathtt{SBM}(\mathbf{y}_n, \pi_{\mathtt{B}} W)$, so we can bound the expectation above using Proposition 5.2, getting

$$\mathcal{E}_k(n, \varepsilon) \le (9n)^{r|\mathtt{B}^{\mathtt{B}(e,k)}|^2(|\mathtt{A}^{\mathtt{B}(e,k)}|+1)} \sum_{W\in\mathcal{W}_n(\alpha\beta,k,\varepsilon)} e^{n(F(W)-F(\pi_{\mathtt{B}} W))} \mathop{\mathbb{P}}_{\sigma\sim \mathtt{S}_n} (W_{\sigma,\mathbf{y}_n^k} = \pi_{\mathtt{B}} W).$$

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

Note $(9n)^{r|\mathrm{B}^{\mathrm{B}(e,k)}|^2(|\mathbb{A}^{\mathrm{B}(e,k)}|+1)} \le e^{o_{n\to\infty}(n)}$. Fix $\delta > 0$. By continuity of F (Lemma 3.1), for all small enough ε (possibly depending on k), we have

$$\mathcal{E}_k(n,\varepsilon) \le e^{n(F_\mu(T,\alpha^k|\beta^k)+\delta+o_{n\to\infty}(1))} \sum_{W\in\mathcal{W}_n(\alpha\beta,k,\varepsilon)} \mathop{\mathbb{P}}_{\sigma\sim\mathrm{S}_n}(W_{\sigma,\mathbf{y}_n^k} = \pi_{\mathrm{B}}W).$$

Bounding each probability by 1, we get

$$\mathcal{E}_k(n,\varepsilon) \le e^{n(F_\mu(T,\alpha^k|\beta^k)+\delta+o_{n\to\infty}(1))}|\mathcal{W}_n(\alpha\beta,k,\varepsilon)|.$$

But

$$|\mathcal{W}_n(\alpha\beta,k,\varepsilon)| \le n^{r|(\mathbb{A}\times\mathrm{B})^{\mathrm{B}(e,k)}|^2} \le e^{o_{n\to\infty}(n)},$$

so this implies

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log\mathcal{E}_k(n,\varepsilon) &\le F_\mu(T,\alpha^k \mid \beta^k)+\delta\\ &\le F_\mu(T,\alpha^k \mid \beta^{k_2})+\delta\end{aligned}$$

for any $k_2 \ge k$, by monotonicity under splitting. Taking the limit as $k_2\to\infty$ followed by the infimum over ε (which takes δ to 0) and k gives

$$\inf_{\varepsilon,k}\limsup_{n\to\infty}\frac{1}{n}\log\mathcal{E}_k(n,\varepsilon) \le f_\mu(T,\alpha\mid\beta).$$

Since

$$\begin{aligned}&\inf_{O\ni(\alpha\beta)^G_*\mu}\limsup_{n\to\infty}\frac{1}{n}\log\mathop{\mathbb{E}}_{\sigma\sim\mathrm{S}_n}|\{\mathbf{x}\in\mathbb{A}^n:(\mathbf{x},\mathbf{y}_n)\in\Omega(\sigma,O)\}|\\ &\quad\le\inf_{\varepsilon}\limsup_{n\to\infty}\frac{1}{n}\log\mathcal{E}_k(n,\varepsilon)\end{aligned}$$

for every k, this completes the upper bound.

6.2. *Lower bound.* Fix $k\in\mathbb{N}$. To estimate

$$\mathcal{E} := \mathop{\mathbb{E}}_{\sigma\sim\mathrm{S}_n}|\{\mathbf{x}\in\mathbb{A}^n:(\mathbf{x},\mathbf{y}_n)\in\Omega_k^*(\sigma,\alpha\beta,\varepsilon)\}|$$

we bound below using the expected size of

$$\mathcal{X}_k(\sigma,\alpha\beta,\varepsilon\mid\mathbf{y}_n) := \{\mathbf{X}\in(\mathbb{A}^{\mathrm{B}(e,k)})^n:(\mathbf{X},\mathbf{y}_n^k)\in\Omega_0^*(\sigma,(\alpha\beta)^k,\varepsilon)\}.$$

This is not a true lower bound but, by equation (7.1) below, there are constants C,d,c independent of n such that

$$\begin{aligned}|\mathcal{X}_k(\sigma,\alpha\beta,\varepsilon\mid\mathbf{y}_n)| &\le C\exp(nd\varepsilon+n\mathrm{H}(2|\mathrm{B}(e,k)|\varepsilon))\\ &\quad\cdot|\{\mathbf{x}\in\mathbb{A}^n:(\mathbf{x},\mathbf{y}_n)\in\Omega_k^*(\sigma,\alpha\beta,\varepsilon)\}|.\end{aligned}$$

The 'error' factor has an exponential growth rate which vanishes as $\varepsilon\to 0$, so will not be a problem.

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

We now find a lower bound for the expectation of $|\mathcal{X}_k|$. Applying Proposition 5.2 as above, we have

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\sigma\sim \mathtt{S}_n} |\mathcal{X}_k(\sigma, \alpha\beta, \varepsilon \mid \mathbf{y}_n)| \\
&\quad = \sum_{W\in\mathcal{W}_n(\alpha\beta,k,\varepsilon)} \mathop{\mathbb{E}}_{\sigma\sim \mathtt{S}_n} |\{\mathbf{X} \in (\mathtt{A}^{\mathtt{B}(e,k)})^n : W_{\sigma,(\mathbf{X},\mathbf{y}_n^k)} = W\}| \\
&\quad \geq \sum_{W\in\mathcal{W}_n(\alpha\beta,k,\varepsilon)} \exp[n(F(W) - F(\pi_{\mathtt{B}}W) - o_n(1))] \mathop{\mathbb{P}}_{\sigma\sim \mathtt{S}_n} (\pi_{\mathtt{B}}W = W_{\sigma,\mathbf{y}_n^k}).
\end{aligned}$$

For any $\delta > 0$, for small enough $\varepsilon > 0$ (independent of n), by continuity of F this is at least

$$\exp[n(F_\mu(\alpha^k \mid \beta^k) - \delta - o_n(1))] \sum_{W\in\mathcal{W}_n(\alpha\beta,k,\varepsilon)} \mathop{\mathbb{P}}_{\sigma\sim \mathtt{S}_n} (\pi_{\mathtt{B}}W = W_{\sigma,\mathbf{y}_n^k}).$$

We give a lower bound for the sum by first rewriting it as

$$\sum_{W_{\mathtt{B}} \text{ denom.-}n \ \mathtt{B}^{\mathtt{B}(e,k)}\text{-weight}} |\{W \in \mathcal{W}_n(\alpha\beta, k, \varepsilon) : \pi_{\mathtt{B}}W = W_{\mathtt{B}}\}| \cdot \mathop{\mathbb{P}}_{\sigma\sim \mathtt{S}_n} (W_{\sigma,\mathbf{y}_n^k} = W_{\mathtt{B}}).$$

Fix $\eta > 0$. By Lemma 2.3, for all large enough n the $\mathtt{B}$-weight $W_{\sigma_n,\mathbf{y}_n}$ can be extended to a $\mathtt{B}^{\mathtt{B}(e,k)}$-weight $W_{\mathtt{B}}$ with $d(W_{\mathtt{B}}, W_{\beta^k}) \leq \eta$; to apply the lemma we can think of the extended weight $W_{\mathtt{B}}$ as having alphabet $\mathtt{B}^{\mathtt{B}(e,k)\setminus\{e\}} \times \mathtt{B}$, and recall that we assume $\lim_{n\to\infty} d(W_{\sigma_n,\mathbf{y}_n}, W_\beta) = 0$. Choose $\sigma, \mathbf{Y}$ such that $W_{\sigma,\mathbf{Y}} = W_{\mathtt{B}}$. Since $\pi_e W_{\mathtt{B}} = W_{\sigma_n,\mathbf{y}_n}$, it must be that $\pi_e\mathbf{Y}$ is a permutation of $\mathbf{y}_n$: they must assign labels with the same frequencies since

$$p_{\pi_e\mathbf{Y}}(\cdot) = (\pi_e W_{\mathtt{B}})(\cdot) = W_{\sigma_n,\mathbf{y}_n}(\cdot) = p_{\mathbf{y}_n}(\cdot).$$

Therefore we can choose $\sigma, \mathbf{Y}$ such that $\pi_e\mathbf{Y} = \mathbf{y}_n$.

Let $\widetilde{W}_{\mathtt{B}} = W_{\sigma,\mathbf{y}_n^k} = W_{\sigma,(\pi_e\mathbf{Y})^k}$. By Proposition 2.5,

$$d(\widetilde{W}_{\mathtt{B}}, W_{\beta^k}) \leq d(\widetilde{W}_{\mathtt{B}}, W_{\mathtt{B}}) + d(W_{\mathtt{B}}, W_{\beta^k}) \leq 2r|\mathtt{B}(e,k)|\eta + \eta.$$

So, as long as η is small enough and n is large enough (depending on ε, k), by Lemma 2.3,

$$|\{W \in \mathcal{W}_n(\alpha\beta, k, \varepsilon) : \pi_{\mathtt{B}}W = W_{\mathtt{B}}\}| \geq 1.$$

Now consider the probability appearing in the $\widetilde{W}_{\mathtt{B}}$ term:

$$\mathop{\mathbb{P}}_{\sigma\sim \mathtt{S}_n} (W_{\sigma,\mathbf{y}_n^k} = \widetilde{W}_{\mathtt{B}}) = \frac{|\{\sigma : W_{\sigma,\mathbf{y}_n^k} = \widetilde{W}_{\mathtt{B}}\}|}{|\{\sigma : W_{\sigma,\mathbf{y}_n} = W_{\sigma_n,\mathbf{y}_n}\}|}.$$

By symmetry in the choice of $\mathbf{y}$ with the correct letter frequencies (any two $\mathbf{y}$ with the same $p_{\mathbf{y}}$ are related by a permutation of $[n]$, so have the same number of σ which give a particular weight), we can write this as

$$\begin{aligned}
\mathop{\mathbb{P}}_{\sigma\sim \mathtt{S}_n} (W_{\sigma,\mathbf{y}_n^k} = \widetilde{W}_{\mathtt{B}}) &= \frac{|\{(\sigma,\mathbf{y}) : W_{\sigma,\mathbf{y}^k} = \widetilde{W}_{\mathtt{B}}\}|}{|\{(\sigma,\mathbf{y}) : W_{\sigma,\mathbf{y}} = W_{\sigma_n,\mathbf{y}_n}\}|} \\
&= \frac{|\{(\sigma,\mathbf{Y}) : W_{\sigma,\mathbf{Y}} = \widetilde{W}_{\mathtt{B}}\}|}{|\{(\sigma,\mathbf{y}) : W_{\sigma,\mathbf{y}} = W_{\sigma_n,\mathbf{y}_n}\}|} \quad \text{(Lemma 5.3)}
\end{aligned}$$

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

$$\begin{aligned}
&= \frac{Z_n(\widetilde{W}_{\mathtt{B}})}{Z_n(W_{\sigma_n,\mathbf{y}_n})} \quad \text{(definition of } Z_n\text{)} \\
&\geq \exp(n[F(\widetilde{W}_{\mathtt{B}}) - F(W_{\sigma_n,\mathbf{y}_n})]) \cdot (3\sqrt{n})^{-r(|\mathtt{B}^{\mathtt{B}(e,k)}|^2 - |\mathtt{B}|)} \quad \text{(Prop. 5.1)} \\
&= \exp(n[F(\widetilde{W}_{\mathtt{B}}) - F(W_{\sigma_n,\mathbf{y}_n}) - o(1)]).
\end{aligned}$$

By continuity of F, we then get

$$\mathop{\mathbb{P}}_{\sigma\sim \mathtt{s}_n}(W_{\sigma,\mathbf{y}_n^k} = \widetilde{W}_{\mathtt{B}}) \geq \exp n(F_\mu(\beta^k) - F_\mu(\beta) - 2\delta + o(1))$$

for all large enough n and small enough η (again depending on k, ε), with $\delta > 0$ the same as chosen above. Since $\beta_*^G\mu$ is a Markov chain, $F_\mu(\beta^k) = F_\mu(\beta)$ [5, Theorem 6.1].

Putting this all together, for any $k \in \mathbb{N}$, for all $\delta > 0$, we have

$$\mathop{\mathbb{E}}_{\sigma\sim \mathtt{s}_n} |\mathcal{X}_k(\sigma, \alpha\beta, \varepsilon \mid \mathbf{y}_n)| \geq \exp[n(F_\mu(\alpha^k \mid \beta^k) - 3\delta - o(1))]$$

for all large enough n and small enough $\varepsilon > 0$.

It follows that for any $k \in \mathbb{N}$,

$$\inf_{\varepsilon} \limsup_{n\to\infty} \frac{1}{n} \log \mathop{\mathbb{E}}_{\sigma\sim \mathtt{s}_n} |\{\mathbf{x} \in \mathtt{A}^n : (\mathbf{x}, \mathbf{y}_n) \in \Omega_k^*(\sigma, \alpha\beta, \varepsilon)\}| \geq F_\mu(T, \alpha^k \mid \beta^k).$$

Taking the limit as $k \to \infty$ gives the desired bound, using Corollary 3.3 and that the family of pseudometrics $\{d_k^* : k \in \mathbb{N}\}$ generates the weak* topology.

7. *Proof of Theorem B*

Let $W_n = W_{\sigma_n,\mathbf{y}_n^{m_n}}$, so that

$$\mathtt{s}_n = \mathtt{SBM}(\mathbf{y}_n, W_n).$$

Note that, by definition of $\mathtt{s}_n$,

$$\mathop{\mathbb{P}}_{\sigma\sim \mathtt{s}_n}(W_{\sigma,\mathbf{y}_n^{m_n}} = W_n) = 1.$$

LEMMA 7.1. *With W_n as just defined in terms of m_n, σ_n, and $\mathbf{y}_n$, we have*

$$\lim_{n\to\infty} F(W_n) = f_\mu(T, \beta).$$

Proof. The assumption in the theorem statement that $d_{m_n}^*(P_{\mathbf{y}_n}^{\sigma_n}, \beta_*^G\mu) = O(1/\log n)$ implies the existence of a constant C such that

$$d(W_n, W_{\beta^{m_n}}) \leq \frac{C}{\log n}.$$

By Lemma 3.1 we have

$$|F(W_{\sigma,\mathbf{y}^{m_n}}) - F(W_{\beta^{m_n}})| \leq 4r\left(\mathrm{H}\left(\frac{C}{\log n}\right) + \frac{C}{\log n}|\mathrm{B}(e, m_n)| \log|\mathtt{B}|\right) = o(1)$$

using that $m_n = o(\log\log n)$. Since m_n approaches infinity as n goes to infinity we have $f_\mu(T, \beta) = \lim_{n\to\infty} F(W_{\beta^{m_n}})$, so the result follows. □

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

LEMMA 7.2. *If* $m_n = o(\log\log n)$, *then for any* $k > 0$ *and* $\varepsilon > 0$ *we have* $|\mathtt{B}^{\mathrm{B}(e,m_n)}|^k = o(n^\varepsilon)$.

Proof. This is certainly true if $|\mathtt{B}| = 1$; assume therefore that $|\mathtt{B}| \geq 2$.

Our assumption $m_n = o(\log\log n)$ guarantees that

$$(2r-1)^{m_n} < \frac{r-1}{r}\frac{\varepsilon}{k\log|\mathtt{B}|}\log n$$

for all large enough n. Therefore

$$|\mathrm{B}(e,m_n)| = \frac{r(2r-1)^{m_n}-1}{r-1} < \frac{\varepsilon}{k\log|\mathtt{B}|}\log n.$$

This inequality can be rearranged to give

$$|\mathtt{B}^{\mathrm{B}(e,m_n)}|^k < n^\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, the result follows. □

In the remainder of this section we prove Theorem B by first proving the right-hand side is an upper bound for the left, then proving it is also lower bound.

7.1. *Upper bound.* Just as in the proof of the upper bound in Theorem A, for each $k \in \mathbb{N}$ and $\varepsilon > 0$ we have

$$\inf_{O \ni (\alpha\beta)^G_* \mu} \limsup_{n\to\infty} \frac{1}{n}\log \mathop{\mathbb{E}}_{\sigma\sim \mathtt{s}_n} |\{\mathbf{x} \in \mathtt{A}^n : (\mathbf{x}, \mathbf{y}_n) \in \Omega(\sigma, O)\}| \leq \limsup_{n\to\infty} \frac{1}{n}\log \mathcal{E}_k(n,\varepsilon),$$

where

$$\begin{aligned}\mathcal{E}_k(n,\varepsilon) &:= \mathop{\mathbb{E}}_{\sigma\sim \mathtt{s}_n} |\{\mathbf{X} \in (\mathtt{A}^{\mathrm{B}(e,k)})^n : (\mathbf{X}, \mathbf{y}_n^k) \in \Omega_0^*(\sigma, (\alpha\beta)^k, \varepsilon)\}| \\ &= \mathop{\mathbb{E}}_{\sigma\sim \mathtt{s}_n} |\{\mathbf{X} \in (\mathtt{A}^{\mathrm{B}(e,k)})^n : d(W_{\sigma,(\mathbf{X},\mathbf{y}_n^k)}, W_{(\alpha\beta)^k}) < \varepsilon)\}|.\end{aligned}$$

We assume that n is large enough that $m_n \geq k$.

Since $\mathtt{s}_n$ is $\mathtt{SBM}(\sigma_n, \mathbf{y}_n, m_n)$ rather than $\mathtt{SBM}(\sigma_n, \mathbf{y}_n, k)$, we cannot apply Proposition 5.2 directly to this expression. We get around this as follows. Let

$$\mathcal{W}_n(m,m') := \{W_{\sigma,(\mathbf{X},\mathbf{y}^{m'})} : \sigma \in \mathrm{Hom}(G, \mathrm{Sym}(n)), \mathbf{X} \in (\mathtt{A}^{\mathrm{B}(e,m)})^n, \mathbf{y} \in \mathtt{B}^n\}.$$

All elements of this set are denominator-n $\mathtt{A}^{\mathrm{B}(e,m)} \times \mathtt{B}^{\mathrm{B}(e,m')}$-weights; we avoid the question of exactly which weights are in this set, but call such weights *attainable*. For $k \leq m$ and $k' \leq m'$ let

$$\mathcal{W}_n(m,m';\alpha\beta,k,k';\varepsilon) = \{W \in \mathcal{W}_n(m,m') : d(\pi_{k,k'}W, W_{\alpha^k\beta^{k'}}) < \varepsilon\}$$

denote the set of such weights whose appropriate marginal is within ε of the $(\mathtt{A}^{\mathrm{B}(e,k)} \times \mathtt{B}^{\mathrm{B}(e,k')})$-weight $W_{\alpha^k\beta^{k'}}$. For now we take $m = k = k'$, but we will need more generality below. Then

$$\mathcal{E}_k(n,\varepsilon) = \mathop{\mathbb{E}}_{\sigma\sim \mathtt{s}_n} \sum_{W\in\mathcal{W}_n(k,m_n;\alpha\beta,k,k;\varepsilon)} |\{\mathbf{X} \in (\mathtt{A}^{\mathrm{B}(e,k)})^n : W_{\sigma,(\mathbf{X},\mathbf{y}_n^{m_n})} = W\}|,$$

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

so we can apply Proposition 5.2 to get

$$\mathcal{E}_k(n,\varepsilon) \le (9n)^{r|\mathrm{B}^{\mathrm{B}(e,m_n)}|^2(|\mathrm{A}^{\mathrm{B}(e,k)}|+1)} \sum_{W\in\mathcal{W}_n(k,m_n;\alpha\beta,k,k;\varepsilon)} e^{n(F(W)-F(\pi_{\mathrm{B}}W))}\mathbf{1}_{\{\pi_{\mathrm{B}}W=W_n\}}.$$

By Lemma 7.2 we have $(9n)^{r|\mathrm{B}^{\mathrm{B}(e,m_n)}|^2(|\mathrm{A}^{\mathrm{B}(e,k)}|+1)} \le e^{o_{n\to\infty}(n)}$. Using this and Lemma 7.1, we have

$$\mathcal{E}_k(n,\varepsilon) \le \sum_{W\in\mathcal{W}_n(k,m_n;\alpha\beta,k,k;\varepsilon)} e^{n(F(W)-f(T,\beta)+o_{n\to\infty}(1))}\mathbf{1}_{\{\pi_{\mathrm{B}}W=W_n\}},$$

where the little o is uniform over all terms in the sum. Here we use the assumption that $f_\mu(T,\beta)$ is finite.

By definition of $\mathcal{W}_n(k,m_n)$, for any $W\in\mathcal{W}_n(k,m_n;\alpha\beta,k,k;\varepsilon)$ we can pick $\sigma\in\mathrm{Hom}(G,\mathrm{Sym}(n))$, $\mathbf{X}\in(\mathrm{A}^{\mathrm{B}(e,k)})^n$, and $\mathbf{y}\in\mathrm{B}^n$ so that $W=W_{\sigma,(\mathbf{X},\mathbf{y}^{m_n})}$. Then since $\mathbf{X}\mathbf{y}^{m_n}$ is a splitting of $\mathbf{X}\mathbf{y}^k$, by Lemma 3.2 we have

$$F(W)=F_{\mathrm{Unif}([n])}(\sigma,\mathbf{X}\mathbf{y}^{m_n})\le F_{\mathrm{Unif}([n])}(\sigma,\mathbf{X}\mathbf{y}^k)=F(\pi_{k,k}W),$$

where here $F_{\mathrm{Unif}([n])}(\sigma,\mathbf{X}\mathbf{y}^{m_n})$ is F of the observable $\mathbf{X}\mathbf{y}^{m_n}$ on the measure-preserving system $([n],\mathrm{Unif}([n]),\sigma)$ (we shift to this notation from weights in order to apply the splitting lemma). By continuity of F, for all small enough ε (depending on k) we have

$$F(\pi_{k,k}W)\le F(W_{(\alpha\beta)^k})+\delta=F_\mu(T,(\alpha\beta)^k)+\delta.$$

Along with the above, this implies that

$$\mathcal{E}_k(n,\varepsilon)\le e^{n(F(T,(\alpha\beta)^k)-f(T,\beta)+o_n(1)+\delta)}\sum_{W\in\mathcal{W}_n(k,m_n;\alpha\beta,k,k;\varepsilon)}\mathbf{1}_{\{\pi_{\mathrm{B}}W=W_n\}}.$$

Bounding all terms in the sum by 1, we get

$$\mathcal{E}_k(n,\varepsilon)\le e^{n(F(T,(\alpha\beta)^k)-f_\mu(T,\beta)+o_n(1)+\delta)}\,|\mathcal{W}_n(k,m_n;\alpha\beta,k,k;\varepsilon)|.$$

Using Lemma 7.2, we have

$$|\mathcal{W}_n(k,m_n;\alpha\beta,k,k;\varepsilon)|\le|\mathcal{W}_n(k,m_n)|\le n^{r|\mathrm{A}^{\mathrm{B}(e,k)}\times\mathrm{B}^{\mathrm{B}(e,m_n)}|^2}\le e^{o_{n\to\infty}(n)},$$

so this implies

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathcal{E}_k(n,\varepsilon)\le F_\mu(T,(\alpha\beta)^k)-f_\mu(T,\beta)+\delta.$$

Taking the infimum over ε and k, and using the chain rule for f (Corollary 3.5, again using the assumption that $f_\mu(T,\beta)$ is finite), gives

$$\inf_{\varepsilon,k}\limsup_{n\to\infty}\frac{1}{n}\log\mathcal{E}_k(n,\varepsilon)\le f_\mu(T,\alpha\beta)-f_\mu(T,\beta)=f_\mu(T,\alpha\mid\beta).$$

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

Since

$$\inf_{\mathcal{O}\ni(\alpha\beta)^G_*\mu} \limsup_{n\to\infty} \frac{1}{n} \log \mathop{\mathbb{E}}_{\sigma\sim \mathtt{s}_n} |\{\mathbf{x}\in \mathtt{A}^n : (\mathbf{x},\mathbf{y}_n)\in \Omega(\sigma,\mathcal{O})\}|$$
$$\leq \inf_{\varepsilon} \limsup_{n\to\infty} \frac{1}{n} \log \mathcal{E}_k(n,\varepsilon),$$

for every k, this completes the upper bound.

7.2. *Lower bound.* In this section we denote

$$\mathcal{X}_{k_1,k_2}(\sigma,\alpha\beta,\varepsilon \mid \mathbf{y}) := \{\mathbf{X}\in (\mathtt{A}^{\mathrm{B}(e,k_1)})^n : (\mathbf{X},\mathbf{y}^{k_2})\in \Omega_0^*(\sigma,\alpha^{k_1}\beta^{k_2},\varepsilon)\},$$
$$\Omega_k^*(\sigma,\alpha\beta,\varepsilon \mid \mathbf{y}) := \{\mathbf{x}\in \mathtt{A}^n : (\mathbf{x},\mathbf{y})\in \Omega_k^*(\sigma,\alpha\beta,\varepsilon)\}$$

(note the dependence on n is implicitly specified by $\sigma\in \mathrm{Hom}(G,\mathrm{Sym}(n))$ and $\mathbf{y}\in \mathtt{B}^n$), and with $\Sigma = \{\mathtt{s}_n\}_{n=1}^\infty$,

$$\begin{aligned}\mathrm{h}_{\Sigma,\mu}(T,\alpha \mid \beta : k,\varepsilon) &:= \limsup_{n\to\infty} \frac{1}{n} \log \mathop{\mathbb{E}}_{\sigma\sim \mathtt{s}_n} |\{\mathbf{x}\in \mathtt{A}^n : (\mathbf{x},\mathbf{y})\in \Omega_k^*(\sigma,\alpha\beta,\varepsilon)\}| \\ &= \limsup_{n\to\infty} \frac{1}{n} \log \mathop{\mathbb{E}}_{\sigma\sim \mathtt{s}_n} |\Omega_k^*(\sigma,\alpha\beta,\varepsilon \mid \mathbf{y})|.\end{aligned}$$

The following two claims are used to relate the sizes of the sets defined above.

CLAIM 1. *Let $k\leq \min(k_1,k_2)$. For any $\sigma,\mathbf{y}$, we have*

$$\pi_e[\mathcal{X}_{k_1,k_2}(\sigma,\alpha\beta,\varepsilon \mid \mathbf{y})]\subseteq \Omega_k^*(\sigma,\alpha\beta,c\varepsilon \mid \mathbf{y})$$

where $c = 1+|\mathrm{B}(e,k)|$.

Proof. If $(\mathbf{X},\mathbf{y}^{k_2})\in \Omega_0^*(\sigma,\alpha^{k_1}\beta^{k_2},\varepsilon)$, then

$$\pi_{k,k}(\mathbf{X},\mathbf{y}^{k_2})\in \Omega_0^*(\sigma,(\alpha\beta)^k,\varepsilon);$$

this follows from the fact that total variation distance is non-increasing under pushforwards. Applying Proposition 2.5, we get

$$(\pi_e\mathbf{X},\mathbf{y}) = \pi_e(\pi_{k,k}(\mathbf{X},\mathbf{y}^{k_2}))\in \Omega_k^*(\sigma,\alpha\beta,c\varepsilon). \qquad \square$$

CLAIM 2. *Fix $\sigma,\mathbf{y}$, and $k\leq \min(k_1,k_2)$. As established in the previous claim, we can consider π_e as a map from $\mathcal{X}_{k_1,k_2}(\sigma,\alpha\beta,\varepsilon \mid \mathbf{y})$ to $\Omega_k^*(\sigma,\alpha\beta,c\varepsilon \mid \mathbf{y})$. There are constants C,d independent of n such that π_e is at most $C\exp(nd\varepsilon + n\mathrm{H}(2|\mathrm{B}(e,k)|\varepsilon))$-to-one.*

Proof. If $\Omega_k^*(\sigma,\alpha\beta,c\varepsilon \mid \mathbf{y})$ is empty, then the claim is vacuously true. Otherwise, fix $\mathbf{x}\in \Omega_k^*(\sigma,\alpha\beta,c\varepsilon \mid \mathbf{y})$. If $\mathbf{X}\in \pi_e^{-1}\{\mathbf{x}\}$, then $\pi_e(\mathbf{X},\mathbf{y}^k) = (\mathbf{x},\mathbf{y})$. Claim 3 in the proof of Proposition 3.2 of [2] gives an upper bound of the desired form for the number of such pairs $(\mathbf{X},\mathbf{y}^k)$, and therefore the number of such $\mathbf{X}$. $\square$

Claim 2 implies that

$$|\mathcal{X}_{k_1,k_2}(\sigma,\alpha\beta,\varepsilon \mid \mathbf{y})|\leq C\exp(nd\varepsilon + n\mathrm{H}(2|\mathrm{B}(e,k)|\varepsilon))\cdot |\Omega_k^*(\sigma,\alpha\beta,c\varepsilon \mid \mathbf{y})|, \qquad (7.1)$$

where C,d are independent of n.

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

We now find a lower bound for the expectation of $|\mathcal{X}|$. Fix $k_1, k_2 \in \mathbb{N}$, and suppose n is large enough that $m_n \geq \max(k_1, k_2)$. Using Proposition 5.2 and Lemma 7.2, we have

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\sigma\sim \mathtt{s}_n} |\mathcal{X}_{k_1,k_2}(\sigma, \alpha\beta, \varepsilon \mid \mathbf{y}_n)| \\
&\quad = \sum_{W\in\mathcal{W}_n(k_1,m_n;\alpha\beta,k_1,k_2;\varepsilon)} \mathop{\mathbb{E}}_{\sigma\sim \mathtt{s}_n} |\{\mathbf{X} \in (\mathtt{A}^{\mathrm{B}(e,k_1)})^n : W_{\sigma,(\mathbf{X},\mathbf{y}_n^{m_n})} = W\}| \\
&\quad \geq \sum_{W\in\mathcal{W}_n(k_1,m_n;\alpha\beta,k_1,k_2;\varepsilon)} \exp[n(F(W) - F(\pi_{\mathrm{B}} W) - o_n(1))]\mathbf{1}_{\{\pi_{\mathrm{B}} W = W_{\sigma,\mathbf{y}_n^{m_n}}\}} \\
&\quad \geq \inf_{W\in\mathcal{W}_n(k_1,m_n;\alpha\beta,k_1,k_2;\varepsilon)} \exp[n(F(W) - F(\pi_{\mathrm{B}} W) - o_n(1))] \\
&\qquad \times \sum_{W\in\mathcal{W}_n(k_1,m_n;\alpha\beta,k_1,k_2;\varepsilon)} \mathbf{1}_{\{\pi_{\mathrm{B}} W = W_{\sigma,\mathbf{y}_n^{m_n}}\}}.
\end{aligned}$$

We bound the infimum below as follows. Given any $W \in \mathcal{W}_n(k_1, m_n; \alpha\beta, k_1, k_2; \varepsilon)$, we can let $\mathbf{X}, \mathbf{y}, \sigma$ be such that $W = W_{\sigma,(\mathbf{X},\mathbf{y}^{m_n})}$. Then by Lemma 3.2 and continuity of F,

$$\begin{aligned}
F(W) - F(\pi_{\mathrm{B}} W) &= F(\sigma, \mathbf{X}|\mathbf{y}^{m_n}) \\
&\geq F(\sigma, \mathbf{X}|\mathbf{y}^{k_2}) \\
&= F(\pi_{k_1,k_2} W) - F(\pi_{\mathrm{B}}\pi_{k_1,k_2} W) \\
&\geq F_\mu(T, \alpha^{k_1}|\beta^{k_2}) - \delta
\end{aligned}$$

for any $\delta > 0$ for all small enough ε (with 'small enough' dependent only on k_1, k_2). This implies that the infimum is bounded below by

$$\exp[n(F_\mu(T, \alpha^{k_1}|\beta^{k_2}) - o_n(1) - \delta)].$$

We bound the sum below by first rewriting it as

$$|\{W \in \mathcal{W}_n(k_1, m_n; \alpha\beta, k_1, k_2; \varepsilon) : \pi_{\mathrm{B}} W = W_{\sigma,\mathbf{y}_n^{m_n}}\}|.$$

The following claim, then, implies that the sum is bounded below by 1.

CLAIM 3. *For all large enough n,*

$$\{W \in \mathcal{W}_n(k_1, m_n; \alpha\beta, k_1, k_2; \varepsilon) : \pi_{\mathrm{B}} W = W_{\sigma,\mathbf{y}_n^{m_n}}\} \neq \varnothing.$$

Proof. By Lemma 2.3, if

$$n > 680|\mathtt{A}^{\mathrm{B}(e,k_1)} \times \mathtt{B}^{\mathrm{B}(e,m_n)}|^2 r/\varepsilon$$

and $d(W_{\sigma,\mathbf{y}_n^{m_n}}, W_{\beta^{m_n}}) < \varepsilon/530r$ then there is a $(\mathtt{A}^{\mathrm{B}(e,k_1)} \times \mathtt{B}^{\mathrm{B}(e,m_n)})$-weight W with $\pi_{\mathrm{B}} W = W_{\sigma,\mathbf{y}_n^{m_n}}$ and $d(W, W_{\alpha^{k_1}\beta^{m_n}}) < \varepsilon$. By definition of $\mathtt{s}_n$ and Lemma 7.2, both conditions are met for all large enough n.

The claim will follow if we show that W is attainable.

With W as chosen above, by Proposition 2.4 we can choose $\tilde{\sigma} \in \mathrm{Hom}(G, \mathrm{Sym}(n))$, $\tilde{\mathbf{X}} \in (\mathtt{A}^{\mathrm{B}(e,k_1)})^n$, and $\tilde{\mathbf{Y}} \in (\mathtt{B}^{\mathrm{B}(e,m_n)})^n$ such that $W = W_{\tilde{\sigma},(\tilde{\mathbf{X}},\tilde{\mathbf{Y}})}$.

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

Let $\tilde{\mathbf{y}} = \pi_e \tilde{\mathbf{Y}} \in \mathtt{B}^n$. To complete the proof we show that $\tilde{\mathbf{y}}^{m_n} = \tilde{\mathbf{Y}}$, that is,

$$\tilde{\mathbf{y}}(\tilde{\sigma}(g)i) = (\tilde{\mathbf{Y}}(i))_g$$

for all $i \in [n]$ and $g \in \mathrm{B}(e, m_n)$. We prove this by induction on the word length $|g|$.

The base case $|g| = 0$ (that is, $g = e$) follows immediately from the definition of $\tilde{\mathbf{y}}$.

For the inductive step, write $g = ht$ with $|h| = |g| - 1$ and $t \in \{s_1^{\pm 1}, \ldots, s_r^{\pm 1}\}$. Then, assuming the result holds for h,

$$\tilde{\mathbf{y}}(\tilde{\sigma}(g)i) = \tilde{\mathbf{y}}(\tilde{\sigma}(h)\tilde{\sigma}(t)i) = (\tilde{\mathbf{Y}}(\tilde{\sigma}(t)i))_h.$$

Now since $W_{\tilde{\sigma},\tilde{\mathbf{Y}}} = W_{\sigma_n,\mathbf{y}_n^{m_n}}$, we can pick $j \in [n]$ such that

$$\tilde{\mathbf{Y}}(i) = \mathbf{y}_n^{m_n}(j) \quad \text{and} \quad \tilde{\mathbf{Y}}(\tilde{\sigma}(t)i) = \mathbf{y}_n^{m_n}(\sigma(t)j).$$

This implies

$$(\tilde{\mathbf{Y}}(\tilde{\sigma}(t)i))_h = (\mathbf{y}_n^{m_n}(\sigma(t)j))_h = \mathbf{y}_n(\sigma(g)j) = (\mathbf{y}_n^{m_n}(j))_g = (\tilde{\mathbf{Y}}(i))_g. \qquad \square$$

Hence, for all large enough n, we have

$$\mathop{\mathbb{E}}_{\sigma \sim \mathtt{S}_n} |\mathcal{X}_{k_1,k_2}(\sigma, \alpha\beta, \varepsilon \mid \mathbf{y}_n)| \geq \exp[n(F_\mu(T, \alpha^{k_1} \mid \beta^{k_2}) - o_n(1) - \delta)],$$

and therefore

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathop{\mathbb{E}}_{\sigma \sim \mathtt{S}_n} |\mathcal{X}_{k_1,k_2}(\sigma, \alpha\beta, \varepsilon \mid \mathbf{y}_n)| \geq F_\mu(T, \alpha^{k_1} \mid \beta^{k_2}) - \delta.$$

Combining this lower bound with equation (7.1) and the definition of $\mathrm{h}_{\Sigma,\mu}(T, \alpha \mid \beta : k, c\varepsilon)$, we get

$$d\varepsilon + \mathrm{H}(2|\mathrm{B}(e,k)|\varepsilon) + \mathrm{h}_{\Sigma,\mu}(T, \alpha \mid \beta : k, c\varepsilon) \geq F_\mu(T, \alpha^{k_1} \mid \beta^{k_2}) - \delta.$$

Taking the infimum in ε then letting δ go to zero gives

$$\inf_{\varepsilon} \limsup_{n\to\infty} \frac{1}{n} \log \mathop{\mathbb{E}}_{\sigma \sim \mathtt{S}_n} |\{\mathbf{x} \in \mathtt{A}^n : (\mathbf{x}, \mathbf{y}_n) \in \Omega_k^*(\sigma, \alpha\beta, \varepsilon)\}| \geq F_\mu(T, \alpha^{k_1} \mid \beta^{k_2})$$

for $k \leq \min(k_1, k_2)$. First take $k_2 \to \infty$, then $k_1 \to \infty$, then take the infimum over k. We get

$$\begin{aligned} f_\mu(T, \alpha \mid \beta) &\leq \inf_{\varepsilon,k} \limsup_{n\to\infty} \frac{1}{n} \log \mathop{\mathbb{E}}_{\sigma \sim \mathtt{S}_n} |\{\mathbf{x} \in \mathtt{A}^n : (\mathbf{x}, \mathbf{y}_n) \in \Omega_k^*(\sigma, \alpha\beta, \varepsilon)\}| \\ &= \inf_{O \ni (\alpha\beta)^G_* \mu} \limsup_{n\to\infty} \frac{1}{n} \log \mathop{\mathbb{E}}_{\sigma \sim \mathtt{S}_n} |\{\mathbf{x} \in \mathtt{A}^n : (\mathbf{x}, \mathbf{y}_n) \in \Omega(\sigma, O)\}| \end{aligned}$$

where the last line follows because the collection of pseudometrics $\{d_k^* : k \in \mathbb{N}\}$ generates the weak* topology on $\mathrm{Prob}((\mathtt{A} \times \mathtt{B})^G)$.

8. *Proof of Theorem C*

By analogy with sofic entropy, we denote $\Sigma := \{\mathtt{s}_n\}_{n=1}^\infty$ and denote the left-hand side of the formula in the theorem statement by $\mathrm{h}_{\Sigma,\mu}(T, \alpha)$.

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

Endow $\mathrm{Prob}(\mathtt{A}^G)$ with the metric

$$d(\lambda, \nu) := \sum_{r=1}^{\infty} 2^{-r} d^{\mathrm{B}(e,r)}(\lambda, \nu).$$

Note that this induces the weak* topology (where $\mathtt{A}$ is given the discrete topology and $\mathtt{A}^G$ the product topology).

Writing $\mu_{\mathtt{A}} = \alpha_*^G \mu \in \mathrm{Prob}(\mathtt{A}^G)$, we then have

$$\mathrm{h}_{\Sigma,\mu}(T, \alpha) = \inf_{\varepsilon>0} \limsup_{n\to\infty} \frac{1}{n} \log \mathop{\mathbb{E}}_{\sigma\sim \mathtt{s}_n} |\{\mathbf{x} \in \mathtt{A}^n : d(P_{\mathbf{x}}^{\sigma}, \mu_{\mathtt{A}}) < \varepsilon\}|.$$

We will similarly denote $\mu_{\mathtt{B}} = \beta_*^G \mu \in \mathrm{Prob}(\mathtt{B}^G)$.

8.1. *Lower bound.* Let $\lambda \in \mathrm{Prob}((\mathtt{A} \times \mathtt{B})^G)$ be any joining of (the shift systems with respective measures) $\mu_{\mathtt{A}}$ and $\mu_{\mathtt{B}}$. Then, for any $\mathbf{x} \in \mathtt{A}^n$ and $\mathbf{y} \in \mathtt{B}^n$, we have

$$d(P_{\mathbf{x}}^{\sigma}, \mu_{\mathtt{A}}) \le d(P_{(\mathbf{x},\mathbf{y})}^{\sigma}, \lambda),$$

where d is defined on $\mathrm{Prob}((\mathtt{A} \times \mathtt{B})^G)$ analogously to the definition given on $\mathrm{Prob}(\mathtt{A}^G)$ above. This inequality holds because total variation distance is non-increasing under pushforwards. Consequently,

$$\mathrm{h}_{\Sigma,\mu}(T, \alpha) \ge \inf_{\varepsilon>0} \limsup_{n\to\infty} \frac{1}{n} \log \mathop{\mathbb{E}}_{\sigma\sim \mathtt{s}_n} |\{\mathbf{x} \in \mathtt{A}^n : d(P_{(\mathbf{x},\mathbf{y}_n)}^{\sigma}, \lambda) < \varepsilon\}| = f_{\lambda}(S, \mathtt{a} \mid \mathtt{b}).$$

Taking the supremum over joinings λ gives the lower bound.

8.2. *Upper bound.* For $\varepsilon > 0$, let

$$\mathtt{J}_{\varepsilon} := \{\lambda \in \mathrm{Prob}^S((\mathtt{A} \times \mathtt{B})^G) : d(\mathtt{a}_*^G \lambda, \mu_{\mathtt{A}}) < \varepsilon \text{ and } d(\mathtt{b}_*^G \lambda, \mu_{\mathtt{B}}) < \varepsilon\}$$

be the set of shift-invariant 'approximate joinings' of $\mu_{\mathtt{A}}$ and $\mu_{\mathtt{B}}$. Since $\mathrm{Prob}((\mathtt{A} \times \mathtt{B})^G)$ is compact, for each $\varepsilon > 0$ there exist $\lambda_1, \ldots, \lambda_m \in \mathtt{J}_{\varepsilon}$ such that

$$\mathtt{J}_{\varepsilon} \subseteq \bigcup_{i=1}^{m} \mathrm{B}(\lambda_i, \varepsilon).$$

By definition of $\mathtt{s}_n$ we have $\mathbb{P}_{\sigma\sim \mathtt{s}_n}(d(P_{\mathbf{y}_n}^{\sigma}, \mu_{\mathtt{B}}) < \varepsilon) = 1$ for all large enough n. Therefore,

$$\begin{aligned}
\mathrm{h}_{\Sigma,\mu}(T, \alpha) &= \inf_{\varepsilon} \limsup_{n\to\infty} \frac{1}{n} \log \mathop{\mathbb{E}}_{\sigma\sim \mathtt{s}_n} |\{\mathbf{x} \in \mathtt{A}^n : P_{(\mathbf{x},\mathbf{y}_n)}^{\sigma} \in \mathtt{J}_{\varepsilon}\}| \\
&\le \inf_{\varepsilon} \limsup_{n\to\infty} \frac{1}{n} \log \sum_{i=1}^{m} \mathop{\mathbb{E}}_{\sigma\sim \mathtt{s}_n} |\{\mathbf{x} \in \mathtt{A}^n : P_{(\mathbf{x},\mathbf{y}_n)}^{\sigma} \in \mathrm{B}(\lambda_i, \varepsilon)\}| \\
&= \inf_{\varepsilon} \max_{1\le i\le m} \limsup_{n\to\infty} \frac{1}{n} \log \mathop{\mathbb{E}}_{\sigma\sim \mathtt{s}_n} |\{\mathbf{x} \in \mathtt{A}^n : P_{(\mathbf{x},\mathbf{y}_n)}^{\sigma} \in \mathrm{B}(\lambda_i, \varepsilon)\}| \\
&\le \inf_{\varepsilon} \sup_{\lambda\in \mathtt{J}_{\varepsilon}} \limsup_{n\to\infty} \frac{1}{n} \log \mathop{\mathbb{E}}_{\sigma\sim \mathtt{s}_n} |\{\mathbf{x} \in \mathtt{A}^n : P_{(\mathbf{x},\mathbf{y}_n)}^{\sigma} \in \mathrm{B}(\lambda, \varepsilon)\}|.
\end{aligned}$$

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

Note that the entire expression in the infimum is decreasing as $\varepsilon \to 0$, so we may replace the infimum with a limit. Rather than taking a continuous limit we write

$$\mathrm{h}_{\Sigma,\mu}(T,\alpha) \leq \lim_{m\to\infty} \sup_{\lambda\in \mathsf{J}_{1/m}} \limsup_{n\to\infty} \frac{1}{n} \log \mathop{\mathbb{E}}_{\sigma\sim \mathtt{s}_n} |\{\mathbf{x}\in \mathtt{A}^n : P^{\sigma}_{(\mathbf{x},\mathbf{y}_n)} \in \mathrm{B}(\lambda, 1/m)\}|.$$

For each m, pick $\lambda_m \in \mathsf{J}_{1/m}$ to get within $1/m$ of the supremum. Then the right-hand side is equal to

$$\lim_{m\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \mathop{\mathbb{E}}_{\sigma\sim \mathtt{s}_n} |\{\mathbf{x}\in \mathtt{A}^n : P^{\sigma}_{(\mathbf{x},\mathbf{y}_n)} \in \mathrm{B}(\lambda_m, 1/m)\}|. \tag{$*$}$$

Let λ_{m_j} be a subsequence with weak* limit λ_0. By weak* continuity of pushforwards under projection we have $\lambda_0 \in \mathsf{J}(\mu_{\mathtt{A}}, \mu_{\mathtt{B}})$. Now, for any $\delta > 0$, for all large enough j we have both $1/m_j < \delta/2$ and $d(\lambda_{m_j}, \lambda_0) < \delta/2$, so by the triangle inequality

$$\mathrm{B}(\lambda_{m_j}, 1/m_j) \subseteq \mathrm{B}(\lambda_0, \delta).$$

It follows that the expression in $(*)$, and hence $h_{\Sigma}(\alpha)$, is bounded above by

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathop{\mathbb{E}}_{\sigma\sim \mathtt{s}_n} |\{\mathbf{x}\in \mathtt{A}^n : P^{\sigma}_{(\mathbf{x},\mathbf{y}_n)} \in \mathrm{B}(\lambda_0, \delta)\}|.$$

Taking the infimum over δ shows that

$$h_{\Sigma}(\mu,\alpha) \leq f_{\lambda_0}(S, \mathtt{a} \mid \mathtt{b}) \leq \sup_{\lambda\in \mathsf{J}(\mu_{\mathtt{A}},\mu_{\mathtt{B}})} f_{\lambda}(S, \mathtt{a} \mid \mathtt{b}).$$

9. *Proof of Proposition 1.1*

All sequences of interest are of the form

$$\mathtt{s}_n = \mathtt{SBM}(\sigma_n, \mathbf{y}_n, m_n) = \mathrm{Unif}(\{\sigma \in \mathrm{Hom}(G, \mathrm{Sym}(n)) : W_{\sigma,\mathbf{y}_n^{m_n}} = W_n\})$$

with $\mathbf{y}_n \in \mathtt{B}^n$, $\sigma_n \in \mathrm{Sym}(n)$, $m_n = o(\log\log n)$, and where W_n is the $\mathtt{B}^{\mathrm{B}(e,m_n)}$-weight $W_{\sigma_n,\mathbf{y}_n^{m_n}}$. In the case of Theorem A we simply have $m_n = 0$ for all n.

The theorem will follow from the following lemma.

LEMMA 9.1. *Let ζ_n denote the uniform measure on* $\mathrm{Hom}(G, \mathrm{Sym}(n))$. *Then, for any finite $D \subset G$ and $\delta > 0$, there exists $\varepsilon > 0$ such that*

$$\mathop{\mathbb{P}}_{\sigma\sim\zeta_n} (\sigma \textit{ is } (D,\delta)\textit{-sofic}) \geq 1 - n^{-\varepsilon n}$$

for all large enough n.

This can be proven by making superficial changes to the proof of the similar result [1, Lemma 3.1].

To prove Proposition 1.1, it now suffices to show that, for any $\varepsilon > 0$,

$$\mathop{\mathbb{P}}_{\sigma\sim\zeta_n} (W_{\sigma,\mathbf{y}_n^{m_n}} = W_n) \geq n^{-\varepsilon n}$$

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

for all large enough n. To do this, first note that the left-hand side here depends only on the vector $p_{\mathbf{y}_n} \in \mathrm{Prob}(\mathtt{B})$ of letter frequencies. Therefore,

$$\begin{aligned}\mathop{\mathbb{P}}_{\sigma\sim\zeta_n}(\text{there exists } \mathbf{y}\in \mathtt{B}^n \text{ s.t. } W_{\sigma,\mathbf{y}^{m_n}} = W_n) &\le \sum_{\mathbf{y}:p_{\mathbf{y}}=p_{\mathbf{y}_n}} \mathop{\mathbb{P}}_{\sigma\sim\zeta_n}(W_{\sigma,\mathbf{y}^{m_n}}=W_n)\\ &= \exp\{n\mathrm{H}(p_{\mathbf{y}_n})+o(n)\}\mathop{\mathbb{P}}_{\sigma\sim\zeta_n}(W_{\sigma,\mathbf{y}_n^{m_n}}=W_n).\end{aligned}$$

But by Proposition 2.5, if $\sigma \in \mathrm{Hom}(G,\mathrm{Sym}(n))$ and $\mathbf{Y}\in(\mathtt{B}^{\mathrm{B}(e,m_n)})^n$ are such that $W_{\sigma,\mathbf{Y}} = W_n = W_{\sigma_n,\mathbf{y}_n^{m_n}}$, then the projection $\mathbf{Y}_e \in \mathtt{B}^n$ satisfies $(\mathbf{Y}_e)^{m_n}=\mathbf{Y}$. Therefore, for each σ,

$$|\{\mathbf{Y}\in(\mathtt{B}^{\mathrm{B}(e,m_n)})^n : W_{\sigma,\mathbf{Y}}=W_n\}| = |\{\mathbf{y}\in\mathtt{B}^n : W_{\sigma,\mathbf{y}^{m_n}}=W_n\}|.$$

Hence,

$$\begin{aligned}\mathop{\mathbb{E}}_{\sigma\sim\zeta_n}|\{\mathbf{Y}\in(\mathtt{B}^{\mathrm{B}(e,m_n)})^n : W_{\sigma,\mathbf{Y}}=W_n\}| &= \mathop{\mathbb{E}}_{\sigma\sim\zeta_n}|\{\mathbf{y}\in\mathtt{B}^n : W_{\sigma,\mathbf{y}^{m_n}}=W_n\}|\\ &\le |\mathtt{B}|^n \mathop{\mathbb{P}}_{\sigma\sim\zeta_n}(\text{there exists } \mathbf{y}\in\mathtt{B}^n \text{ s.t. } W_{\sigma,\mathbf{y}^{m_n}}=W_n).\end{aligned}$$

Combining these last few statements, we see that

$$\mathop{\mathbb{P}}_{\sigma\sim\zeta_n}(W_{\sigma,\mathbf{y}_n^{m_n}}=W_n) \ge \exp\{-2n\log|\mathtt{B}|+o(n)\}\mathop{\mathbb{E}}_{\sigma\sim\zeta_n}|\{\mathbf{Y}\in(\mathtt{B}^{\mathrm{B}(e,m_n)})^n : W_{\sigma,\mathbf{Y}}=W_n\}|.$$

We can ignore the first factor here since it only decays exponentially fast. By Proposition 5.1,

$$\mathop{\mathbb{E}}_{\sigma\sim\zeta_n}|\{\mathbf{Y}\in(\mathtt{B}^{\mathrm{B}(e,m_n)})^n : W_{\sigma,\mathbf{Y}}=W_n\}| = \frac{Z_n(W_n)}{(n!)^r} \ge (3\sqrt{n})^{-r|\mathtt{B}^{\mathrm{B}(e,m_n)}|^2} e^{F(W_n)n} n^{(1-r)/2}.$$

The third factor is clearly not a problem and can also be ignored. For the first factor,

$$\frac{1}{n\log n}\log(3\sqrt{n})^{-r|\mathtt{B}^{\mathrm{B}(e,m_n)}|^2} = -r\frac{|\mathtt{B}^{\mathrm{B}(e,m_n)}|^2}{n}\frac{\log 3\sqrt{n}}{\log n}\to 0 \quad \text{as } n\to\infty$$

using Lemma 7.2. For the second factor, first note that by definition of $F(W_n)$ we have

$$\begin{aligned}F(W_n) &= (1-2r)\mathrm{H}(W_n(\cdot)) + \sum_{i\in[r]}\mathrm{H}(W_n(\cdot,\cdot;i))\\ &\ge -2r\mathrm{H}(W_n(\cdot))\\ &\ge -2r\log|\mathtt{B}^{\mathrm{B}(e,m_n)}|.\end{aligned}$$

So

$$\frac{1}{n\log n}\log e^{F(W_n)n} = \frac{F(W_n)}{\log n} \ge -2r\frac{\log|\mathtt{B}^{\mathrm{B}(e,m_n)}|}{\log n}\to 0 \quad \text{as } n\to\infty,$$

again using Lemma 7.2. This implies that for every $\varepsilon>0$ we have

$$(3\sqrt{n})^{-r|\mathtt{B}^{\mathrm{B}(e,m_n)}|^2} e^{F(W_n)n} \ge n^{-\varepsilon n}$$

for all large enough n, which implies the result.

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

10. *Proof of Lemma 2.3*

We show how to construct a denominator-n weight $W_{\mathtt{AB}}$ that has a given $\mathtt{B}$-marginal $W_{\mathtt{B}}$ and is close to a given $(\mathtt{A} \times \mathtt{B})$-weight W whose $\mathtt{B}$-marginal $\pi_{\mathtt{B}} W$ is close to $W_{\mathtt{B}}$. As in the theorem statement, we assume

$$d(\pi_{\mathtt{B}} W, W_{\mathtt{B}}) < \delta.$$

To minimize the appearance of factors of $\frac{1}{2}$, in this section we work with the ℓ^1 distance on weights, which is twice the distance defined above. Therefore the previous assumption becomes

$$d_1(\pi_{\mathtt{B}} W, W_{\mathtt{B}}) = \sum_{i \in [r]} \sum_{b, b' \in \mathtt{B}} |\pi_{\mathtt{B}} W(b, b'; i) - W_{\mathtt{B}}(b, b'; i)| < 2\delta.$$

We fix distinguished elements $a_0 \in \mathtt{A}$ and $b_0 \in \mathtt{B}$ which will be referred to throughout this section.

10.1. *The vertex measure.* We first define the weight's vertex measure by

$$W_{\mathtt{AB}}((a, b)) = \frac{1}{n} \lfloor n \cdot W((a, b)) \rfloor \quad a \in \mathtt{A} \setminus \{a_0\},\ b \in \mathtt{B},$$
$$W_{\mathtt{AB}}((a_0, b)) = W_{\mathtt{B}}(b) - \sum_{a \neq a_0} W_{\mathtt{AB}}((a, b)) \quad b \in \mathtt{B}.$$

See Table 1.

Note that $|W_{\mathtt{AB}}((a, b)) - W((a, b))| \leq 1/n$ for $a \neq a_0$ and

$$|W_{\mathtt{AB}}((a_0, b)) - W((a_0, b))| \leq |W_{\mathtt{B}}(b) - \pi_{\mathtt{B}} W(b)| + |\mathtt{A}|/n.$$

Therefore the ℓ^1 distance between the vertex measures is

$$\sum_{a,b} |W_{\mathtt{AB}}((a, b)) - W((a, b))| \leq |\mathtt{A}||\mathtt{B}|/n + \sum_{b \in \mathtt{B}} (|W_{\mathtt{B}}(b) - \pi_{\mathtt{B}} W(b)| + |\mathtt{A}|/n)$$
$$\leq 2\delta + 2|\mathtt{A}||\mathtt{B}|/n.$$

10.1.1. *Nonnegativity.* The terms defined by rounding down W using the floor function are guaranteed to be non-negative, but the others are not. In the following we show how to repair any negativity.

TABLE 1. Picking entries of the vertex measure $W_{\mathtt{AB}}(\cdot)$. First choose entries of the form $W_{\mathtt{AB}}((a, b))$ for $a \neq a_0$ by rounding down $W((a, b))$, then fill in the first column in a way that guarantees the correct $\mathtt{B}$-marginal.

	a_0	a_1	$\cdots$
b_0	$\rightarrow$	$\lfloor \cdot \rfloor$	$\lfloor \cdot \rfloor$
b_1	$\rightarrow$	$\lfloor \cdot \rfloor$	$\lfloor \cdot \rfloor$
$\vdots$	$\rightarrow$	$\lfloor \cdot \rfloor$	$\lfloor \cdot \rfloor$

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

Let $-R/n$ denote the sum of all negative terms in the vertex measure. Since W contains only non-negative terms, we have

$$\mathbf{1}_{\{W_{\mathtt{AB}}((a,b))<0\}} \cdot |W_{\mathtt{AB}}((a,b))| \leq |W_{\mathtt{AB}}((a,b)) - W((a,b))| \quad \text{for all } a, b.$$

Therefore

$$R/n \leq \sum_{b\in \mathtt{B}} |W_{\mathtt{AB}}((a_0,b)) - W((a_0,b))| \leq 2\delta + |\mathtt{A}||\mathtt{B}|/n.$$

Suppose there is some $b \in \mathtt{B}$ such that $W_{\mathtt{AB}}((a_0,b)) < 0$. Since $W_{\mathtt{AB}}$ has denominator n, we must have $W_{\mathtt{AB}}((a_0,b)) \leq -1/n$. By construction, we have

$$\sum_{a\in A} W_{\mathtt{AB}}((a,b)) = W_{\mathtt{B}}(b) \geq 0,$$

so there exists some $a^+ \in \mathtt{A}$ with $W_{\mathtt{AB}}((a^+,b)) \geq 1/n$. Increase $W_{\mathtt{AB}}((a_0,b))$ by $1/n$ and decrease $W_{\mathtt{AB}}((a^+,b))$ by $1/n$.

The number of times we must repeat this step before all terms are non-negative is exactly R, and each step moves the measure by ℓ^1 distance $2/n$; therefore the final edited vertex measure is distance at most $2R/n$ from the original $W_{\mathtt{AB}}$. If we now let $W_{\mathtt{AB}}$ denote the new, non-negative vertex measure, by the above bound on R/n we get

$$\sum_{a,b} |W_{\mathtt{AB}}((a,b)) - W((a,b))| \leq 6\delta + 4|\mathtt{A}||\mathtt{B}|/n.$$

10.2. *The* $\mathtt{B}$ *half-marginal.* For the purposes of this construction we use the $\mathtt{B}$ 'half-marginal', which we denote by

$$W(b,(a',b');i) := \sum_{a\in \mathtt{A}} W((a,b),(a',b');i).$$

This is an element of $\mathrm{Prob}((\mathtt{B}\times(\mathtt{A}\times\mathtt{B}))^r)$.

Before constructing the edge measure of $W_{\mathtt{AB}}$, in this section we first construct what will be its half-marginal.

For each $i \in [r]$, $b, b' \in \mathtt{B}$, and $a' \in \mathtt{A}$, we define

$$W_{\mathtt{AB}}(b,(a',b');i) = \frac{1}{n}\lfloor n \cdot W(b,(a',b');i)\rfloor \quad \text{for } a' \neq a_0,\ b \neq b_0, \tag{10.1}$$

$$W_{\mathtt{AB}}(b,(a_0,b');i) = W_{\mathtt{B}}(b,b';i) - \sum_{a'\neq a_0} W_{\mathtt{AB}}(b,(a',b');i) \quad \text{for } b \neq b_0, \tag{10.2}$$

$$W_{\mathtt{AB}}(b_0,(a',b');i) = W_{\mathtt{AB}}((a',b')) - \sum_{b\neq b_0} W_{\mathtt{AB}}(b,(a',b');i). \tag{10.3}$$

See Table 2 for a representation of which terms are defined by each equation.

The definition of the terms in (10.3) ensures that

$$\sum_{b\in \mathtt{B}} W_{\mathtt{AB}}(b,(a',b');i) = W_{\mathtt{AB}}((a',b')) \quad \text{for all } a', b', i.$$

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

TABLE 2. A diagram of how the half-marginal $W_{\mathtt{AB}}(\cdot,(\cdot,\cdot);i)$ is chosen if $\mathtt{A}=\{a_0,a_1,a_2\}$ and $\mathtt{B}=\{b_0,b_1,b_2\}$. First obtain the entries marked $\lfloor\cdot\rfloor$ by rounding down W. Then choose the entries marked $\rightarrow$ according to equation (10.2) which ensures that the $\mathtt{B}$-marginal is $W_{\mathtt{B}}$. Then choose the entries marked $\downarrow$ according to equation (10.3) which ensures that the vertex weight is the one we chose above.

	(a_0,b_0)	(a_1,b_0)	(a_2,b_0)	(a_0,b_1)	(a_1,b_1)	(a_2,b_1)	(a_0,b_2)	(a_1,b_2)	(a_2,b_2)
b_0	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
b_1	$\rightarrow$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\rightarrow$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\rightarrow$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$
b_2	$\rightarrow$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\rightarrow$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\rightarrow$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$

This will ensure that $W_{\mathtt{AB}}$ has the correct vertex measure. Note also that, by line (10.2),

$$\sum_{a'\in\mathtt{A}} W_{\mathtt{AB}}(b,(a',b');i) = W_{\mathtt{B}}(b,b';i) \quad \text{for all } b\in\mathtt{B}\setminus\{b_0\} \text{ and } b'\in\mathtt{B}.$$

Using this and definition (10.3), we also get

$$\sum_{a'\in\mathtt{A}} W_{\mathtt{AB}}(b_0,(a',b');i) = W_{\mathtt{B}}(b_0,b';i).$$

This will ensure that the $\mathtt{B}$-marginal of $W_{\mathtt{AB}}$ is $W_{\mathtt{B}}$.

We show now that the half-marginal $W_{\mathtt{AB}}(\cdot,(\cdot,\cdot);i)$ is ℓ^1-close to $W(\cdot,(\cdot,\cdot);i)$ by considering separately the contributions to the ℓ^1 distance from terms defined using equations (10.1), (10.2), and (10.3).

(10.1) terms: Each of the terms of $W_{\mathtt{AB}}$ defined using the floor in equation (10.1) is distance at most $1/n$ from the corresponding term of W; therefore the total contribution of these terms to the ℓ^1 distance is

$$\sum_{\substack{b\in\mathtt{B}\setminus\{b_0\}\\ a'\in\mathtt{A}\setminus\{a_0\},b'\in\mathtt{B}\\ i\in[r]}} |W_{\mathtt{AB}}(b,(a',b');i) - W(b,(a',b');i)| \le |\mathtt{A}||\mathtt{B}|^2 r/n.$$

(10.2) terms: By the triangle inequality,

$$\begin{aligned}
&|W_{\mathtt{AB}}(b,(a_0,b');i) - W(b,(a_0,b');i)| \\
&\quad = \left|\left(W_{\mathtt{B}}(b,b';i) - \sum_{a'\neq a_0} W_{\mathtt{AB}}(b,(a',b');i)\right)\right. \\
&\qquad \left. - \left(\pi_{\mathtt{B}} W(b,b';i) - \sum_{a'\neq a_0} W(b,(a',b');i)\right)\right| \\
&\quad \le |W_{\mathtt{B}}(b,b';i) - \pi_{\mathtt{B}} W(b,b';i)| \\
&\qquad + \sum_{a'\neq a_0} |W_{\mathtt{AB}}(b,(a',b');i) - W(b,(a',b');i)|.
\end{aligned}$$

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

The total contribution of such terms is therefore

$$\begin{aligned}
&\sum_{\substack{b\in \mathtt{B}\setminus\{b_0\},\ b'\in \mathtt{B}\\ i\in[r]}} |W_{\mathtt{AB}}(b,(a_0,b');i) - W(b,(a_0,b');i)| \\
&\leq \overbrace{\sum_{\substack{b\in \mathtt{B}\setminus\{b_0\},\ b'\in \mathtt{B}\\ i\in[r]}} |W_{\mathtt{B}}(b,b';i) - (\pi_{\mathtt{B}})_* W(b,b';i)|}^{\leq d_1(W_{\mathtt{B}},\pi_{\mathtt{B}}W)} \\
&\quad + \overbrace{\sum_{\substack{b\in \mathtt{B}\setminus\{b_0\}\\ a'\in \mathtt{A}\setminus\{a_0\},\ b'\in \mathtt{B}\\ i\in[r]}} |W_{\mathtt{AB}}(b,(a',b');i) - W(b,(a',b');i)|}^{=\text{contribution from (10.1) terms}} \\
&\leq 2\delta + |\mathtt{A}||\mathtt{B}|^2 r/n.
\end{aligned}$$

(10.3) terms: Again applying the triangle inequality,

$$\begin{aligned}
&|W_{\mathtt{AB}}(b_0,(a,b');i) - W(b_0,(a,b');i)| \\
&\quad \leq |W_{\mathtt{AB}}((a,b')) - W((a,b'))| \\
&\qquad + \sum_{b\neq b_0} |W_{\mathtt{AB}}(b,(a,b');i) - W(b,(a,b');i)|.
\end{aligned}$$

Summing over all $a \in \mathtt{A}$, $b' \in \mathtt{B}$ and $i \in [r]$, we see that the total contribution of such terms is bounded by

$$\begin{aligned}
&\sum_{\substack{a\in \mathtt{A}, b'\in \mathtt{B}\\ i\in[r]}} \Bigg[|W_{\mathtt{AB}}((a,b')) - W((a,b'))| \\
&\qquad + \sum_{b\neq b_0} |W_{\mathtt{AB}}(b,(a,b');i) - W(b,(a,b');i)| \Bigg] \\
&= \sum_{i\in[r]} \overbrace{\sum_{\substack{a\in \mathtt{A}\\ b\in \mathtt{B}}} |W_{\mathtt{AB}}((a,b)) - W((a,b))|}^{\text{vertex measure}} \\
&\quad + \overbrace{\sum_{\substack{b\in \mathtt{B}\setminus\{b_0\}\\ a'\in \mathtt{A}\setminus\{a_0\},\ b'\in \mathtt{B}\\ i\in[r]}} |W_{\mathtt{AB}}(b,(a',b');i) - W(b,(a',b');i)|}^{\text{(10.1) terms}} \\
&\quad + \overbrace{\sum_{\substack{b\in \mathtt{B}\setminus\{b_0\},\ b'\in \mathtt{B}\\ i\in[r]}} |W_{\mathtt{AB}}(b,(a_0,b');i) - W(b,(a_0,b');i)|}^{\text{(10.2) terms}} \\
&\leq r\cdot[6\delta + 4|\mathtt{A}||\mathtt{B}|/n] + [|\mathtt{A}||\mathtt{B}|^2 r/n] + [2\delta + |\mathtt{A}||\mathtt{B}|^2 r/n] \\
&\leq 8r\delta + 6|\mathtt{A}||\mathtt{B}|^2 r/n.
\end{aligned}$$

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

Adding up the contributions of the three types of terms, we see that the ℓ^1 distance between the half-marginals of W and $W_{\mathtt{AB}}$ is bounded by

$$10r\delta + 8|\mathtt{A}||\mathtt{B}|^2 r/n.$$

10.2.1. *Nonnegativity.* Again, the preceding construction does not guarantee that all terms are non-negative. In the following we describe how to correct negativity.

Let $-R/n$ be the sum of all negative terms of the half-marginal. As above, we get

$$R/n \leq 10r\delta + 7|\mathtt{A}||\mathtt{B}|^2 r/n.$$

Suppose there is some $b_- \in B$, $(a'_-, b'_-) \in \mathtt{A} \times \mathtt{B}$, and $i \in [r]$ such that $W_{\mathtt{AB}}(b_-, (a'_-, b'_-); i) < 0$. Then $W_{\mathtt{AB}}(b_-, (a'_-, b'_-); i) \leq -1/n$. Since

$$\sum_{a' \in \mathtt{A}} W_{\mathtt{AB}}(b_-, (a', b'_-); i) = W_{\mathtt{B}}(b_-, b'_-; i) \geq 0$$

and

$$\sum_{b \in \mathtt{B}} W_{\mathtt{AB}}(b, (a'_-, b'_-); i) = W_{\mathtt{AB}}((a'_-, b'_-)) \geq 0$$

there exist $a'_+ \in \mathtt{A}$ and $b_+ \in \mathtt{B}$ such that

$$W_{\mathtt{AB}}(b_-, (a'_+, b'_-); i) \geq 1/n \quad \text{and} \quad W_{\mathtt{AB}}(b_+, (a'_-, b'_-); i) \geq 1/n.$$

Decrease both of these terms by $1/n$, and increase both $W_{\mathtt{AB}}(b_-, (a'_-, b'_-); i)$ and $W_{\mathtt{AB}}(b_+, (a'_+, b'_-); i)$ by $1/n$. This moves the half-marginal by ℓ^1 distance $4/n$:

$$\sum_{a' \in \mathtt{A}} W_{\mathtt{AB}}(b, (a', b'); i) = W_{\mathtt{B}}(b, b'; i) \quad \text{and} \quad \sum_{b \in \mathtt{B}} W_{\mathtt{AB}}(b, (a', b'); i) = W_{\mathtt{AB}}((a', b')).$$

This step must be done at most R times to eliminate all negative entries, so the final half-marginal satisfies

$$\sum_{i \in [r]} \sum_{b \in \mathtt{B}} \sum_{(a', b') \in \mathtt{A} \times \mathtt{B}} |W_{\mathtt{AB}}(b, (a', b'); i) - W(b, (a', b'); i)|$$
$$\leq (10r\delta + 8|\mathtt{A}||\mathtt{B}|^2 r/n) + R \cdot 4/n \leq 50r\delta + 36|\mathtt{A}||\mathtt{B}|^2 r/n.$$

10.3. *The edge measure.* Finally, we define the edge measure of $W_{\mathtt{AB}}$ by

$$W_{\mathtt{AB}}((a, b), (a', b'); i) = \frac{1}{n} \lfloor n \cdot W((a, b), (a', b'); i) \rfloor$$
$$\text{for } a \neq a_0 \text{ and } (a', b') \neq (a_0, b_0), \tag{10.4}$$

$$W_{\mathtt{AB}}((a_0, b), (a', b'); i) = W_{\mathtt{AB}}(b, (a', b'); i) - \sum_{a \neq a_0} W_{\mathtt{AB}}((a, b), (a', b'); i)$$
$$\text{for } (a', b') \neq (a_0, b_0), \tag{10.5}$$

$$W_{\mathtt{AB}}((a, b), (a_0, b_0); i) = W_{\mathtt{AB}}((a, b)) - \sum_{(a', b') \neq (a_0, b_0)} W_{\mathtt{AB}}((a, b), (a', b'); i). \tag{10.6}$$

See Table 3.

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

TABLE 3. A diagram of how the edge measure $W_{\mathtt{AB}}((\cdot,\cdot),(\cdot,\cdot);i)$ is chosen if $\mathtt{A}=\{a_0,a_1,a_2\}$ and $\mathtt{B}=\{b_0,b_1,b_2\}$. First obtain the entries marked $\lfloor\cdot\rfloor$ by rounding down entries of W. Then choose entries marked $\downarrow$ according to equation (10.5), which ensures that the $\mathtt{B}$ half-marginal is the one chosen above. Then choose entries marked $\rightarrow$ according to equation (10.6), which ensures that the vertex measure is the one chosen above.

	(a_0,b_0)	(a_1,b_0)	(a_2,b_0)	(a_0,b_1)	(a_1,b_1)	(a_2,b_1)	(a_0,b_2)	(a_1,b_2)	(a_2,b_2)
(a_0,b_0)	$\rightarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
(a_1,b_0)	$\rightarrow$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$
(a_2,b_0)	$\rightarrow$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$
(a_0,b_1)	$\rightarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
(a_1,b_1)	$\rightarrow$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$
(a_2,b_1)	$\rightarrow$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$
(a_0,b_2)	$\rightarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
(a_1,b_2)	$\rightarrow$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$
(a_2,b_2)	$\rightarrow$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$	$\lfloor\cdot\rfloor$

It follows from this definition that $W_{\mathtt{AB}}$ is a (signed) weight with $\mathtt{B}$-marginal $W_{\mathtt{B}}$.

We now check that $W_{\mathtt{AB}}$ is ℓ^1-close to W. We consider separately the contribution to the ℓ^1 distance of terms defined in equations (10.4), (10.5), and (10.6).

(10.4) terms: Each term of $W_{\mathtt{AB}}$ defined using the floor function in equation (10.4) is distance at most $1/n$ from the corresponding W term. The total contribution of these terms to the ℓ^1 distance is therefore at most $|\mathtt{A}|^2|\mathtt{B}|^2 r/n$.

(10.5) terms: Applying the triangle inequality to terms defined in equation (10.5),

$$\begin{aligned}
&|W_{\mathtt{AB}}((a_0,b),(a',b');i) - W((a_0,b),(a',b');i)| \\
&\quad\le |W_{\mathtt{AB}}(b,(a',b');i) - W(b,(a',b');i)| \\
&\qquad + \sum_{a\neq a_0} |W_{\mathtt{AB}}((a,b),(a',b');i) - W((a,b),(a',b');i)| \\
&\quad\le |W_{\mathtt{AB}}(b,(a',b');i) - W(b,(a',b');i)| + |\mathtt{A}|/n.
\end{aligned}$$

By the ℓ^1 bound on the distance between the half-marginals, the total contribution of all such terms is therefore at most

$$\begin{aligned}
&\sum_{i\in[r]}\sum_{b}\sum_{(a',b')\neq(a_0,b_0)} (|W_{\mathtt{AB}}(b,(a',b');i) - W(b,(a',b');i)| + |\mathtt{A}|/n) \\
&\quad\le [50r\delta + 36|\mathtt{A}|^2|\mathtt{B}|^2 r/n] + |\mathtt{A}|^2|\mathtt{B}|^2 r/n \\
&\quad= 50r\delta + 37|\mathtt{A}|^2|\mathtt{B}|^2 r/n.
\end{aligned}$$

(10.6) terms: Applying the triangle inequality to terms defined in equation (10.6),

$$\begin{aligned}
&|W_{\mathtt{AB}}((a,b),(a_0,b_0);i) - W_{\mathtt{AB}}((a,b),(a_0,b_0);i)| \\
&\quad\le |W_{\mathtt{AB}}((a,b)) - W((a,b))| \\
&\qquad + \sum_{(a',b')\neq(a_0,b_0)} |W_{\mathtt{AB}}((a,b),(a',b');i) - W((a,b),(a',b');i)|.
\end{aligned}$$

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

Therefore the total contribution of all such terms is

$$\begin{aligned}
&\sum_{i\in[r]}\sum_{a,b}|W_{\mathtt{AB}}((a,b),(a_0,b_0);i)-W_{\mathtt{AB}}((a,b),(a_0,b_0);i)|\\
&\quad=\sum_{i\in[r]}\sum_{a,b}\Bigg[|W_{\mathtt{AB}}((a,b))-W((a,b))|\\
&\qquad+\sum_{(a',b')\neq(a_0,b_0)}|W_{\mathtt{AB}}((a,b),(a',b');i)-W((a,b),(a',b');i)|\Bigg]\\
&\quad=\overbrace{\sum_{i\in[r]}\sum_{a,b}|W_{\mathtt{AB}}((a,b))-W((a,b))|}^{\text{vertex measure}}\\
&\qquad+\overbrace{\sum_{i\in[r]}\sum_{a\neq a_0}\sum_{b}\sum_{(a',b')\neq(a_0,b_0)}|W_{\mathtt{AB}}((a,b),(a',b');i)-W((a,b),(a',b');i)|}^{\text{(10.4) terms}}\\
&\qquad+\overbrace{\sum_{i\in[r]}\sum_{b}\sum_{(a',b')\neq(a_0,b_0)}|W_{\mathtt{AB}}((a_0,b),(a',b');i)-W((a_0,b),(a',b');i)|}^{\text{(10.5) terms}}\Bigg]\\
&\quad\leq r\cdot[6\delta+3|\mathtt{A}||\mathtt{B}|/n]+[|\mathtt{A}|^2|\mathtt{B}|^2r/n]+[50r\delta+37|\mathtt{A}|^2|\mathtt{B}|^2r/n]\\
&\quad\leq 56r\delta+41|\mathtt{A}|^2|\mathtt{B}|^2r/n.
\end{aligned}$$

Summing up the contributions from terms of all three types, we get that

$$d_1(W_{\mathtt{AB}},W)\leq 106r\delta+79|\mathtt{A}|^2|\mathtt{B}|^2r/n.$$

10.3.1. *Nonnegativity.* We can modify a solution with negative entries to get a non-negative one similarly to above. Let $-R/n$ be the sum of all negative entries; then

$$R/n\leq 106r\delta+78|\mathtt{A}|^2|\mathtt{B}|^2r/n.$$

Suppose there is some entry

$$W_{\mathtt{AB}}((a_-,b_-),(a'_-,b'_-);i)\leq -1/n.$$

We want to increment this term by $1/n$ without affecting the vertex measure or the $\mathtt{B}$ marginal. Since

$$\sum_{(a',b')\in\mathtt{A}\times\mathtt{B}}W_{\mathtt{AB}}((a_-,b_-),(a',b');i)=W_{\mathtt{AB}}((a_-,b_-))\geq 0$$

there exists some $(a'_+,b'_+)\in\mathtt{A}\times\mathtt{B}$ such that $W_{\mathtt{AB}}((a_-,b_-),(a'_+,b'_+);i)\geq 1/n$; similarly, since

$$\sum_{a\in A}W_{\mathtt{AB}}((a,b_-),(a',b'_-);i)=W_{\mathtt{AB}}(b_-,(a'_-,b'_-);i)\geq 0$$

there exists some a_+ such that $W_{\mathtt{AB}}((a_+,b_-),(a'_-,b'_-);i)\geq 1/n$. Increase

$$W_{\mathtt{AB}}((a_-,b_-),(a'_-,b'_-);i)\quad\text{and}\quad W_{\mathtt{AB}}((a_+,b_-),(a'_+,b'_+);i)$$

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

by $1/n$, and decrease

$$W_{\mathtt{AB}}((a_-, b_-), (a'_+, b'_+); i) \quad \text{and} \quad W_{\mathtt{AB}}((a_+, b_-), (a'_-, b'_-); i)$$

by $1/n$. This moves the weight by ℓ^1 distance $4/n$.

Since R is the maximum number of times we need to do this before there are no more negative entries, the final weight satisfies

$$d_1(W_{\mathtt{AB}}, W) \leq 106r\delta + 79|\mathtt{A}|^2|\mathtt{B}|^2 r/n + 4R/n \leq 530r\delta + 391|\mathtt{A}|^2|\mathtt{B}|^2 r/n.$$

To simplify, we write

$$d_1(W_{\mathtt{AB}}, W) \leq 530r(\delta + |\mathtt{A} \times \mathtt{B}|^2/n),$$

or

$$d(W_{\mathtt{AB}}, W) \leq 265r(\delta + |\mathtt{A} \times \mathtt{B}|^2/n).$$

Acknowledgements. Thanks go to Tim Austin for suggesting that results like Theorems B and C should hold, for many helpful discussions, and for providing comments on earlier drafts. Thanks also go to Ben Hayes for sharing helpful references to the operator algebras literature, and to Lewis Bowen and Brandon Seward for helpful conversations. Thanks go to an anonymous referee for many helpful comments on an earlier draft. This material is based upon work supported by the National Science Foundation under grant no. DMS-1855694.

References

[1] D. Airey, L. Bowen and F. Lin. A topological dynamical system with two different positive sofic entropies. *Trans. Amer. Math. Soc.* B **9**(2) (2022), 35–98.
[2] L. Bowen. The ergodic theory of free group actions: entropy and the f-invariant. *Groups Geom. Dyn.* **4** (2010), 419–432.
[3] L. Bowen. A measure-conjugacy invariant for free group actions. *Ann. of Math. (2)* **171**(2) (2010), 1387–1400.
[4] L. Bowen. Measure conjugacy invariants for actions of countable sofic groups. *J. Amer. Math. Soc.* **23**(1) (2010), 217–217.
[5] L. Bowen. Non-abelian free group actions: Markov processes, the Abramov–Rohlin formula and Yuzvinskii's formula. *Ergod. Th. & Dynam. Sys.* **30**(6) (2010), 1629–1663.
[6] L. Bowen. Sofic entropy and amenable groups. *Ergod. Th. & Dynam. Sys.* **32**(2) (2012), 427–466.
[7] L. Bowen and Y. Gutman. Nonabelian free group actions: Markov processes, the Abramov–Rohlin formula and Yuzvinskii's formula – Corrigendum. *Ergod. Th. & Dynam. Sys.* **33**(2) (2013), 643–645.
[8] A. Coja-Oghlan, M. Hahn-Klimroth, P. Loick, N. Müller, K. Panagiotou and M. Pasch. Inference and mutual information on random factor graphs. *38th International Symposium on Theoretical Aspects of Computer Science (STACS 2021) (Leibniz International Proceedings in Informatics (LIPIcs), 187)*. Ed. M. Bläser and B. Monmege. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2021, pp. 24:1–24:15.
[9] T. M. Cover and J. A. Thomas. *Elements of Information Theory*, 2nd edn. Wiley-Interscience, Hoboken, NJ, 2006.
[10] K. Dykema, D. Kerr and M. Pichot. Sofic dimension for discrete measured groupoids. *Trans. Amer. Math. Soc.* **366**(2) (2013), 707–748.
[11] B. Hayes. Relative entropy and the Pinsker product formula for sofic groups. *Groups Geom. Dyn.* **15** (2016), 413–463.
[12] D. Kerr and H. Li. Entropy and the variational principle for actions of sofic groups. *Invent. Math.* **186**(3) (2011), 501–558.

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press

[13] D. S. Ornstein and B. Weiss. Entropy and isomorphism theorems for actions of amenable groups. *J. Anal. Math.* **48**(1) (1987), 1–141.
[14] L. Păunescu. On sofic actions and equivalence relations. *J. Funct. Anal.* **261**(9) (2011), 2461–2485.
[15] S. Popa. Independence properties in subalgebras of ultraproduct Π_1 factors. *J. Funct. Anal.* **266**(9) (2014), 5818–5846.

https://doi.org/10.1017/etds.2022.27 Published online by Cambridge University Press