TYPICALITY AND ENTROPY OF PROCESSES ON INFINITE TREES

ÁGNES BACKHAUSZ[1,2], CHARLES BORDENAVE[3], AND BALÁZS SZEGEDY[2]

ABSTRACT. Consider a uniformly sampled random d-regular graph on n vertices. If d is fixed and n goes to ∞ then we can relate typical (large probability) properties of such random graph to a family of invariant random processes (called "typical" processes) on the infinite d-regular tree T_d. This correspondence between ergodic theory on T_d and random regular graphs is already proven to be fruitful in both directions. This paper continues the investigation of typical processes with a special emphasis on entropy. We study a natural notion of micro-state entropy for invariant processes on T_d. It serves as a quantitative refinement of the notion of typicality and is tightly connected to the asymptotic free energy in statistical physics. Using entropy inequalities, we provide new sufficient conditions for typicality for edge Markov processes. We also extend these notions and results to processes on unimodular Galton-Watson random trees.

ABSTRACT. [Fr] On considère un graphe d-régulier aléatoire avec n sommets uniformément distribué. Si d est fixé et n diverge, nous pouvons alors relié les propriétés typiques (de grande probabilité) d'un tel graphe aléatoire avec une famille de processus aléatoires invariants (dénommés processus "typiques") sur l'arbre d-régulier infini T_d. Cette correspondance entre théorie ergodique sur T_d et graphes réguliers aléatoires s'est déjà révélée fructueuse dans les deux directions. Ce papier poursuit l'investigation des processus typiques avec un accent mis sur l'entropie. Nous y étudions une notion naturelle d'entropie micro-état pour les processus invariant sur T_d. Elle sert de rafinement quantitatif à la notion de typicalité et elle est intimement reliée à l'energie libre asymptotique en physique statistique. Au moyen d'inégalités entropiques, nous démontrons des nouvelles conditions suffisantes de typicalité pour des processus markovien sur les arêtes de l'arbre. Nous étendons aussi ces notions et résultats à des processus sur des arbres de Galton-Watson unimodulaires.

1. INTRODUCTION

1.1. **Typical processes and sofic entropy.** Random d-regular graphs have been extensively studied over the past 50 years [10, 21, 28, 32]. Sophisticated methods from probability theory, combinatorics and statistical physics have been successfully used to uncover many of their properties such as independence ratio, the density of a maximal cut or its spectral gap [18, 31, 21]. The recently emerging theory of graph limits [8, 24, 22, 26] gives a new, limiting point of views on the subject. It turns out that many of the crucial properties of random d-regular graphs for d fixed and n going to infinity can also be studied in the framework of ergodic theory on the infinite d-regular tree T_d [2]. An illustration of the power of this method is the proof of the Gaussianity of the almost eigenvectors of random d-regular graphs [3, 11]. The proof is a combination of ergodic theory on T_d with information theoretic methods.

In this paper we are interested in a question which is also related to the limit of random regular graphs, namely, to the family of typical processes on T_d introduced in [2]. These objects arise as the local limits of vertex-colored random d-regular graphs (see formal definition below). These typical processes contain useful information on the structure of d-regular graphs. For example, the classical fact from [9] that the independence ratio is separated from $1/2$ is equivalent to the fact that the alternating coloring of the infinite d-regular tree is not typical. Several necessary conditions have been formulated for typical processes in the last years. Some of them are about the covariance structure, others are entropy inequalities [4, 5, 15]. However, general sufficient conditions for a process to be typical are less common.

In this paper we obtain sufficient conditions for the typicality of processes on T_d, by studying a new micro-state entropy. This entropy measures in some sense the number of finite approximations of a process

[1] ELTE EÖTVÖS LORÁND UNIVERSITY, BUDAPEST, HUNGARY, FACULTY OF SCIENCE AND ALFRÉD RÉNYI INSTITUTE OF MATHEMATICS

[2] ALFRÉD RÉNYI INSTITUTE OF MATHEMATICS

[3]INSTITUT DE MATHÉMATIQUES DE MARSEILLE

on a random instance of a large d-regular graph. This entropy is tightly connected to Bowen's sofic entropy in measured group theory, see [13].

Our approach has another connection to the convergence of random graphs. While graph limit theory shows great promise in a variety of questions related to random d-regular graphs, it also revealed an intriguing open problem. It is believed that for fix d and $n \to \infty$ we have that a random d-regular graph on n vertices is convergent (in probability) in the local-global sense of [24] and also right-convergent in the sense of [22], this last notion of convergence relies on a deep statistical physics theory, see the monograph [27]. These general conjectures are a common strengthening of a large variety of conjectures many of which are already proven. For example the convergence of the independence ratio is proven in [6, 17] and the convergence of asymptotic free energies of a large class of statistical physics models on random d-regular graphs in [29, 14]. In this paper, we introduce an upper and a lower version of the micro-state entropies. Equality of these entropies imply the convergence of random regular graphs. For a family of processes we can establish this equality. This will lead to the sufficient conditions mentioned above, and this shows that our results might lead to a deep understanding of the structure of random regular graphs.

1.2. **Definitions.** We now formalize the main definitions. Let $d \geq 3$ be an integer and T_d the infinite d-regular tree (all vertices have d neighbors). Let M be a finite set. A process X on T_d is a random variable on M^{T_d}. This process (or its law) is *invariant* if the law of X is invariant by all automorphisms of the tree T_d. We denote by $\mathcal{I}_d(M)$ the set of invariant probability measures on M^{T_d}.

We now define the Benjamini-Schramm topology. A pair (G, f) formed by a graph $G = (V, E)$ and a map $f : V \to M$ will be called a colored graph with color set M. A rooted colored graph is a triple (G, f, o) formed by a connected colored graph (G, f) and a distinguished vertex o of G, called the root. Two rooted colored graphs (G, f, o) and (G', f', o') are isomorphic if there is an isomorphism of G and G' which preserves the colors and the roots. An equivalence class of rooted colored graphs is called an unlabeled rooted colored graph in combinatorics.

Unlabeled rooted graphs give the proper setup for defining a meaningful notion of convergence. It is however more convenient to work with rooted labeled graphs instead of unlabeled rooted graphs. To this end, we now define a randomized canonical graph in each equivalence class. We define the set of finite integer sequences as

$$\mathbb{N}^f = \cup_{k \geq 0} \mathbb{N}^k \tag{1.1}$$

where $\mathbb{N}^0 = \{o\}$ and $\mathbb{N} = \{1, 2, \ldots, \}$ by convention. The tree T_d can be classically built on a subset of $\mathbb{N}^f$ as follows. The root of the tree is o, its d-neighbors of o are $V_1 = \{1, \ldots, d\} \subset \mathbb{N}$, the neighbors of $i \in V_1$ are o and $\{(i, 1), \ldots, (i, d-1)\} \in \mathbb{N}^2$ and so on. More generally, if (G, o) is a rooted graph, the breadth-first search tree started at the root o, where ties between vertices are broken uniformly at random defines a random graph (G', o) on a subset of $\mathbb{N}^f$ whose law depends only the equivalence class of (G, o): a vertex at distance k from the root receives a label in $\mathbb{N}^k$, if $(i_1, \ldots, i_{k-1})$ is the label of its parent in the search tree, it has the label $(i_1, \ldots, i_{k-1}, j)$ if it the j-th offspring of its parent in the random ordering. We call this random rooted graph, the *randomly labeled rooted graph* associated to (G, o). Conversely, we will say that a random labeled colored graph (G, f, o) on a subset of $\mathbb{N}^f$ is randomly labeled if its law is equal to the law of the randomly labeled rooted colored graph associated to its unlabeled rooted colored graph. By definition if $X \in \mathcal{I}_d(M)$ then (T_d, X, o) is randomly labeled.

Recall that a graph is locally finite if all its vertices have a finite number of neighbors. We denote by $\mathcal{G}^\bullet$ (respectively $\mathcal{G}_M^\bullet$) the set of locally finite graphs (respectively locally finite colored graphs on the color set M) on the vertex set $\mathbb{N}^f$ rooted at o which are admissible in the sense that they are realizable as a breadth-first search labeling of a locally finite graph. The sets $\mathcal{G}^\bullet$ and $\mathcal{G}_M^\bullet$ are complete separable metric spaces when equipped with the distance $\mathrm{d}(g, g') = \sum_{r \geq 0} 2^{-r} \mathbf{1}_{(g)_r \neq (g')_r}$ where $\mathbf{1}$ is the indicator function and g_r is the restriction of g to the vertices at distance at most r from the root. We denote by $\mathcal{P}(\mathcal{G}_M^\bullet)$ the set of probability measures on $\mathcal{G}_M^\bullet$. We equip $\mathcal{P}(\mathcal{G}_M^\bullet)$ with a distance, also denoted by d, which generates the weak topology on $\mathcal{P}(\mathcal{G}_M^\bullet)$ (for example, the Lévy-Prohorov distance).

If (G, f) is a locally finite colored graph and v is a vertex of G then we denote by $\mathrm{distr}_{G,v}(f)$ the law in $\mathcal{P}(\mathcal{G}_M^\bullet)$ of the randomly labeled rooted graph $((G, f)(v), v)$ where $(G, f)(v)$ is the restriction of (G, f) to the connected component of G containing v. The law $\mathrm{distr}_{G,v}$ in $\mathcal{P}(\mathcal{G}^\bullet)$ is defined similarly for a graph G and a vertex v. Finally, if G is finite with vertex set V, we may define the probability measures in $\mathcal{P}(\mathcal{G}^\bullet)$ and

$\mathcal{P}(\mathcal{G}_M^\bullet)$:

$$\text{distr}_G = \frac{1}{|V|} \sum_{v \in V} \text{distr}_{G,v} \quad \text{and} \quad \text{distr}_G(f) = \frac{1}{|V|} \sum_{v \in V} \text{distr}_{G,v}(f).$$

For integers $n \geq d+1$ with nd even, the set $\mathcal{G}_n(d)$ of simple d-regular graphs on the vertex set $[n] = \{1, \ldots, n\}$ is not empty. For each integer $n \geq d+1$ with nd even, let G_n be a uniformly distributed random graph on $\mathcal{G}_n(d)$. Almost-surely, the probability distribution distr_{G_n} converges as n goes to infinity to the Dirac mass at T_d rooted at o (it is a consequence of the fact that the number of cycles of length k in G_n is $O(1)$ for any fixed k, see [10]). It can further be checked that, a.s. if μ is an accumulation point of $\text{distr}_{G_n}(f_n)$ for some sequence of colorings $f_n \in M^n$, then $\mu \in \mathcal{I}_d(M)$. This motivates the following definition introduced in [2].

Definition 1.1 (Typical process). *A measure $\mu \in \mathcal{I}_d(M)$ is* weakly typical *if*

$$\lim_{\epsilon \to 0} \limsup_{n \to \infty} \mathbb{P}(\exists f \in M^n : \mathrm{d}(\text{distr}_{G_n}(f), \mu) \leq \epsilon) = 1.$$

It is strongly typical *if*

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \mathbb{P}(\exists f \in M^n : \mathrm{d}(\text{distr}_{G_n}(f), \mu) \leq \epsilon) = 1.$$

Note that it is apparent from the definition that being typical does not depend on the choice of the distance d which generates the weak topology. We note also that the definition in [2] is slightly different but it turns out to be equivalent by using some measure concentration phenomena (see [3, Section 5]).

We may also define a notion of micro-state entropy of $\mu \in \mathcal{I}_d(M)$ as follows. For $\mu \in \mathcal{I}_d(M)$ and $r \geq 0$, the probability measure μ_r is defined as the restriction of μ to the vertices at distance at most r from the root. Similarly, for $f \in M^n$, we define $\text{distr}_G(f)_r$ as the restriction of the distribution $\text{distr}_G(f)$ to the set of vertices at distance at most r from the root. For any $\epsilon > 0$, if $G \in \mathcal{G}_n(d)$, we define

$$\mathcal{F}_G(\mu, r, \epsilon) = \{f \in M^n : \mathrm{d}(\text{distr}_G(f)_r, \mu_r) \leq \epsilon\}. \tag{1.2}$$

This is the set of coloring functions f on G which are ϵ close to μ_r in the Benjamini-Schramm sense. Let G_n be a uniformly distributed random graph on $\mathcal{G}_n(d)$ with $n \geq d-1$ and nd even. Roughly speaking the sofic entropy of μ is a limit of

$$H_{G_n}(\mu, r, \epsilon) = \frac{1}{n} \log |\mathcal{F}_{G_n}(\mu, r, \epsilon)|, \tag{1.3}$$

in $n \to \infty$ and then in $\epsilon \to 0$ and $r \to \infty$. Notice that random regular graphs could be replaced by other locally tree-like graph sequences, but as the expectation of $|\mathcal{F}_{G_n}(\mu, r, \epsilon)|$ is easier to compute in our case, we use the definition based on random regular graphs. Since $H_{G_n}(\mu, r, \epsilon)$ is a random variable in $\{-\infty\} \cup [0, \infty)$ some care is needed. More formally, we fix $0 < \alpha < 1$ and consider the following median value of $H_{G_n}(\mu, r, \epsilon)$,

$$h_n(\mu, r, \epsilon, \alpha) = \sup \{h \in \{-\infty\} \cup [0, \infty) : \mathbb{P}(H_{G_n}(\mu, r, \epsilon) \geq h) \geq \alpha\}.$$

Since h_n is non-decreasing in ϵ, we may define the upper and lower entropies of μ as

$$\bar{h}(\mu, r, \alpha) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} h_n(\mu, r, \epsilon, \alpha) \quad \text{and} \quad \underline{h}(\mu, r, \alpha) = \lim_{\epsilon \to 0} \liminf_{n \to \infty} h_n(\mu, r, \epsilon, \alpha).$$

These entropies are extended real numbers in $\{-\infty\} \cup [0, \infty)$. This entropy can be interpreted as a version of Bowen's sofic entropy, see the survey [13].

The sofic entropies $\bar{h}(\mu, r, \alpha)$ and $\underline{h}(\mu, r, \alpha)$ do not depend on $\alpha \in (0, 1)$. Note that they do not depend neither on the choice of the distance d (in the sense that if two distances are topologically equivalent, the corresponding quantities $\bar{h}(\mu, r, \alpha)$ and $\underline{h}(\mu, r, \alpha)$ are equal). We shall prove the following.

Lemma 1.2. *Let $\mu \in \mathcal{I}_d(M)$ and $r \geq 0$. The function $\alpha \mapsto (\bar{h}(\mu, r, \alpha), \underline{h}(\mu, r, \alpha))$ is constant on $(0, 1)$.*

By Lemma 1.2, we may consider the common value of the entropies: for all $\alpha \in (0, 1)$,

$$\bar{h}(\mu, r) = \bar{h}(\mu, r, \alpha) \quad \text{and} \quad \underline{h}(\mu, r) = \underline{h}(\mu, r, \alpha).$$

By construction, $\bar{h}(\mu, r)$ and $\underline{h}(\mu, r)$ are the growth rates of the number of coloring of a random d-regular graph whose r-neighborhood is close to μ_r. Finally, since $\bar{h}(\mu, r)$ and $\underline{h}(\mu, r)$ are non-increasing in r, we may define the *upper and lower sofic entropies* as

$$\bar{h}(\mu) = \lim_{r \to \infty} \bar{h}(\mu, r) \quad \text{and} \quad \underline{h}(\mu) = \lim_{r \to \infty} \underline{h}(\mu, r).$$

Taking the limit as $\alpha \to 1$, for $h \geq 0$, the inequality $\bar{h}(\mu) \geq h$ is equivalent to the existence of a vanishing sequence (ϵ_n) and such that

$$\limsup_{n \to \infty} \mathbb{P}(H_{G_n}(\mu, 1/\epsilon_n, \epsilon_n) \geq (h - \epsilon_n)_+) = 1,$$

where $(x)_+ = x \vee 0$, and similarly for $\underline{h}(\mu)$. Since for $\nu, \mu \in \mathcal{P}(\mathcal{G}_M^\bullet)$, $\mathrm{d}(\nu_r, \mu_r)$ converges to $\mathrm{d}(\nu, \mu)$ as $r \to \infty$, the sofic entropy is thus closely related to the typicality:

Lemma 1.3. *Let $\mu \in \mathcal{I}_d(M)$. We have $\bar{h}(\mu) \geq 0$ (resp. $\underline{h}(\mu) \geq 0$) if and only if μ is weakly (resp. strongly) typical.*

This work is notably motivated the following conjecture which is connected to the notion of right convergence, see [22] and Subsection 1.6 below.

Conjecture 1.4. *For all $\mu \in \mathcal{I}_d(M)$, we have $\underline{h}(\mu) = \bar{h}(\mu)$. In particular, μ is weakly typical if and only if it is strongly typical.*

In this work, we will compute the entropy $\underline{h}(\mu) = \bar{h}(\mu)$ for a large class of invariant measures $\mu \in \mathcal{I}_d(M)$. This class is a class of processes where a second moment method can be applied. In a subsequent work, we will use more advanced statistical physics methods to refine our criterion. This conjecture might also be related to the Guerra's interpolation method [23] developed in the context of random graphs notably in [20, 6, 22, 29, 14, 25].

1.3. **Annealed entropy.** If $r \geq 0$ is an integer and S is a subset of T_d, we define $B_r(S)$ the subset of vertices of T_d at distance at most r from a vertex in S. For ease of notation, for $r \geq 1$, we define $S_r = B_r(o)$ as the ball of radius r around the root of T_d and $E_r = B_{r-1}(\{o, 1\})$ is the set of vertices at distance $r - 1$ from the edge $\{o, 1\}$ of T_d.

If X is an invariant process with law $\mu \in \mathcal{I}_d(M)$ and $r \geq 1$ integer, we set

$$\Sigma_r(X) = \Sigma_r(\mu) = H(X_{S_r}) - \frac{d}{2}H(X_{E_r}), \tag{1.4}$$

where X_S is the restriction of X to the subset $S \subset T_d$ and H is the usual Shannon entropy: if Y is a random variable taking value a finite set F, then

$$H(Y) = -\sum_{x \in F} \mathbb{P}(Y = x) \ln \mathbb{P}(Y = x).$$

By abusing notation, we denote entropy in the same way for both random variables and their distributions. It follows from [3, 12] (see Subsection 2.2 below for details) that $\Sigma_r(\mu)$ is non-increasing in r. We may thus define

$$\Sigma(\mu) = \lim_{r \to \infty} \Sigma_r(\mu).$$

Note that the law of X_{S_r} is μ_r and that the law of X_{E_r} is a marginal of μ_r. The quantities $\Sigma_r(\mu)$ and $\Sigma(\mu)$ will be called the annealed entropy of μ_r and μ, the reason will be clear in the forthcoming Subsection 2.3. The following first moment bound is essentially contained in [2, 12].

Theorem 1.5. *For any $\mu \in \mathcal{I}_d(M)$ and integer $r \geq 1$, we have*

$$\bar{h}(\mu, r) \leq \Sigma_r(\mu) \quad and \quad \bar{h}(\mu) \leq \Sigma(\mu).$$

As a corollary, we recover the "star-edge inequality" of [2, 3] which is a necessary condition of typicality: if μ is weakly typical then by Lemma 1.3, $\bar{h}(\mu) \geq 0$ and thus, by Theorem 1.5, $\Sigma(\mu) \geq 0$.

Corollary 1.6 ([2, 3]). *If $\mu \in \mathcal{I}_d(M)$ is a weakly typical process then $\Sigma(\mu) \geq 0$.*

The main result of this paper is a matching lower bound for a large class of invariant processes. To this end, we first recall the notion of coupling restricted to our setting. Let M_1 and M_2 be two finite sets, X_1 and X_2 be two random variables on $M_1^{T_d}$ and $M_2^{T_d}$ with respective laws μ_1 and μ_2. A coupling of μ_1 and μ_2 is a distribution ν on $(M_1 \times M_2)^{T_d}$ such that if $Y = (Y_1, Y_2)$ has law ν, Y_i has law μ_i for $i = 1, 2$. If X_i is an invariant process for $i = 1, 2$, we say that ν or Y is an invariant coupling if $\nu \in \mathcal{I}_d(M_1 \times M_2)$.

Theorem 1.7. *Let $\mu \in \mathcal{I}_d(M)$. For any integer $r \geq 1$, if all invariant couplings ν of μ and μ satisfy $\Sigma_r(\nu) \leq 2\Sigma_r(\mu)$ then $\underline{h}(\mu, r) = \bar{h}(\mu, r) = \Sigma_r(\mu)$. In particular, if the above condition holds for an increasing sequence of integers $(r_k)_{k \geq 1}$ then $\underline{h}(\mu) = \bar{h}(\mu) = \Sigma(\mu)$.*

We note that the bound $\Sigma_r(\nu) \leq 2\Sigma_r(\mu)$ is attained for the independent coupling $Y = (X_1, X_2)$ with X_i independent with law μ. Note also that if ν is the trivial coupling of μ and μ, that is $Y = (X, X)$ with X with law μ, we find $\Sigma_r(\nu) = \Sigma_r(\mu)$. Hence, under the condition of Theorem 1.7, we have $\Sigma_r(\mu) \leq 2\Sigma_r(\mu)$ or equivalently $\Sigma_r(\mu) \geq 0$. As a corollary, by Lemma 1.3, we thus obtain the following sufficient condition for typicality.

Corollary 1.8. *Let $\mu \in \mathcal{I}_d(M)$ and $(r_k)_{k \geq 1}$ an increasing sequence of integers be such that for all invariant couplings ν of μ and μ and all $k \geq 1$ we have $\Sigma_{r_k}(\nu) \leq 2\Sigma_{r_k}(\mu)$. Then μ is strongly typical.*

1.4. **Edge-Markov processes.** There is a specific class of processes in $\mathcal{I}_d(M)$ for which it is possible to improve on Theorem 1.7, the edge-Markov processes defined as follows. As above, for integer $r \geq 0$, $B_r(S)$ is the r-neighborhood of a subset S in T_d. For $r \geq 1$, recall that $S_r = B_r(o)$ and $E_r = B_{r-1}(\{o, 1\})$.

Definition 1.9 (Edge-Markov process). *A probability measure in M^{T_d} is edge-Markov if conditioned on the value at an edge, the processes on the left and right subtrees of that edge are independent.*

More generally, for integer $r \geq 1$, a probability measure on M^{T_d} is r-Markov if conditioned on the value at $B_{r-1}(e)$, the $(r-1)$-neighborhood of an edge e, the processes on the left and right subtrees of e are independent (for $r = 1$ we recover the edge-Markov process).

Let $\mathcal{I}_{d,r}(M)$ denote the set of probability measures on M^{S_r} that are invariant by automorphisms of S_r and whose restriction to E_r is invariant by switching the two sides of the edge $\{o, 1\}$. If $\mu \in \mathcal{I}_d(M)$, then, μ_r, its restriction to S_r, is in $\mathcal{I}_{d,r}(M)$. Conversely, the following lemma is easy to see.

Lemma 1.10. *Let $r \geq 1$ be an integer and $p \in \mathcal{I}_{d,r}(M)$. Then there is a unique r-Markov process $\mu(p) \in \mathcal{I}_d(M)$ such that the marginal of $\mu(p)$ on S_r is equal to p.*

If $p \in \mathcal{I}_{d,r}(M)$, we define $\Sigma(p) = \Sigma_r(\mu(p)) = H(X_{S_r}) - (d/2)H(X_{E_r})$ as in Equation (1.4), where X has law p. As above, if p_1 and p_2 are probability measures on $M_1^{S_r}$ and $M_2^{S_r}$, a coupling of p_1 and p_2 is a probability measure on $M_1^{S_r} \times M_2^{S_r}$ whose marginals are p_1 and p_2. The following theorem is a strengthening of Theorem 1.7 for egde-Markov processes.

Theorem 1.11. *Let $r \geq 1$ be an integer and $p \in \mathcal{I}_{d,r}(M)$. If for all couplings $q \in \mathcal{I}_{d,r}(M^2)$ of p and p, we have $\Sigma(q) \leq 2\Sigma(p)$ then $\underline{h}(\mu(p)) = \bar{h}(\mu(p)) = \Sigma(p)$ and $\mu(p)$ is strongly typical.*

Theorem 1.11 provides an easy to check criteria for typicality for edge Markov processes. In the course of the proof, we will need an important maximizing property satisfied by edge Markov processes (a closely related characterization can be found in [12, Theorem 1.3] and [3, Lemma 10.1]).

Lemma 1.12. *Let $X \in \mathcal{I}_d(M)$ and $r \geq 1$. We have*

$$\Sigma_{r+1}(X) \leq \Sigma_r(X)$$

with equality if and only if $X_{S_{r+1}}$ is a r-Markov process on S_{r+1}.

Notice that we also get that $\Sigma(X) = \Sigma_r(X)$ holds for r-Markov processes.

1.5. **Vertex-Markov processes.** There is a subclass of edge-Markov processes for which the annealed entropy takes a particularly simple form.

Definition 1.13 (Vertex-Markov process). *Let T be a tree, a probability measure in M^T is vertex-Markov if conditioned on the value at a vertex, the processes on the pending subtrees of that vertex are independent.*

Let $\mathcal{I}_e(M)$ denote the set of probability measures on M^{E_1} that are invariant by switching the two sides of the edge $\{o, 1\}$. If $\mu \in \mathcal{I}_d(\mu)$ then its restriction to E_1 is in $\mathcal{I}_e(M)$. Conversely, if $p \in \mathcal{I}_e(M)$, there exists a unique vertex-Markov process whose restriction to E_1 is p. We denote the law of this process by $\mu(p)$. If $X \in \mathcal{I}_d(M)$, we define

$$\Sigma_e(X) = \frac{d}{2}H(X_{E_1}) - (d-1)H(X_o).$$

If $p \in \mathcal{I}_e(M)$, we set $\Sigma(p) = \Sigma_e(\mu(p))$. Vertex-Markov processes satisfy the following extremal property.

Lemma 1.14. *If $X \in \mathcal{I}_d(M)$ then*

$$\Sigma_1(X) \leq \Sigma_e(X),$$

with equality if and only if X_{S_1} is a vertex-Markov process on S_1.

Combined with Theorem 1.11, the above lemma implies the following corollary.

Theorem 1.15. *Let $p \in \mathcal{I}_e(M)$. If for all couplings $q \in \mathcal{I}_e(M^2)$ of p and p we have $\Sigma(q) \leq 2\Sigma(p)$, then $\underline{h}(\mu(p)) = \bar{h}(\mu(p)) = \Sigma(p)$ and $\mu(p)$ is strongly typical.*

Proof. Let $p' = \mu(p)_1 \in \mathcal{I}_{d,1}(M)$ be the law of $\mu(p)$ restricted to S_1. Let q' be an invariant coupling of p' and p' and let q be its restriction to E_1. By construction $q \in \mathcal{I}_e(M^2)$. Moreover, by Lemma 1.14, $\Sigma(q') \leq \Sigma_1(\mu(q')) = \Sigma_e(\mu(q)) = \Sigma(q)$. It follows that Theorem 1.15 is a consequence of Theorem 1.11 applied to $r = 1$ and $p' = \mu(p)_1$. □

1.6. **Application to factor graphs and combinatorial optimization.** In this paragraph, we discuss a basic connection between asymptotic free energy of factor graphs and sofic entropy. This may serve an extra motivation for studying the sofic entropy.

Let M be a finite set, $r \geq 1$ be an integer and let φ be a function on the set of rooted unlabeled M-colored graphs of radius r taking value in $(0, \infty)$. If $G \in \mathcal{G}_n(d)$,

$$Z_G = \sum_{f \in M^n} \prod_{v=1}^n \varphi((G, f, v)_r),$$

where $(G, f, v)_r$ is the rooted colored graph associated to the ball of radius r around v in G. We set

$$\psi = \ln \varphi.$$

The asymptotic free energy is defined as the limit of $(1/n) \ln Z_{G_n}$ where G_n is a uniformly sampled graph in $\mathcal{G}_n(d)$ (provided that the limit exists). By standard concentration inequality (see argument in Theorem 2.1), it is easy to check that if G_n is uniformly sampled in $\mathcal{G}_n(d)$, then, in probability, as n goes to infinity,

$$\frac{1}{n} \ln Z_{G_n} - \mathbb{E}\frac{1}{n} \ln Z_{G_n} \to 0. \tag{1.5}$$

It is straightforward to express the limits of the expected free energy in terms of the entropy. If $p \in \mathcal{I}_{d,r}(M)$, we set for ease of notation $\underline{h}(p) = \underline{h}(\mu(p), r)$ and similarly for $\bar{h}(p)$ (there are the upper and lower growth rates of the number of colorings of G_n whose r-neighborhood is close to p). In the statement below, we use the notation $\langle p, \psi \rangle = \mathbb{E}\psi(X)$ where X has law p.

Lemma 1.16. *For integer $r \geq 1$ and ψ as above, if G_n is uniformly distributed on $\mathcal{G}_n(d)$ (with nd even and $n \geq d+1$), we have*

$$\sup_{p \in \mathcal{I}_{d,r}(M)} (\underline{h}(p) + \langle p, \psi \rangle) \leq \liminf_{n \to \infty} \mathbb{E}\frac{1}{n} \ln Z_{G_n} \leq \limsup_{n \to \infty} \mathbb{E}\frac{1}{n} \ln Z_{G_n} \leq \sup_{p \in \mathcal{I}_{d,r}(M)} (\bar{h}(p) + \langle p, \psi \rangle).$$

In particular, if Conjecture 1.4 holds true, then it would automatically imply the convergence of the expected free energy for all functions ψ. Note also that Theorem 1.7 can be used to obtain a lower bound expected free energy while Theorem 1.5 can be used to get an upper bound. With proper technical conditions, it is possible to extend Lemma 1.16 to some hard-constrained models, that is to some functions $\varphi = e^\psi$ which take value in $[0, \infty)$. For simplicity, we will however not discuss in details this possibility here.

In the same vein, in combinatorial optimization problems, we are often interested in the computation of a graph functional of the form:

$$L_G = \max_{f \in M^n} \sum_{v=1}^n \psi((G, f, v)_r),$$

with $r \geq 1$ and ψ as above. Again, it is easy to check that if G_n is uniformly sampled in $\mathcal{G}_n(d)$, we have, in probability,

$$\frac{L_{G_n}}{n} - \frac{\mathbb{E}L_{G_n}}{n} \to 0.$$

The following statement is a corollary of Lemma 1.16. It shows that typical processes are intimately connected to the computation of limits of $\mathbb{E}L_{G_n}/n$.

Lemma 1.17. *For integer $r \geq 1$ and ψ as above, if G_n is uniformly distributed on $\mathcal{G}_n(d)$ (with nd even and $n \geq d+1$), we have*

$$\sup_{p \in \mathcal{I}_{d,r}(M):\underline{h}(p) \geq 0} \langle p, \psi \rangle \leq \liminf_{n \to \infty} \frac{\mathbb{E}L_{G_n}}{n} \leq \limsup_{n \to \infty} \frac{\mathbb{E}L_{G_n}}{n} \leq \sup_{p \in \mathcal{I}_{d,r}(M):\bar{h}(p) \geq 0} \langle p, \psi \rangle.$$

Again, we observe that Theorem 1.7 can be used to obtain a lower bound on $\mathbb{E}L_{G_n}/n$ and Theorem 1.5 an upper bound.

Remark 1.18. We conclude this paragraph by mentioning that, for $r = 1$, there is a simplification of Lemma 1.16 and Lemma 1.17 for functions ψ of the form

$$\psi((G, f, v)_1) = \psi_0(f(v)) + \sum_{u:u \sim v} \psi_1(f(v), f(u))$$

where the supremum in Lemma 1.16 and Lemma 1.17 is taken over $p \in \mathcal{I}_e(M)$ instead of $p \in \mathcal{I}_{d,1}(M)$ and the entropic term is given by $\underline{h}(p) = \sup \underline{h}(q)$ where the supremum is over all $q \in \mathcal{I}_{d,1}(M)$ whose restriction to E_1 is p, and similarly for $\bar{h}(p)$. This can be useful because it reduces the dimension of the underlying optimization problem. In that case, Theorem 1.15 can be used to give lower bounds.

Organization of the paper. The remainder of this text is organized as follows. In Section 2, we will establish the key properties of sofic and annealed entropies. In Section 3, we will prove the main results of this paper. In the final Section 4, we will extend our framework and main results to invariant processes on unimodular Galton-Watson trees.

2. PROPERTIES OF SOFIC AND ANNEALED ENTROPIES

2.1. **Concentration of entropy: proof of Lemma 1.2.** Let G_n be a uniformly distributed random graph on $\mathcal{G}_n(d)$ with $n \geq d-1$ and nd even. Recall the definition of $H_{G_n}(\mu, r, \epsilon)$ in (1.3). The aim of this subsection is to establish the following concentration result.

Theorem 2.1. *Let $r \geq 0$, $\mu \in \mathcal{I}_d(M)$, $h \in \{-\infty\} \cup [0, \infty)$. For all functions $\delta : (0, \infty) \to (0, \infty]$ with $\lim_{\epsilon \to 0} \delta(\epsilon) = 0$, we have the following: if for all $\epsilon > 0$,*

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(H_{G_n}(\mu, r, \epsilon) \geq h) \geq -\delta(\epsilon), \tag{2.1}$$

then $\bar{h}(\mu, r, \alpha) \geq h$ for all $0 < \alpha < 1$. Conversely, for all functions δ as above such that $\lim_{\epsilon \to 0} \epsilon^{-2}\delta(\epsilon) = 0$: if for all $\epsilon > 0$,

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(H_{G_n}(\mu, , r, \epsilon) \leq h) \geq -\delta(\epsilon) \tag{2.2}$$

then $\bar{h}(\mu, r, \alpha) \leq h$ for all $0 < \alpha < 1$. Finally, the same claims hold with $\liminf$ and $\underline{h}$ replacing $\limsup$ and $\bar{h}$.

It is immediate to check that Lemma 1.2 is a corollary of Theorem 2.1. Beware of the asymmetry between the lower and upper bound. We believe that it is a caveat of our proof. It is ultimately due to the fact that $H_{G_n}(\mu, r, \epsilon)$ can be equal to $-\infty$.

The proof of Theorem 2.1 makes a detour through a relaxation of the entropy. Fix $\mu \in \mathcal{I}_d(M)$ and $r \geq 0$. If G is in $\mathcal{G}_d(n)$, we define for $\beta > 0$,

$$Z_G(\beta) = \sum_{f \in M^n} e^{-n\beta \mathrm{d}(\mathrm{distr}_G(f)_r, \mu_r)}.$$

We start the proof of the proposition with a concentration inequality.

Lemma 2.2. *Let G_n be uniformly distributed on $\mathcal{G}_n(d)$ with dn even and $n \geq d-1$. There exists a constant C depending on (d, r) and a deterministic number $s_n(\beta)$ depending on (n, d, r, β) such that for any $t > 0$, we have*

$$\mathbb{P}\left(\left| \frac{1}{n} \ln Z_{G_n}(\beta) - s_n(\beta) \right| \geq t \right) \leq C \exp(-nt^2/(C\beta)^2).$$

Proof. The proof follows a standard path. By classical contiguity results, it is enough to establish the claim for the configuration model (see Bollobás [10, Section 2.4]). Recall that the configuration model is the graph (with possible loops and multiple edges) obtained as follows. We attach to each vertex in $[n]$, d half-edges. We sample a matching m on the set $\vec{E}$ of nd half-edges uniformly at random (recall that a matching is an involution without fixed point). Finally, we form a d-regular graph $G = G(m)$ by creating an edge for each pair of matched half-edges.

Let us say that two matchings m, m' differ by a switch if there exists (a, b, c, d) in $\vec{E}$ such that $m(e) = m'(e)$ for all $e \in \vec{E} \backslash \{a, b, c, d\}$ and $m(a) = b$, $m'(a) = c$, $m(c) = d$, $m'(b) = d$. If m and m' differ by a switch, we claim that for any $f \in M^n$,

$$d(\mathrm{distr}_{G(m)}(f)_r, \mathrm{distr}_{G(m')}(f)_r) \leq C_0 \frac{4d(d-1)^{r-1}}{n} = \frac{\theta}{n},$$

where C_0 is the diameter of $\mathcal{P}(\mathcal{G}_M^\bullet)$ for the distance d. Indeed, we have $\mathrm{distr}_{G,v}(f)_r = \mathrm{distr}_{G',v'}(f')_r$ if the rooted subgraphs $(G, f, v)_r$ and $(G', f', v')_r$ are isomorphic. Notably, $\mathrm{distr}_{G(m),v}(f)_r = \mathrm{distr}_{G(m'),v}(f)_r$ unless v is at distance at most r from an edge in the symmetric difference of $G(m)$ and $G(m')$.

We deduce that

$$e^{-\beta\theta} \leq \frac{Z_{G(m')}}{Z_{G(m)}} \leq e^{\beta\theta}$$

and

$$\left| \ln Z_{G(m)} - \ln Z_{G(m')} \right| \leq \beta\theta.$$

From [32, Theorem 2.19], if m is a uniformly sampled matching on $\vec{E}$, we get

$$\mathbb{P}\left(\left| \frac{1}{n} \ln Z_{G(m)} - \mathbb{E}\frac{1}{n} \ln Z_{G(m)} \right| \geq t \right) \leq 2 \exp(-nt^2/(2d\theta^2\beta^2)).$$

The conclusion follows with $s_n(\beta) = \mathbb{E}\frac{1}{n} \ln Z_{G(m)}(\beta)$. $\qquad\square$

We are ready for the proof of Theorem 2.1.

Proof of Theorem 2.1. Recall the definition of $\mathcal{F}_G(\mu, r, \epsilon)$ in (1.2). Since μ and r are fixed, we write simply $\mathcal{F}_G(\epsilon)$ and set $F_G(\epsilon) = |\mathcal{F}_G(\epsilon)|$. For any $\epsilon > 0$, we have

$$\ln Z_G(\beta) \geq \ln F_G(\epsilon) - n\beta\epsilon, \tag{2.3}$$

where we have used that $d(\mathrm{distr}_G(f)_r, \mu_r) \leq \epsilon$ for all $f \in \mathcal{F}_G(\epsilon)$. The other way around, if $f \notin \mathcal{F}_G(\epsilon)$, we have $d(\mathrm{distr}_G(f)_r, \mu_r) \geq \epsilon$. Hence,

$$\ln Z_G(\beta) \leq \ln \left(|M|^n e^{-n\beta\epsilon} + F_G(\epsilon) \right) \leq \ln 2 + (\ln F_G(\epsilon)) \vee (n(\ln |M| - \beta\epsilon)). \tag{2.4}$$

We may now prove the first claim of the theorem. Assume that (2.1) holds for some $h \geq 0$ (if $h = -\infty$, there is nothing to prove). Let $h_1 < h$ and $\mathcal{E} = \{H_{G_n}(\mu, r, \epsilon) \geq h\}$. On the event $\mathcal{E}$, from (2.3), we have for all $\beta > 0$,

$$\frac{1}{n} \ln Z_{G_n}(\beta) \geq h - \beta\epsilon.$$

There exists β_ϵ such that, as $\epsilon \to 0$, $\beta_\epsilon \epsilon \to 0$, $\beta_\epsilon \to \infty$ and $\beta_\epsilon^2 \delta(\epsilon) \to 0$. For this choice of β, $(t_\epsilon/\beta_\epsilon)^2 \gg \delta(\epsilon)$ for some $t_\epsilon \to 0$. It follows from Lemma 2.2 and (2.1) that for any h_2, h_3 such that $h_1 < h_2 < h_3 < h$, $s_n(\beta_\epsilon) \geq h_3$ for all n large enough (depending on ϵ), because the event $\mathcal{E}$ overlaps with the event from Lemma 2.2 when β_ϵ satisfies this condition. Let $0 < \alpha < 1$. Applying again Lemma 2.2, we deduce that the event $\mathcal{E}_2 = \{\frac{1}{n} \ln Z_{G_n}(\beta_\epsilon) \geq h_2\}$ has probability greater than α for all n large enough.

Now, there exists η_ϵ such that, as $\epsilon \to 0$, $\eta_\epsilon \to 0$ and $\beta_\epsilon \eta_\epsilon \to \infty$ (for example $\eta_\epsilon = 1/\sqrt{\beta_\epsilon}$). We apply (2.4) with $\epsilon = \eta_\epsilon$. We get on the event $\mathcal{E}_2$, if ϵ is small enough,

$$H_{G_n}(\mu, r, \eta_\epsilon) \geq \frac{1}{n} \ln Z_{G_n}(\beta_\epsilon) - \frac{1}{n} \ln 2 \geq h_2 - \frac{1}{n} \ln 2.$$

The right-hand side is larger than h_1 if n is large enough. We deduce that $\bar{h}(\mu, r, \alpha) \geq h_1$, since h_1 can be arbitrarily close to h, the first claim follows.

The second claim is proven similarly. Since $H_{G_n}(\mu, r, \epsilon)$ takes value in $\{-\infty\} \cup [0, \infty)$, we have $H_{G_n}(\mu, r, \epsilon) \leq -1$ if and only if $H_{G_n}(\mu, r, \epsilon) = -\infty$. We may thus prove the second claim with $h \geq -1$. Assume that (2.2)

holds for some $\delta(\epsilon)$ which will be defined later on. We now set $\mathcal{E} = \{H_{G_n}(\mu, r, \epsilon) \leq h\}$. On the event $\mathcal{E}$, from (2.4), we have for all $\beta > 0$,

$$\frac{1}{n} \ln Z_{G_n}(\beta) \leq \frac{1}{n} \ln 2 + h \vee (\ln|M| - \beta \epsilon).$$

If $\beta \geq \beta_\epsilon = (\ln|M| + 1)/\epsilon$ then we get, for all n large enough

$$\frac{1}{n} \ln Z_{G_n}(\beta) \leq \frac{1}{n} \ln 2 + h.$$

For our choice of $\delta(\epsilon)$, we have $\delta(\epsilon)\beta_\epsilon^2 \to 0$ as $\epsilon \to 0$. Then, if $h_1 > h_2 > h_3 > h$ and ϵ is small enough, we deduce by Lemma 2.2 that

$$s_n(\beta_\epsilon) \leq h_3$$

for all n large enough. We apply again Lemma 2.2 and deduce that the event $\mathcal{E}_2 = \{\frac{1}{n} \ln Z_{G_n}(\beta_\epsilon) \leq h_2\}$ has probability greater than α for all n large enough. Finally, from (2.3), we have on the event $\mathcal{E}_2$,

$$H_{G_n}(\mu, r, \epsilon^2) \leq h_2 + \beta_\epsilon \epsilon^2$$

The latter is less than h_1 for all ϵ small enough. The second claim follows. Obviously, the same argument works with $\liminf$ and $\underline{h}(\mu, r, \alpha)$. $\qquad\square$

2.2. Maximizers of the annealed entropy: proofs of Lemma 1.12 and Lemma 1.14.
In this subsection, we prove Lemma 1.12 and Lemma 1.14. If X, Y are discrete random variables, we recall that the relative entropy of X given Y is

$$H(X|Y) = -\sum_{x,y} \mathbb{P}((X,Y) = (x,y)) \ln \mathbb{P}(X = x | Y = y),$$

where $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$ is the usual conditional probability (if $\mathbb{P}(B) = 0$, $\mathbb{P}(A|B)$ takes an arbitrary value). In other words, $H(X|Y)$ is the average over Y of the entropy of the conditional law of X given Y. We will repeatedly use that

$$H(X,Y) = H(Y) + H(X|Y) \quad \text{and} \quad H(X|(Y,Y')) \leq H(X|Y), \tag{2.5}$$

with equality if and only if X conditioned on Y is independent of Y'.

We start with the proof of Lemma 1.12.

Proof of Lemma 1.12. The following fact is useful. For a given integer $r \geq 1$, we introduce the finite set $N = M^{S_{r-1}}$ where as usual $S_{r-1} = B_{r-1}(o)$. We consider the map Ψ from M^{T_d} to N^{T_d} which maps x to $\Psi(x)$ such that for $v \in T_d$, $\Psi(x)_v$ is the restriction of x to $S_{r-1}(v)$ (composed by a given isomorphism from $S_{r-1}(v)$ to S_{r-1}). If X is a process on T_d then for all integers $t \geq 0$, we have $\Sigma_{t+r}(X) = \Sigma_{t+1}(\Psi(X))$. Moreover, if X is a r-Markov process, then $\Psi(X)$ is an edge-Markov process.

As a byproduct, it is sufficient to prove Theorem 1.12 with $r = 1$: we should check that

$$\Sigma_2(X) \leq \Sigma_1(X)$$

with equality if and only if X_{S_2} is an edge Markov process. Note that the above inequality can be equivalently written as

$$H(X_{S_2}) - H(X_{S_1}) - \frac{d}{2} H(X_{E_2}) + \frac{d}{2} H(X_{E_1}) \leq 0. \tag{2.6}$$

To check that (2.6) holds, we need some extra notation. We denote by $L = \{2, \ldots, d\}$ the left side of E_2 along $E_1 = \{o, 1\}$. We also set $L_i = \{(i, 1), \ldots, (i, d-1)\}$ with $i = 1, \ldots, d$. We have $S_1 = L \cup E_1$ and thus, from (2.5)

$$H(S_1) = H(E_1) + H(L|E_1),$$

where for ease of notation, for sets S, T, we write $H(S)$ and $H(S|T)$ in place of $H(X_S)$ and $H(X_S|X_T)$. Similarly, since $E_2 = S_1 \cup L_1$,

$$H(E_2) = H(S_1) + H(L_1|S_1) = H(E_1) + H(L|E_1) + H(L_1|S_1).$$

Finally, since S_2 is the disjoint union of E_2 and $\cup_{i=2}^d L_i$, we have,

$$H(S_2) = H(E_2) + H(\cup_{i=2}^d L_i | E_2).$$

The last three identities imply that Equation (2.6) is equivalent to

$$H(\cup_{i=2}^d L_i | E_2) - \left(\frac{d}{2} - 1\right) H(L_1 | S_1) - \frac{d}{2} H(L | E_1) \le 0. \tag{2.7}$$

Using the invariance, we deduce from (2.5) that

$$H(\cup_{i=2}^d L_i | E_2) \le (d-1) H(L_2 | E_2)$$

with equality if and only if there is conditional independence of the X_{L_i}'s given X_{E_2}. Now, since E_2 contains S_1, we get

$$H(L_2 | E_2) \le H(L_2 | S_1) = H(L_1 | S_1),$$

with equality in case of conditional independence of X_{L_2} and $X_{E_2 \setminus S_1}$ given X_{S_1}. It follows that the left-hand side of (2.7) is upper bounded by

$$\frac{d}{2} H(L_1 | S_1) - \frac{d}{2} H(L | E_1). \tag{2.8}$$

From the invariance of X by switching the two sides of e, we get $H(L | E_1) = H(L_1 | E_1)$ and thus (2.8) is equal to

$$\frac{d}{2} (H(L_1 | S_1) - H(L_1 | E_1)).$$

Using again (2.5), since $E_1 \subset S_1$, this last expression is always non-positive with equality if and only if X_{L_1} is conditionally independent of X_{S_1} given X_e. This proves that (2.6) holds. By considering the case of equality, it is then easy to check that it implies that X_{S_2} is an edge Markov process. It concludes the proof of Lemma 1.12. $\qquad \square$

We now prove Lemma 1.14.

Proof of Lemma 1.14. Let $X \in \mathcal{I}_d(M)$. From (2.5), with the notation used in the proof of Lemma 1.12, we have

$$H(S_1) = H(o) + H(S_1 | o) \le H(o) + d H(E_1 | o),$$

with equality if the variables $(X_o, X_i)_{1 \le i \le d}$ conditioned on X_o are independent. Using (2.5) again, we get

$$H(S_1) \le d H(E_1) - (d-1) H(o).$$

So finally, $\Sigma_1(X) = H(S_1) - \frac{d}{2} H(E_1) \le \frac{d}{2} H(E_1) - (d-1) H(o) = \Sigma_e(X)$ as requested. $\qquad \square$

2.3. **Combinatorial characterization of the annealed entropy.** In this subsection, we give a combinatorial interpretation of the annealed entropy $\Sigma_r(\mu)$. Recall that $\mathcal{G}_n(d)$ is the set of simple d-regular graphs on the vertex set $[n]$. For $\mu \in \mathcal{I}_d(M)$, $r \ge 0$ integer and $\epsilon > 0$, we define the set of colored graphs whose r-neighborhood is close to μ_r as

$$\begin{aligned}
\mathcal{G}_n(\mu, r, \epsilon) &= \{(G, f) : G \in \mathcal{G}_n(d), f \in M^n, d(\mathrm{distr}_G(f)_r, \mu_r) \le \epsilon\} \\
&= \bigsqcup_{G \in \mathcal{G}_n(d)} \mathcal{F}_G(\mu, r, \epsilon),
\end{aligned}$$

where $\sqcup$ is the disjoint union and $\mathcal{F}_G(\mu, r, \epsilon)$ was defined in (1.2). We then set

$$\Sigma_n(\mu, r, \epsilon) = \frac{1}{n}(\log |\mathcal{G}_n(\mu, r, \epsilon)| - \log |\mathcal{G}_n(d)|) = \frac{1}{n} \log \mathbb{E}|\mathcal{F}_{G_n}(\mu, r, \epsilon)|, \tag{2.9}$$

where the expectation is with respect to the random graph G_n uniformly distributed on $\mathcal{G}_n(d)$. In comparison with the definition of $H_{G_n}(\mu, r, \epsilon)$ in (1.3), $\Sigma_n(\mu, r, \epsilon)$ appears as an annealed quantity in the sense that there is an average over the randomness of G_n inside the logarithm. The following theorem asserts that $\Sigma_n(\mu, r, \epsilon)$ is close to $\Sigma_r(\mu)$ as n goes to infinity and ϵ goes to 0.

Theorem 2.3. *Let $\mu \in \mathcal{I}_d(M)$ and $r \ge 1$ integer. We have*

$$\lim_{\epsilon \to 0} \liminf_{n \to \infty} \Sigma_n(\mu, r, \epsilon) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \Sigma_n(\mu, r, \epsilon) = \Sigma_r(\mu).$$

Proof. One side of this identity can be found in [3, Lemma 6.2]. We will however give a proof which relies on [16] which is a generalization of [12] to colored graphs. This is interesting because it connects [12, 16] to the entropic inequalities found in [2, 3]. First, a classical result of Bender and Canfield [7] implies that

$$\frac{1}{n} \log |\mathcal{G}_n(d)| = \frac{d}{2} \log n - s(d) - \log(d!) + o(1), \tag{2.10}$$

where $s(d) = d/2 - (d/2) \log d$. On the other hand, Proposition 5 and Proposition 6 in Delgosha and Anantharam [16] imply that

$$\lim_{\epsilon \to 0} \liminf_{n \to \infty} \left(\frac{1}{n} \log |\mathcal{G}_n(\mu, r, \epsilon)| - \frac{d}{2} \log n \right) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \left(\frac{1}{n} \log |\mathcal{G}_n(\mu, r, \epsilon)| - \frac{d}{2} \log n \right) = J_r(\mu),$$

where $J_r(\mu)$ has an explicit formula that we now describe (the same formula appears in [12]).

We define $\widetilde{\mathcal{T}}_{r-1}^\bullet$ as the set of unlabeled colored rooted $(d-1)$-ary trees of depth $r-1$. An element $g = (t, t') \in \widetilde{\mathcal{E}}_r = \widetilde{\mathcal{T}}_{r-1}^\bullet \times \widetilde{\mathcal{T}}_{r-1}^\bullet$ can be seen as an unlabeled coloring of E_r rooted on the oriented edge $(o, 1)$. For $g = (t, t')$ in $\widetilde{\mathcal{E}}_r$ and X a coloring of T_d, we then define $N_X(g)$ as the number of neighbors v of the root such that X restricted to $E_r(o, v) = B_{r-1}(\{o, v\})$ is isomorphic to g: more precisely such that the restriction of X to $(d-1)$-ary tree rooted at o (respectively v) in $E_r(o, v) \backslash \{o, v\}$ is isomorphic to t (respectively t'). By construction

$$\sum_{g \in \widetilde{\mathcal{E}}_r} N_X(g) = \deg(o) = d. \tag{2.11}$$

If X is a random coloring of T_d with μ, we then define a probability measure on $\widetilde{\mathcal{E}}_r$ by, for all $g \in \widetilde{\mathcal{E}}_r$:

$$\pi_\mu(g) = \frac{\mathbb{E}[N_X(g)]}{d},$$

where the expectation is with respect to the randomness of X. We have

$$J_r(\mu) = -s(d) + H(\widetilde{X}_{S_r}) - \frac{d}{2} H(\pi_\mu) - \sum_{g \in \widetilde{\mathcal{E}}_r} \mathbb{E}[\log(N_X(g)!)],$$

where $\widetilde{X}_{S_r}$ is the rooted unlabeled coloring associated to X_{S_r}. As a sanity check, if M is a singleton, then $J_r(\mu) = -s(d) - \log(d!)$ and we retrieve Equation (2.10). Moreover, in view of Equation (2.10), the theorem follows from the claim

$$J_r(\mu) = -s(d) - \log(d!) + H(X_{S_r}) - \frac{d}{2} H(X_{E_r}). \tag{2.12}$$

The expression (2.12) is obtained by putting random labeling on an unlabeled rooted coloring and following the effect on the Shannon entropy. We first observe that, since X is invariant, for any $g \in \widetilde{\mathcal{E}}_r$, we have

$$\mathbb{P}(X_{E_r} \simeq g) = \frac{1}{d} \sum_{v=1}^{d} \mathbb{P}(X_{E_r(o,v)} \simeq g) = \pi_\mu(g).$$

It follows that π_μ is the law of $\widetilde{X}_{E_r}$ defined as the unlabeled coloring associated to X_{E_r} rooted at the oriented edge $(o, 1)$. Besides, since X is invariant, X_{E_r} is in one-to-one correspondence with the triple $(\widetilde{X}_{E_r}, \sigma, \sigma')$ where, given $\widetilde{X}_{E_r}$, σ and σ' are independent and σ is a uniform random labeling of $\widetilde{X}_{E_r}$ restricted to E_r^o, the $(d-1)$-ary tree rooted at o in $E_r \backslash \{o, 1\}$, and similarly for σ'. From the relative entropy identity (2.5), we find that

$$H(X_{E_r}) = H(\pi_\mu) + 2K,$$

where K is the relative entropy of σ given $\widetilde{X}_{E_r^o}$.

Secondly, we observe that X_{S_r} is in one-to-one correspondence with the vector $Y = (X_{E_r^1}, \ldots, X_{E_r^d})$ where E_r^k is the $(d-1)$-ary tree rooted at k in $E_r \backslash \{o, k\}$. It follows that $H(Y) = H(X_{S_r})$. Also, if $\widetilde{Y} = (\widetilde{X}_{E_r^1}, \ldots, \widetilde{X}_{E_r^d})$, we find from what precedes and the invariance of X that

$$H(X_{S_r}) = H(Y) = H(\widetilde{Y}) + dK.$$

Finally, the difference between $\widetilde{Y}$ and $\widetilde{X}_{S_r}$ is that the neighbors of o are ordered (or labeled) in $\widetilde{Y}$. We deduce from Lemma 2.4 below that

$$H(\widetilde{Y}) = H(\widetilde{X}_{S_r}) - \sum_{g \in \widetilde{\mathcal{E}}_r} \mathbb{E}[\log(N_X(g)!)] + \log(d!).$$

This concludes the proof of (2.12). $\qquad\square$

In the proof of Theorem 2.3, we have used the following elementary lemma. Recall that a vector is exchangeable if its law is invariant by any permutation of its coordinates.

Lemma 2.4. *Let F be a finite set and $Z = (Z_1, \ldots, Z_n)$ a random exchangeable vector in F^n. The counting measure $N_Z = \sum_{i=1}^{n} \delta_{Z_i}$ associated to Z satisfies*

$$H(Z) = H(N_Z) - \sum_{x \in F} \mathbb{E}[\log N_Z(x)!] + \log n!.$$

Proof. We consider the equivalence class on F^n, $z \sim z'$ if z and z' are equal up to a permutation of the coordinates of z. We have $z \sim z'$ if and only if $N_z = N_{z'}$. Moreover, the number of vectors in the equivalence class of z is given by the multinomial formula:

$$\frac{n!}{\prod_{x \in F} N_z(x)!}.$$

Using the exchangeability of Z, we deduce that

$$\mathbb{P}(Z = z) = \frac{\prod_{x \in F} N_z(x)!}{n!} \sum_{z' \sim z} \mathbb{P}(Z = z') = \frac{\prod_{x \in F} N_z(x)!}{n!} \mathbb{P}(N_Z = N_z).$$

It then remains to use the relative entropy formula (2.5). $\qquad\square$

3. Proofs of main results

3.1. **First moment method: proof of Theorem 1.5.** Let $r \geq 1$, $\epsilon > 0$ and G_n be uniformly sampled on $\mathcal{G}_n(d)$. From Markov inequality, for any real h,

$$\mathbb{P}(H_{G_n}(\mu, r, \epsilon) \geq h) = \mathbb{P}(|\mathcal{F}_{G_n}(\mu, r, \epsilon)| \geq e^{nh}) \leq e^{-nh}\mathbb{E}|\mathcal{F}_{G_n}(\mu, r, \epsilon)|.$$

In particular, we find

$$\frac{1}{n} \log \mathbb{P}(H_{G_n}(\mu, r, \epsilon) \geq h) \leq \Sigma_n(\mu, r, \epsilon) - h.$$

From Theorem 2.3, we deduce the large deviations bound

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(H_{G_n}(\mu, r, \epsilon) \geq h) \leq \Sigma_r(\mu) - h + \delta(\epsilon),$$

where $\delta(\epsilon)$ goes to 0 as $\epsilon \to 0$. If $h > \Sigma_r(\mu)$, the right-hand side of the above expression is negative for all ϵ small enough. We deduce in particular that for all ϵ small enough, $\mathbb{P}(H_{G_n}(\mu, r, \epsilon) \geq h)$ converges to 0. By Lemma 1.2, this proves that $\overline{h}(\mu, r) < h$. It concludes the proof of Theorem 1.5. $\qquad\square$

3.2. **Second moment method: proof of Theorem 1.7.** Let $\mu \in \mathcal{I}_d(M)$, $r \geq 1$ and set $p = \mu_r \in \mathcal{I}_{d,r}(M)$. In view of Theorem 1.5, we should prove that

$$\underline{h}(\mu, r) \geq \Sigma(p). \tag{3.1}$$

For ease of notation, we write $\mathcal{F}_G(p, \epsilon)$ in place of $\mathcal{F}_G(\mu, r, \epsilon)$ (since this depends of μ only through $\mu_r = p$). The Paley-Zygmund inequality implies that

$$
\begin{aligned}
\mathbb{P}\left(H_{G_n}(\mu, r, \epsilon) \geq \Sigma_n(\mu, r, \epsilon) - \frac{1}{n} \right) &= \mathbb{P}\left(|\mathcal{F}_{G_n}(p, \epsilon)| \geq e^{-1}\mathbb{E}|\mathcal{F}_{G_n}(p, \epsilon)|\right) \\
&\geq (1 - e^{-1})^2 \frac{(\mathbb{E}|\mathcal{F}_{G_n}(p, \epsilon)|)^2}{\mathbb{E}|\mathcal{F}_{G_n}(p, \epsilon)|^2} \\
&= (1 - e^{-1})^2 \frac{\exp(2n\Sigma_n(\mu, r, \epsilon))}{\mathbb{E}|\mathcal{F}_{G_n}(p, \epsilon)|^2}.
\end{aligned}
$$

12

Since $\mu_r = p$, we have $\Sigma(p) = \Sigma_r(\mu)$ and, by Theorem 2.3,

$$\liminf_{n \to \infty} \Sigma_n(\mu, r, \epsilon) \geq \Sigma(p) - \delta(\epsilon),$$

where $\delta(\epsilon)$ goes to 0 as $\epsilon \to 0$. We deduce that if we manage to prove that

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{E}|\mathcal{F}_{G_n}(p, \epsilon)|^2 \leq 2\Sigma(p) + \delta'(\epsilon), \tag{3.2}$$

where $\delta'(\epsilon)$ goes to 0 as $\epsilon \to 0$, then we would get that

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}(H_{G_n}(\mu, r, \epsilon) \geq \Sigma(p) - 2\delta(\epsilon)) \geq -2\delta(\epsilon) - \delta'(\epsilon).$$

From Equation (2.1) in Theorem 2.1, this would imply that $\underline{h}(\mu, r) \geq \Sigma(p)$ as claimed in (3.1).

It thus remains to prove Equation (3.2). For concreteness, we may assume that the chosen distance d generating the weak topology is the total variation distance. To that end, let $\epsilon > 0$ and $\mathcal{N}_\varepsilon$ be an ε-net on the set of invariant coupling $q \in \mathcal{I}_{d,r}(M^2)$ of p and p. Given a graph $G \in \mathcal{G}_n(d)$, consider two colorings of G with color set M whose r-neighborhood statistics are at most at total variation distance ε from p. The number of such pairs is $|\mathcal{F}_G(p, \varepsilon)|^2$. On the other hand, each pair is in fact a coloring of G on M^2. Then its r-neighborhood statistics is an element $q' \in \mathcal{I}_{d,r}(M^2)$. Since both marginals of q' are at most at total variation distance ε from p, there is a measure $q^* \in \mathcal{I}_{d,r}(M^2)$ whose total variation distance is at most 2ε from q' and whose both marginals are exactly p (for each marginal of q', there is an invariant coupling of this marginal and p such that the two colorings are equal with probability $1 - \epsilon$). Therefore there exists an element in the ε-net, $q \in \mathcal{N}_\epsilon$ such that the distance of q from the original pair of coloring is at most 3ε. We conclude that

$$|\mathcal{F}_G(p, \varepsilon)|^2 \leq \sum_{q \in \mathcal{N}_\varepsilon} |\mathcal{F}_G(q, 3\varepsilon)|.$$

This implies that

$$\mathbb{E}|\mathcal{F}_{G_n}(p, \epsilon)|^2 \leq \frac{1}{|\mathcal{G}_n(d)|} \sum_{G \in \mathcal{G}_n(d)} \sum_{q \in \mathcal{N}_\varepsilon} |\mathcal{F}_G(q, 3\varepsilon)| = \sum_{q \in \mathcal{N}_\varepsilon} \exp(n\Sigma_n(\mu(q), r, 3\epsilon)).$$

It follows that,

$$\frac{1}{n} \log \mathbb{E}|\mathcal{F}_{G_n}(p, \epsilon)|^2 \leq \max_{q \in \mathcal{N}_\epsilon} \Sigma_n(\mu(q), r, 3\epsilon) + \frac{1}{n} \log |\mathcal{N}_\varepsilon|.$$

By Theorem 2.3, we deduce that, for some function $\delta'(\epsilon)$ going to 0 with $\epsilon \to 0$,

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{E}|\mathcal{F}_{G_n}(p, \epsilon)|^2 \leq \max_{q \in \mathcal{N}_\epsilon} \Sigma(q) + \delta'(\epsilon).$$

By assumption for any $q \in \mathcal{I}_{d,r}(M^2)$, $\Sigma(q) \leq 2\Sigma(p)$. We thus have proved that (3.2) holds. $\qquad \square$

3.3. **Proof of Theorem 1.11.** In view of Theorem 1.5 and Theorem 1.7, it remains to prove that for any $t \geq r$,

$$\underline{h}(\mu(p), t) \geq \Sigma_r(\mu(p)) = \Sigma(p).$$

Let $q \in \mathcal{I}_{d,t}(M^2)$ be an invariant coupling of $(\mu(p))_t$ and $(\mu(p))_t$. Then q_r is an invariant coupling of p and p (since $((\mu(p))_t)_r = \mu(p)_r = p$ by construction). By Lemma 1.12, we have

$$\Sigma(q) = \Sigma(\mu(q)) \leq \Sigma(\mu(\mu(q)_r)) = \Sigma(q_r).$$

By assumption $\Sigma(q_r) \leq 2\Sigma(p)$. However, by Lemma 1.12, we have $\Sigma(p) = \Sigma(\mu(p)_t)$. It follows that

$$\Sigma(q) \leq 2\Sigma(\mu(p)_t).$$

From Equation (3.1) applied to the radius t, this implies that

$$\underline{h}(\mu(p), t) \geq \Sigma(\mu(p)_t).$$

By a last application of Lemma 1.12, the right-hand side of above expression is equal to $\Sigma(p)$. This concludes the proof of Theorem 1.11. $\qquad \square$

3.4. Application to factor graphs: proofs of Lemmas 1.16 and Lemma 1.17.

We start with the proof of Lemma 1.16.

Proof of Lemma 1.16. By construction, we have

$$Z_G = \sum_{f \in M^n} \prod_{v=1}^{n} e^{\psi((G,f,v)_r)} = \sum_{f \in M^n} e^{n \langle \mathrm{distr}_G(f)_r, \psi \rangle}.$$

Let $\epsilon > 0$ and $\mathcal{N}_\varepsilon$ be an ε-net of $\mathcal{I}_{d,r}(M)$. The function $p \to \langle p, \psi \rangle$ being uniformly continuous (since M is finite), there exists a function $\delta(\epsilon) \to 0$ as $\epsilon \to 0$ such that for any probability measure, say q, on rooted colored graphs of radius r, if $\mathrm{d}(q,p) \leq \epsilon$ then $|\langle q, \psi \rangle - \langle p, \psi \rangle| \leq \delta(\epsilon)$. If N_G is the number of colorings such that $\mathrm{distr}_G(f)$ is at distance larger than ϵ from $\mathcal{N}_\varepsilon$, it follows that

$$
\begin{aligned}
Z_G &\leq \sum_{p \in \mathcal{N}_\varepsilon} |\mathcal{F}_G(\mu(p), r, \epsilon)| e^{n\langle p, \psi \rangle + n\delta(\epsilon)} + N_G e^{n\|\psi\|_\infty}, \\
&\leq |\mathcal{N}_\epsilon| \max_{p \in \mathcal{I}_{d,r}(M)} e^{n(H_G(\mu(p),r,\epsilon) + \langle p, \psi \rangle + \delta(\epsilon))} + N_G e^{n\|\psi\|_\infty}.
\end{aligned}
$$

Now, if G_n is uniformly sampled on $\mathcal{G}_n(d)$, then, for any fixed $\epsilon > 0$, $\mathbb{P}(N_{G_n} = 0)$ converges to 1 (since distr_{G_n} converges in probability to a Dirac mass at (T_d, o)). Using (1.5) and taking the limit in n, we find

$$\limsup_{n \to \infty} \mathbb{E} \frac{1}{n} \ln Z_{G_n} \leq \max_{p \in \mathcal{I}_{d,r}(M)} \left(\bar{h}(p) + \langle p, \psi \rangle + \delta'(\epsilon) \right)$$

with $\delta'(\epsilon) \to 0$ as $\epsilon \to 0$. This gives the upper bound in Lemma 1.16.

For the lower bound, we write similarly,

$$
\begin{aligned}
Z_G &\geq \sum_{p \in \mathcal{N}_\varepsilon} |\mathcal{F}_G(\mu(p), r, \epsilon)| e^{n\langle p, \psi \rangle - n\delta(\epsilon)}, \\
&\geq \max_{p \in \mathcal{N}_\varepsilon} e^{n(H_G(\mu(p),r,\epsilon) + \langle p, \psi \rangle - \delta(\epsilon))} \\
&\geq \max_{p \in \mathcal{I}_{d,r}(M)} e^{n(H_G(\mu(p),r,\epsilon) + \langle p, \psi \rangle - 2\delta(\epsilon))}.
\end{aligned}
$$

The conclusion follows easily. $\qquad\square$

Lemma 1.17 is a corollary of Lemma 1.16.

Proof of Lemma 1.17. For $\beta > 0$, let $Z_G(\beta)$ be the factor graph model:

$$Z_G(\beta) = \sum_{f \in M^n} \prod_{v=1}^{n} e^{n\beta\psi((G,f,v)_r)}.$$

By construction, we have

$$|M|^{-n} Z_G(\beta) \leq e^{\beta L_G} \leq Z_G(\beta).$$

By Lemma 1.16, we find

$$\sup_{p \in \mathcal{I}_{d,r}(M)} \left(\frac{\underline{h}(p)}{\beta} + \langle p, \psi \rangle - \frac{\ln|M|}{\beta} \right) \leq \liminf_n \frac{\mathbb{E} L_{G_n}}{n} \leq \limsup_n \frac{\mathbb{E} L_{G_n}}{n} \leq \sup_{p \in \mathcal{I}_{d,r}(M)} \left(\frac{\bar{h}(p)}{\beta} + \langle p, \psi \rangle \right).$$

We recall that $\bar{h}(p)$ and $\underline{h}(p)$ take value in $\{-\infty\} \cup [0, \ln|M|]$. We get

$$\sup_{p \in \mathcal{I}_{d,r}(M): \underline{h}(p) \geq 0} \left(\langle p, \psi \rangle - \frac{\ln|M|}{\beta} \right) \leq \liminf_n \frac{\mathbb{E} L_{G_n}}{n} \leq \limsup_n \frac{\mathbb{E} L_{G_n}}{n} \leq \sup_{p \in \mathcal{I}_{d,r}(M): \underline{h}(p) \geq 0} \left(\frac{\ln|M|}{\beta} + \langle p, \psi \rangle \right).$$

We obtain the statement of the lemma by taking the limit $\beta \to \infty$. $\qquad\square$

4. Extension to processes on unimodular Galton-Watson trees

4.1. An extended setting. We now discuss an extension to processes on random trees. We will focus our attention on *unimodular Galton-Watson trees*. In this section, we fix a probability measure π on integers with positive and finite expectation:

$$d = \sum_{k=0}^{\infty} k\pi(k) > 0.$$

We define $\hat{\nu}$, the size-biased version of ν as the probability measure defined by: for all integers $k \geq 0$

$$\hat{\pi}(k) = \frac{(k+1)\pi(k+1)}{d}.$$

Then, the unimodular Galton-Watson tree with degree distribution π, is the Galton-Watson tree whose vertex set is a subset of $\mathbb{N}^f$ defined in (1.1) such that the root o has a number of offsprings N_o with distribution π indexed by $1, \ldots, N_o$ and all other vertices v have an independent number of offsprings N_v with distribution $\hat{\pi}$ indexed by $(v, 1), \ldots, (v, N_v)$. We will denote by T a realization of this random tree and UGW(π) the law of the rooted tree (T, o). We note that (T, o) is randomly labeled in the sense defined in Subsection 1.1.

For example, if π is a Dirac mass at d then T is the d-regular tree. If π is a Poisson random variable with mean d, then $\hat{\pi} = \pi$ and T is a standard Galton-Watson tree with Poisson offspring distribution.

As its name suggests, the random rooted tree T is unimodular. Recall that a random rooted graph (G, o) is unimodular if for all non-negative functions f on the set of doubly rooted graphs (a connected graph with two ordered distinguished vertices) which are invariant by isomorphisms, we have

$$\mathbb{E} \sum_{v \in V} f(G, o, v) = \mathbb{E} \sum_{v \in V} f(G, v, o). \tag{4.1}$$

where V is the vertex set of G and the expectation is with respect to the randomness of (G, o).

If M is a finite set, an invariant process X on T is defined as a random colored tree (T, X) such that (T, X, o) is unimodular (that is, it satisfies (4.1) with $G = (T, X)$ and f defined on the set of doubly rooted colored graphs which are invariant by isomorphisms). We denote by $\mathcal{I}_\pi(M)$ the set of laws of (T, X) with X invariant colorings of T on the color set M.

Now, in order to define a relevant notion of sofic entropy, we need to choose the ensemble of finite graphs G_n such that distr$_{G_n}$ converges to UGW(π). A natural choice is the family of uniform random graphs with a given degree sequence. Let $d_n = (d_n(1), \ldots, d_n(n))$ be a sequence of integers, indexed by a subset of $\mathbb{N}$, whose sum is even and such that

$$\text{distr}_{d_n} = \frac{1}{n} \sum_{v=1}^{n} \delta_{d_n(v)}$$

converges weakly to π. For technical simplicity, we assume that the degree sequence is uniformly bounded: for some real Δ,

$$\sup_n \max_{1 \leq v \leq n} d_n(v) \leq \Delta. \tag{4.2}$$

Note in particular that (4.2) implies that the support of π is contained in $\{0, \ldots, \Delta\}$. From Erdős-Gallai Theorem [19], for all n large enough, the set $\mathcal{G}_n(d_n)$ of simple graphs with vertex set $[n] = \{1, \ldots, n\}$ such that for all $v \in [n]$, v has degree $d_n(v)$ is not empty. Under these conditions, if G_n is uniformly distributed on $\mathcal{G}_n(d_n)$ then almost surely distr$_{G_n}$ converges to UGW(π), see for example [30]. In the statements below, we will not repeat the above assumptions on the sequence (d_n).

For a given probability measure $\mu \in \mathcal{I}_\pi(M)$ (that is, μ is the law of an invariant coloring (T, X)), we can now reproduce the definition of weakly and typical processes and define the sofic entropy by taking limits of $H_{G_n}(\mu, r, \epsilon)$ defined in (1.3). We do not repeat the definitions since there are identical except that G_n is now a random graph uniformly distributed on $\mathcal{G}_n(d_n)$. For integer $r \geq 1$ and $0 < \alpha < 1$, we define the quantity $\bar{h}(\mu, r, \alpha)$ and $\underline{h}(\mu, r, \alpha)$ exactly as done below (1.3). Lemma 1.2 continues to holds in this more general setting.

Lemma 4.1. *Let $\mu \in \mathcal{I}_\pi(M)$ and $r \geq 0$. The function $\alpha \mapsto (\bar{h}(\mu, r, \alpha), \underline{h}(\mu, r, \alpha))$ is constant on $(0, 1)$.*

Proof. The proof of Theorem 2.1 works verbatim under the assumption (4.2). □

We define $\bar{h}(\mu, r)$ and $\underline{h}(\mu, r)$ as the common value of $\bar{h}(\mu, r, \alpha)$ and $\underline{h}(\mu, r, \alpha)$. The upper and lower sofic entropies $\bar{h}(\mu)$ and $\underline{h}(\mu)$ are the limits in r of $\bar{h}(\mu, r)$ and $\underline{h}(\mu, r)$. Exactly as in Lemma 1.3, as a corollary of Lemma 4.1, we obtain the following claim.

Lemma 4.2. *Let* $\mu \in \mathcal{I}_\pi(M)$. *We have* $\bar{h}(\mu) \geq 0$ *(resp.* $\underline{h}(\mu) \geq 0$*) if and only if* μ *is weakly (resp. strongly) typical.*

4.2. **Annealed entropy.** In this broader setting, the annealed entropy is defined as follows. Let (T, o) be a randomly labeled rooted unimodular tree and X an invariant coloring of T with (T, X) having law μ. The degree of a vertex v of T is denoted by $\deg(v)$. We assume that $d = \mathbb{E}\deg(o) > 0$. As above, if $r \geq 0$ is an integer and S is a subset of the vertices of T, $B_r(S)$ is the subset of vertices of T at distance at most r from S. For $r \geq 1$, we set $S_r = B_r(o)$ and, if $\deg(o) \geq 1$, we set $E_r = B_{r-1}(\{o, 1\})$ (since T is randomly labeled, the neighbors of the root are indexed by $(1, \ldots, \deg(o))$).

We denote by X_{S_r} the colored tree (T, X) restricted to S_r: by construction, X_{S_r} has law μ_r. We also need to define the law of X restricted to E_r but this requires a biasing of the tree T. This is done as follows. A (directed) edge-rooted graph is defined as a pair (G, ρ) formed by a connected graph G and a distinguished directed edge $\rho = (u, v)$ (that is, $\{u, v\}$ is an edge of the graph). Now, we denote by $\vec{\mu}$ the law on colored edge-rooted trees defined by:

$$\vec{\mu}(\cdot) = \frac{1}{d}\mathbb{E}[\deg(o)\mathbf{1}((T, X, (o, 1)) \in \cdot], \tag{4.3}$$

where (T, X) has law μ and $d = \mathbb{E}\deg(o)$. Note that under the probability measure $\vec{\mu}$, o has at least degree 1 and thus $\{o, 1\}$ is an edge of the tree. We denote by $(\vec{T}, \vec{X}, \rho)$, with $\rho = (o, 1)$ a random variable with law $\vec{\mu}$. It is easy to check that Equation (4.1) implies that $\vec{\mu}$ is invariant by switching the two sides of the oriented edge. Moreover, if T has law $\mathrm{UGW}(\pi)$ then $\vec{T}$ is given by two independent Galton-Watson trees with offspring distribution $\hat{\pi}$ whose roots are connected by the root-edge, see [1, Example 1.1]. In particular, if π is a Dirac mass at d, then $T = \vec{T}$.

We denote by $\vec{X}_{E_r}$ the colored tree $(\vec{T}, \vec{X})$ restricted to E_r. The law of $\vec{X}_{E_r}$ is $\vec{\mu}_r$, the restriction of $\vec{\mu}$ to E_r. We observe that $\vec{\mu}_r$ depends on μ only through its marginal μ_r.

Now, if X is an invariant coloring of T with law $\mu \in \mathcal{I}_\pi(M)$ and $r \geq 1$ is an integer, we set

$$\Sigma_r(X) = \Sigma_r(\mu) = H(X_{S_r}) - \frac{d}{2}H(\vec{X}_{E_r}) - H(\pi). \tag{4.4}$$

See Remark 4.5 for an alternative expression which is arguably more natural. Thanks to assumption (4.2) it is immediate that the above entropies are finite as soon as M is finite. Note also that $\Sigma_r(\mu)$ depends on μ only through μ_r. We will check in Lemma 4.10 below that $\Sigma_r(\mu)$ is non-increasing in r. We may thus define

$$\Sigma(\mu) = \lim_{r \to \infty} \Sigma_r(\mu).$$

The quantities $\Sigma_r(\mu)$ and $\Sigma(\mu)$ are the annealed entropies of μ_r and μ. The following lemma generalizes Lemma 4.3.

Theorem 4.3. *For any* $\mu \in \mathcal{I}_\pi(M)$ *and integer* $r \geq 1$, *we have*

$$\bar{h}(\mu, r) \leq \Sigma_r(\mu) \quad and \quad \bar{h}(\mu) \leq \Sigma(\mu).$$

There is also an analog of Theorem 1.7.

Theorem 4.4. *Let* $\mu \in \mathcal{I}_\pi(M)$. *For any integer* $r \geq 1$, *if all invariant couplings* ν *of* μ *and* μ, *we have* $\Sigma_r(\nu) \leq 2\Sigma_r(\mu)$ *then* $\underline{h}(\mu, r) = \bar{h}(\mu, r) = \Sigma_r(\mu)$. *In particular, if the above condition holds for an increasing sequence of integers* $(r_k)_{k \geq 1}$ *then* $\underline{h}(\mu) = \bar{h}(\mu) = \Sigma(\mu)$.

Remark 4.5. The annealed entropy is also given by the formula:

$$\Sigma_r(X) = H(X_{S_r}|T_{S_r}) - \frac{d}{2}H(\vec{X}_{E_r}|\vec{T}_{E_r}),$$

where $H(X|Y) = H(X, Y) - H(Y)$ is the relative entropy. Indeed $\vec{T}$ is the union of two independent copies of T', a Galton-Watson tree with offspring distribution $\hat{\pi}$, while T is the union of N independent copies of T' and with N independent with distribution π. It follows that, $H(\vec{T}_{E_r}) = 2H(T'_{S_{r-1}})$ and $H(T_{S_r}) = H(N) + dH(T'_{S_{r-1}})$ (from (2.5)). In particular, $H(T_{S_r}) - (d/2)H(\vec{T}_{E_r}) = H(N) = H(\pi)$.

4.3. **Markov processes.** There is an extension of Theorem 1.11 and Theorem 1.15 in our extended setting. The previous definitions of Markov processes carry over when conditioned on the random tree. More precisely, we use the following definitions.

Definition 4.6 (Markov process)**.** *Let (T, X) be a random coloring of a random tree T on a finite set M with law μ. For integer $r \geq 1$, X or μ is r-Markov if conditioned on T on $B_{r-1}(e)$ and on the value at $B_{r-1}(e)$, the $(r-1)$-neighborhood of an edge e, the processes on the left and right subtrees of e are independent. Similarly, X or μ is vertex-Markov if conditioned on T and on the value at a vertex, the processes on the pending subtrees of that vertex are independent.*

For integer $r \geq 1$, let $\mathcal{I}_{\pi,r}(M)$ denote the set of laws μ of coloring (T', X') on M which are randomly labeled, such that T' has law $\text{UGW}(\pi)_r$ (the law of the restriction of T to S_r) and such that $\vec{\mu}$ as defined above is invariant by switching the two sides of the oriented edge. If $\mu \in \mathcal{I}_\pi(M)$ then $\mu_r \in \mathcal{I}_\pi(M)$. Conversely, we have the following (see [12, Proposition 1.1]):

Lemma 4.7. *Let $r \geq 1$ integer and $p \in \mathcal{I}_{\pi,r}(M)$. Then there is a unique r-Markov process $\mu(p) \in \mathcal{I}_\pi(M)$ such that the marginal of $\mu(p)$ on S_r is equal to μ.*

If $p \in \mathcal{I}_{\pi,r}(M)$, we define $\Sigma(p) = \Sigma_r(\mu(p))$ as in Equation (4.4). The following theorem is an extension of Theorem 1.11.

Theorem 4.8. *Let $r \geq 1$ be an integer and $p \in \mathcal{I}_{\pi,r}(M)$. If for all couplings $q \in \mathcal{I}_{\pi,r}(M^2)$ of p and p, we have $\Sigma(q) \leq 2\Sigma(p)$ then $\underline{h}(\mu(p)) = \bar{h}(\mu(p)) = \Sigma(p)$ and $\mu(p)$ is strongly typical.*

There is a version of this theorem for vertex-Markov processes. As above, let $\mathcal{I}_e(M)$ denote the set of probability measures on M^{E_1} that are invariant by switching the two sides of the edge $\{o, 1\}$. If $\mu \in \mathcal{I}_\mu(M)$ then the restriction of $\vec{\mu}$ to E_1 is in $\mathcal{I}_e(M)$. Conversely, if $p \in \mathcal{I}_e(M)$, there exists a unique vertex-Markov process $\mu(p)$ in $\mathcal{I}_\pi(M)$ such that $\vec{\mu}$ restricted to E_1 is in $\mathcal{I}_e(M)$. If $(T, X) \in \mathcal{I}_\pi(M)$, we define

$$\Sigma_e(X) = \frac{d}{2}H(\vec{X}_{E_1}) - dH(\vec{X}_o) + H(X_o) - H(\pi).$$

If $p \in \mathcal{I}_e(M)$, we set $\Sigma(p) = \Sigma_e(\mu(p))$. The following theorem is an extension of Theorem 1.15.

Theorem 4.9. *Let $p \in \mathcal{I}_e(M)$. If for all couplings $q \in \mathcal{I}_e(M^2)$ of p and p, we have $\Sigma(q) \leq 2\Sigma(p)$ then $\underline{h}(\mu(p)) = \bar{h}(\mu(p)) = \Sigma(p)$ and $\mu(p)$ is strongly typical.*

In the remainder of the paper, we explain the proofs of Theorem 4.3, Theorem 4.4, Theorem 4.8 and Theorem 4.9. The proofs are entirely similar to the proof of the corresponding results for invariant processes on T_d. We will only sketch the proof and explain the differences.

4.4. **Maximizers of the annealed entropy.** The following lemma is the exact analog of Lemma 1.12 and Lemma 1.14.

Lemma 4.10. *Let $X \in \mathcal{I}_\pi(M)$ and $r \geq 1$. We have*

$$\Sigma_{r+1}(X) \leq \Sigma_r(X),$$

with equality if and only if $X_{S_{r+1}}$ is a r-Markov process on S_{r+1}. Moreover,

$$\Sigma_1(X) \leq \Sigma_e(X),$$

with equality if and only if X_{S_1} is a vertex-Markov process on S_1.

Proof. We start by the first statement. Arguing as in the proof of Lemma 4.10, it is enough to check the inequality for $r = 1$. The inequality $\Sigma_2(X) \leq \Sigma_1(X)$ is equivalent to

$$H(X_{S_2}) - H(X_{S_1}) - \frac{d}{2}H(\vec{X}_{E_2}) + \frac{d}{2}H(\vec{X}_{E_1}) \leq 0. \tag{4.5}$$

Let $\deg_T(o)$ and $\deg_{\vec{T}}(o)$ be the degrees of the root in T and $\vec{T}$. For integer $k \geq 1$, from (4.3), we have, for any event A,

$$\mathbb{P}((\vec{X}, \vec{T}) \in A, \deg_{\vec{T}}(o) = k) = \frac{k}{d}\mathbb{P}((X, T) \in A, \deg_T(o) = k).$$

It follows that $\mathbb{P}(\deg_{\vec{T}}(o) = k) = k\pi(k)/d$ and if $k \geq 1$ is in the support of π,

$$\mathbb{P}((\vec{X}, \vec{T}) \in A | \deg_{\vec{T}}(o) = k) = \mathbb{P}((X, T) \in A | \deg_T(o) = k). \tag{4.6}$$

In other words, (X, T) and $(\vec{X}, \vec{T})$ have the same law when conditioned on the root degree is $k \geq 1$. For a set S of vertices, let us denote by by $H_k(S)$ the entropy of the variable (T, X) restricted to S and conditioned on the event $\deg_T(o) = k$. Similarly, we set $H_k(S|S') = H_k(S, S') - H_k(S')$ is the associated relative entropy.

From (2.5), we may write, for $t = 1, 2$,

$$H(X_{S_t}) = H(\deg_T(o)) + \pi(0)H_0(X_o) + \sum_{k=1}^{\infty} \pi(k)H_k(S_t).$$

and

$$H(\vec{X}_{E_t}) = H(\deg_{\vec{T}}(o)) + \sum_{k=1}^{\infty} \frac{k\pi(k)}{d} H_k(E_t).$$

Hence, (4.5) is equivalent to the claim:

$$\sum_{k=1}^{\infty} \pi(k)\left(H_k(S_2) - H_k(S_1) - \frac{k}{2}H_k(E_2) + \frac{k}{2}H_k(E_1)\right) \leq 0. \tag{4.7}$$

On the event $\deg_T(o) = k$, for $1 \leq i \leq k$, let $L_i = \{(i, 1), \ldots, (i, n_i)\}$ be the offspring of vertex i. Note that the random variables X_{L_i} conditioned on the event $\deg_T(o) = k$ are exchangeable. Then, the computation from (2.6) to (2.8) gives

$$H_k(S_2) - H_k(S_1) - \frac{k}{2}H_k(E_2) + \frac{k}{2}H_k(E_1) \leq \frac{k}{2}H_k(L_1|S_1) - \frac{k}{2}H_k(L|E_1)$$

The left-hand side of (4.7) is thus upper bounded by

$$\sum_{k=1}^{\infty} \frac{k}{2}\pi(k)(H_k(L_1|S_1) - H_k(L|e)) = \frac{d}{2}\left(H(\vec{X}_{L_1}|\vec{X}_{S_1}) - H(\vec{X}_L|\vec{X}_{E_1})\right).$$

We now use the invariance of $\vec{\mu}$ by switching the two sides of the edge $E_1 = \{o, 1\}$. We get $H(\vec{X}_L|\vec{X}_{E_1}) = H(\vec{X}_{L_1}|\vec{X}_{E_1})$. Finally, since $E_1 \subset S_1$, we deduce that the inequalities (4.5)-(4.7) hold. As in the proof of Lemma 1.12, the case of equality is a directed consequence of the case of equality in (2.5).

We now prove the second statement of Lemma 4.10, $\Sigma_1(X) \leq \Sigma_e(X)$. It is equivalent to prove that

$$H(X_{S_1}) - \frac{d}{2}H(\vec{X}_{E_1}) \leq \frac{d}{2}H(\vec{X}_{E_1}) - dH(\vec{X}_o) + H(X_o).$$

Arguing as above, we find that this is equivalent to

$$\sum_{k=1}^{\infty} \pi(k)(H_k(S_1) - kH_k(E_1) - (k-1)H_k(o)) \leq 0.$$

As in the proof of Lemma 1.14, it remains to use that $H_k(S_1) \leq H(o) + kH_k(E_1|o)$ and $H_k(E_1|o) = H_k(E_1) - H_k(o)$. In the case of equality, this implies the conditional independence of $(X_1, \ldots, X_k)$ given X_o and root degree equal to k. $\qquad\square$

4.5. **Combinatorial characterization of the annealed entropy.** In our extended setting, the combinatorial interpretation of the annealed entropy $\Sigma_r(\mu)$ explained in Subsection 2.3 continues to hold. The definition of $\Sigma_n(\mu, r, \epsilon)$ in (2.9) remains unchanged. The following theorem is an extension of Theorem 2.3.

Theorem 4.11. *Let $\mu \in \mathcal{I}_\pi(M)$ and $r \geq 1$ integer. We have*

$$\lim_{\epsilon \to 0} \liminf_{n \to \infty} \Sigma_n(\mu, r, \epsilon) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \Sigma_n(\mu, r, \epsilon) = \Sigma_r(\mu).$$

Proof. Let $s(d) = d/2 - (d/2)\log d$. From $[10$, Theorem 2.16$]$, we have

$$\lim_{n \to \infty} \left(\frac{1}{n}\log|\mathcal{G}_n(d_n)| - \frac{d}{2}\log n\right) = -s(d) + H(\deg(o)) - \mathbb{E}[\log(\deg(o)!)],$$

where $\deg(o)$ has law π. Also, from [12, 16], we have

$$\lim_{\epsilon \to 0} \liminf_{n \to \infty} \left(\frac{1}{n} \log |\mathcal{G}_n(\mu, r, \epsilon)| - \frac{d}{2} \log n \right) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \left(\frac{1}{n} \log |\mathcal{G}_n(\mu, r, \epsilon)| - \frac{d}{2} \log n \right) = J_r(\mu),$$

where $J_r(\mu) = -s(d) + H(\widetilde{X}_{S_r}) - \frac{d}{2} H(\pi_\mu) - \sum_{g \in \widetilde{\mathcal{E}}_r} \mathbb{E}[\log(N_X(g)!)]$ is defined exactly as in Theorem 2.3, the only difference being that (2.11) is replaced by

$$\sum_{g \in \widetilde{\mathcal{E}}_r} N_X(g) = \deg(o).$$

Arguing as in the proof of Theorem 2.3, we have $H(\pi_\mu) = H(\widetilde{X}_{E_r})$ where $\widetilde{X}_{E_r}$ is the unlabeled coloring associated to $\vec{X}_{E_r}$. Hence, the theorem follows by checking that

$$H(\widetilde{X}_{S_r}) - \frac{d}{2} H(\widetilde{X}_{E_r}) = H(X_{S_r}) - \frac{d}{2} H(\vec{X}_{E_r}) + \sum_{g \in \widetilde{\mathcal{E}}_r} \mathbb{E}[\log(N_X(g)!)] - \mathbb{E}[\log(\deg(o)!)]. \tag{4.8}$$

In order to prove that (4.8) holds, we decompose the left-hand side of the possible values of the root-degree. We write

$$H(\widetilde{X}_{S_r}) = H(\pi) + H_0(X_o) + \sum_{k=1}^{\infty} \pi(k) H_k(\widetilde{X}_{S_r}),$$

where H_k is the entropy conditioned on $\deg_T(o) = k$. Similarly,

$$\frac{d}{2} H(\widetilde{X}_{E_r}) = \frac{d}{2} H(\hat{\pi}) + \sum_{k=1}^{\infty} \frac{k}{2} \pi(k) H_k(\widetilde{X}_{E_r}).$$

Let $\mathbb{E}_k[\cdot]$ is the expectation conditioned on $\deg_T(o) = k$. From (4.6), it follows that the identity (4.8) is equivalent to

$$\sum_{k=1}^{\infty} \pi(k) \left(H_k(\widetilde{X}_{S_r}) - \frac{k}{2} H_k(\widetilde{X}_{E_r}) \right) = \sum_{k=1}^{\infty} \pi(k) \left(H_k(X_{S_r}) - \frac{k}{2} H_k(X_{E_r}) + \sum_{g \in \widetilde{\mathcal{E}}_r} \mathbb{E}_k[\log(N_X(g)!)] - \log(k!) \right). \tag{4.9}$$

We denote by E_r^o, the tree rooted at o in $E_r \backslash \{o, 1\}$ and by E_r^1 the tree rooted at 1. Arguing as in the proof of Theorem 2.3, we find

$$H_k(X_{E_r}) = H_k(\widetilde{X}_{E_r}) + K_k + K_k',$$

where K_k is the relative entropy of a random labeling σ of $\widetilde{X}_{E_r^o}$ conditioned on $\deg(o) = k$ and K_k' is the relative entropy of a random labeling σ' of $\widetilde{X}_{E_r^1}$ conditioned on $\deg(o) = k$. Similarly, arguing as in the proof of Theorem 2.3, we get

$$H_k(X_{S_r}) = H_k(\widetilde{X}_{S_r}) + k K_k' - \sum_{g \in \widetilde{\mathcal{E}}_r} \mathbb{E}_k[\log(N_X(g)!)] + \log(k!).$$

We deduce that the right-hand side of (4.9) is equal to

$$\sum_{k=1}^{\infty} \pi(k) \left(H_k(\widetilde{X}_{S_r}) - \frac{k}{2} H_k(\widetilde{X}_{E_r}) \right) + \sum_{k=1}^{\infty} \pi(k) \left(\frac{k}{2} K_k' - \frac{k}{2} K_k \right).$$

Finally, we observe that

$$\sum_{k=1}^{\infty} \pi(k) \left(\frac{k}{2} K_k' - \frac{k}{2} K_k \right) = \frac{d}{2} (H(\sigma) - H(\sigma')).$$

The above expression is equal to 0 because the law $\vec{\mu}$ is invariant by switching the two sides of the edge $\{o, 1\}$. This concludes the proof of Equation (4.8). $\qquad \square$

Remark 4.12. The proof of Theorem 4.11 gives the simplified expression $J_r(\mu) = -s(d) + \Sigma_r(\mu) - \mathbb{E}[\log(\deg(o)!)]$. In particular, the proof of Lemma 4.10 has an interesting corollary for the maximizers of these functions $J_r(\mu)$. Notably, in [12, Theorem 1.3] below (9) we may remove the assumption that ρ_1 has finite support. Consequently, from [12, Corollary 1.4], the annealed entropy $\Sigma(\rho)$ defined in [12] for unimodular random trees with average root degree d is *uniquely* maximized by $\mathrm{UGW}(\mathrm{Poi}(d))$.

4.6. **Proofs of Theorem 4.3, Theorem 4.4, Theorem 4.8 and Theorem 4.9.** As already pointed, the conclusion of Theorem 2.1 holds in our extended setting. We may thus repeat verbatim the proofs in Section 3 and invoke Theorem 4.11 in place of Theorem 2.3 and Lemma 4.10 in place of Lemma 1.12. □

References

[1] D. Aldous and R. Lyons. Processes on unimodular random networks. *Electronic Journal of Probability*, 12:1454–1508, 2007. 16

[2] A. Backhausz and B. Szegedy. On large-girth regular graphs and random processes on trees. *Random Structures Algorithms*, 53(3):389–416, 2018. 1, 3, 4, 11

[3] A. Backhausz and B. Szegedy. On the almost eigenvectors of random regular graphs. *Ann. Probab.*, 47(3):1677–1725, 2019. 1, 3, 4, 5, 11

[4] A. Backhausz, B. Szegedy, and B. Virág. Ramanujan graphings and correlation decay in local algorithms. *Random Structures Algorithms*, 47(3):424–435, 2015. 1

[5] A. Backhausz and B. Virág. Spectral measures of factor of i.i.d. processes on vertex-transitive graphs. *Ann. Inst. Henri Poincaré Probab. Stat.*, 53(4):2260–2278, 2017. 1

[6] M. Bayati, D. Gamarnik, and P. Tetali. Combinatorial approach to the interpolation method and scaling limits in sparse random graphs. *Ann. Probab.*, 41(6):4080–4115, 2013. 2, 4

[7] E. A. Bender and E. Canfield. The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory, Series A*, 24(3):296 – 307, 1978. 11

[8] I. Benjamini and O. Schramm. Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.*, 6:no. 23, 13, 2001. 1

[9] B. Bollobás. The independence ratio of regular graphs. *Proc. Amer. Math. Soc.*, 83(2):433–436, 1981. 1

[10] B. Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001. 1, 3, 8, 18

[11] C. Bordenave. Normalité asymptotique des vecteurs propres d'un graphe régulier aléatoire (d'après Backhausz et Szegedy). Séminaire Bourbaki, 71e année, 2018–2019, No. 1151., 2018. 1

[12] C. Bordenave and P. Caputo. Large deviations of empirical neighborhood distribution in sparse random graphs. *Probab. Theory Related Fields*, 163(1-2):149–222, 2015. 4, 5, 11, 17, 19, 20

[13] L. P. Bowen. A brief introduction of sofic entropy theory. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. III. Invited lectures*, pages 1847–1866. World Sci. Publ., Hackensack, NJ, 2018. 2, 3

[14] A. Coja-Oghlan and W. Perkins. Spin systems on Bethe lattices. *Comm. Math. Phys.*, 372(2):441–523, 2019. 2, 4

[15] E. Csóka, V. Harangi, and B. Virág. Entropy and expansion. *Ann. Inst. Henri Poincaré Probab. Stat.*, 56(4):2428–2444, 2020. 1

[16] P. Delgosha and V. Anantharam. A notion of entropy for stochastic processes on marked rooted graphs. arXiv:1908.00964, 2019. 11, 19

[17] J. Ding, A. Sly, and N. Sun. Maximum independent sets on random regular graphs. *Acta Math.*, 217(2):263–340, 2016. 2

[18] W. Duckworth and N. C. Wormald. On the independent domination number of random regular graphs. *Combin. Probab. Comput.*, 15(4):513–522, 2006. 1

[19] P. Erdős and T. Gallai. Graphs with prescribed degrees of vertices (hungarian). *Mat. Lapok*, 11:264–274, 1960. 15

[20] S. Franz and M. Leone. Replica bounds for optimization problems and diluted spin systems. *J. Statist. Phys.*, 111(3-4):535–564, 2003. 4

[21] J. Friedman. A proof of Alon's second eigenvalue conjecture and related problems. *Mem. Amer. Math. Soc.*, 195(910):viii+100, 2008. 1

[22] D. Gamarnik. Right-convergence of sparse random graphs. *Probab. Theory Related Fields*, 160(1-2):253–278, 2014. 1, 2, 4

[23] F. Guerra. Broken replica symmetry bounds in the mean field spin glass model. *Comm. Math. Phys.*, 233(1):1–12, 2003. 4

[24] H. Hatami, L. Lovász, and B. Szegedy. Limits of locally-globally convergent graph sequences. *Geom. Funct. Anal.*, 24(1):269–296, 2014. 1, 2

[25] M. Lelarge and M. Oulamara. Replica bounds by combinatorial interpolation for diluted spin systems. *J. Stat. Phys.*, 173(3-4):917–940, 2018. 4

[26] L. Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012. 1

[27] M. Mezard and A. Montanari. *Information, Physics, and Computation.* Oxford University Press, Inc., USA, 2009. 2

[28] D. Puder. Expansion of random graphs: new proofs, new results. *Invent. Math.*, 201(3):845–908, 2015. 1

[29] J. Salez. The interpolation method for random graphs with prescribed degrees. *Combin. Probab. Comput.*, 25(3):436–447, 2016. 2, 4

[30] R. van der Hofstad. *Random Graphs and Complex Networks, Volume 2*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, to appear. 15

[31] N. C. Wormald. Differential equations for random processes and random graphs. *Ann. Appl. Probab.*, 5(4):1217–1235, 1995. 1

[32] N. C. Wormald. Models of random regular graphs. In *Surveys in combinatorics, 1999 (Canterbury)*, volume 267 of *London Math. Soc. Lecture Note Ser.*, pages 239–298. Cambridge Univ. Press, Cambridge, 1999. 1, 8