



**HAL**  
open science

## QF-Limits of Structures and the Example of Tree-Semilattices

Pierre Charbit, Lucas Hosseini, Patrice Ossona de Mendez

► **To cite this version:**

Pierre Charbit, Lucas Hosseini, Patrice Ossona de Mendez. QF-Limits of Structures and the Example of Tree-Semilattices. 2015. hal-01150659v1

**HAL Id: hal-01150659**

**<https://hal.science/hal-01150659v1>**

Preprint submitted on 11 May 2015 (v1), last revised 17 Sep 2015 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# QF-LIMITS OF STRUCTURES AND THE EXAMPLE OF TREE-SEMILATTICES

PIERRE CHARBIT, LUCAS HOSSEINI, AND PATRICE OSSONA DE MENDEZ

ABSTRACT. The notion of left convergence for graphs (in relation with homomorphism densities for fixed patterns and Szemerédi's regularity lemma) introduced by Lovász et al. got increasingly studied over the past 10 years. Recently, Nešetřil and Ossona de Mendez introduced a general framework to study the limits of structures based on the converging probability for the structures of the sequence to verify any formula of a given fragment of first order logic for a random assignment of free variables. In this context, a sequence is *quantifier-free convergent* (or QF-convergent) if the probability of any equation being satisfied converges. We will give examples, and show how the QF-convergence of colored tree-semilattices can be related to the left convergence of  $m$ -partite cographs, a generalization of cographs.

## 1. INTRODUCTION

The study of limits of graphs gained recently a major interest [5, 6, 8, 17, 18, 19]. In the framework studied in the aforementioned papers a sequence  $(G_n)_{n \in \mathbb{N}}$  of graphs is said *left-convergent* if, for every (finite) graph  $F$ , the probability

$$t(F, G_n) = \frac{\text{hom}(F, G_n)}{|G_n|^{|F|}}$$

that a random map  $f : V(F) \rightarrow V(G_n)$  is a *homomorphism* (that is a mapping preserving adjacency) converges as  $n$  goes to infinity. (For a graph  $G$ , we denote by  $|G|$  the *order* of  $G$ , that is the number of vertices of  $G$ .)

In this case, the limit object can be represented by means of a *graphon*, that is a measurable function  $W : [0, 1]^2 \rightarrow [0, 1]$ . The definition of the function  $t$  above is extended to graphons by

$$t(F, W) = \int \cdots \int \prod_{ij \in E(F)} W(x_i, x_j) dx_1 \dots dx_p$$

---

*Date:* May 11, 2015.

*2010 Mathematics Subject Classification.* 05C99 (Graph theory).

*Key words and phrases.* Graph limits and Structural limit and Random-free graphon and Tree-order and  $m$ -partite Cograph.

Supported by grant ERCCZ LL-1201 and CE-ITI P202/12/G061, and by the European Associated Laboratory "Structures in Combinatorics" (LEA STRUCO).

Supported by grant ERCCZ LL-1201 and by the European Associated Laboratory "Structures in Combinatorics" (LEA STRUCO).

(where we assume that  $F$  is a graph with vertex set  $\{1, \dots, p\}$ ) and then the graphon  $W$  is the left-limit of a left-convergent sequence of graphs  $(G_n)_{n \in \mathbb{N}}$  if for every graph  $F$  it holds

$$t(F, W) = \lim_{n \rightarrow \infty} t(F, G_n).$$

For  $k$ -regular hypergraphs, the notion of left-convergence extends in the natural way, left-limits — called *hypergraphons* — are measurable functions  $W : [0, 1]^{2^k - 2} \rightarrow [0, 1]$  and have been constructed by Elek and Szegedy using ultraproducts [9] (see also [25]). These limits were also studied by Hoover [13], Aldous [1], and Kallenberg [16] in the setting of exchangeable random arrays (see also [3]).

Recall that for a *signature*  $\sigma$  (that is a set of symbols of relations and functions with their arities), a  $\sigma$ -structure  $\mathbf{A}$  is defined by its *domain* (or *universe*)  $A$  (which is the set of the *elements* of  $\mathbf{A}$ ) and an interpretation in  $A$  of all the relations and functions declared in  $\sigma$ . A  $\sigma$ -structure is *relational* if the signature  $\sigma$  only contains symbols of relations. In some sense, relational structures are natural generalization of  $k$ -uniform hypergraphs. To the opposite, a  $\sigma$ -structure is *functional* (or called an *algebra*) if the signature  $\sigma$  only contains symbols of functions.

In the general framework of  $\sigma$ -structures with finite domain, the notion of *structural convergence*, which is a notion of convergence driven by a fragment  $X$  of first-order logic (in the language of  $\sigma$ -structures), has been introduced by Nešetřil and the third author [21]. In this setting, a sequence  $(\mathbf{A}_n)_{n \in \mathbb{N}}$  of  $\sigma$ -structure is  *$X$ -convergent* if, for every formula  $\phi \in X$  with free variables  $x_1, \dots, x_p$ , the probability

$$(1) \quad \langle \phi, \mathbf{A}_n \rangle = \frac{|\{(v_1, \dots, v_p) \in A_n^p : \mathbf{A}_n \models \phi(v_1, \dots, v_p)\}|}{|A_n|^p}$$

that a random (uniform independent) assignment to the free variables of  $\phi$  of elements of  $A_n$  satisfies  $\phi$  converges as  $n$  goes to infinity. Denoting

$$\phi(\mathbf{A}) = \{(v_1, \dots, v_p) \in A^p : \mathbf{A} \models \phi(v_1, \dots, v_p)\}$$

the set of satisfying assignments of  $\phi$ , Equation (1) rewrites as

$$\langle \phi, \mathbf{A}_n \rangle = \frac{|\phi(\mathbf{A}_n)|}{|A_n|^p}.$$

The value  $\langle \phi, \mathbf{A} \rangle$  is called the *Stone pairing* of the formula  $\phi$  and the  $\sigma$ -structure  $\mathbf{A}$ .

Note that (1) naturally extends to any  $\sigma$ -structure  $\mathbf{A}$ , whose domain is equipped with a probability measure  $\nu_{\mathbf{A}}$  with the property that every  $X$ -definable set is measurable. Precisely, if  $\phi \in X$  has  $p$  free variables then:

$$(2) \quad \langle \phi, \mathbf{A} \rangle = \nu_{\mathbf{A}}^{\otimes p}(\phi(\mathbf{A})) = \Pr\left(\mathbf{A} \models \phi(X_1, \dots, X_p)\right),$$

where  $\nu_{\mathbf{A}}^{\otimes p}$  is the product measure, and where  $X_1, \dots, X_p$  are random elements of  $A_n$  chosen independently with probability  $\nu_A$ . When the probability measure  $\nu_{\mathbf{A}}$  needs to be made explicit we use the notation  $\langle \phi, (\mathbf{A}, \nu_{\mathbf{A}}) \rangle$  instead of  $\langle \phi, \mathbf{A} \rangle$ .

It has been proved in [21] that the limit of an  $X$ -convergent sequence of structures can be represented by means of a probability measure on the Stone space  $S(X)$  dual to the Lindenbaum-Tarski algebra induced by  $X$ , which is invariant under some group action. (This generalizes the point of view of Hoover, Aldous and Kallenberg mentioned above). For more details on this approach, we refer the reader to [21, 23, 22].

Denote by  $\text{QF}_p(\sigma)$  the fragment of all quantifier free formulas with  $p$  free variables (in the language of  $\sigma$ ) and by  $\text{QF}(\sigma) = \bigcup_p \text{QF}_p(\sigma)$  the fragment of all quantifier free formulas. In the following, we shall use  $\text{QF}_p$  and  $\text{QF}$  when the signature  $\sigma$  is clear from context.

In this paper, we shall be specifically interested by *QF-convergence*, which is the convergence driven by the fragment  $\text{QF}$  of quantifier-free formulas.

Because we consider structures with infimum operator  $\wedge$ , we shall use the symbol  $\&$  for the logical conjunction.

## 2. STATEMENT OF THE RESULTS

A *tree-semilattice* is an algebraic structure  $\mathbf{T} = (T, \wedge)$  such that:

- (1)  $(T, \wedge)$  is a meet semi-lattice (i.e. an idempotent commutative semi-group);
- (2)  $\forall x, y, z$  s.t.  $x \wedge z = x$  and  $y \wedge z = y$  it holds  $x \wedge y \in \{x, y\}$ .

Each tree-semilattice  $\mathbf{T}$  canonically defines a partial order on its domain by  $x \leq y$  if  $x \wedge y = x$ . In the case where  $T$  is finite, it is a partial order induced by the ancestor relation of a rooted tree.

It is possible to add finitely many unary relations  $M_1, \dots, M_k$  to the signature of tree-semilattices. In this case, we speak of *colored* tree-semilattices, and we define the color of a vertex  $v$  as the set of the indices of those unary relations it belongs to:  $c(v) = \{i : M_i(v)\}$ .

A *Borel tree-semilattice* is a tree-semilattice  $\mathbf{T}$  on a standard Borel space  $T$ , such that  $\wedge : T \times T \rightarrow T$  is Borel. Note that every finite tree-semilattice is indeed a Borel tree-semilattice.

Our main results concerning QF-convergence of tree-semilattices are as follows:

- every QF-convergent sequence of colored weighted tree-semilattices admits a limit, which is a Borel colored tree-semilattice (Theorem 2), and conversely: every Borel colored tree-semilattice is the limit of some QF-convergent sequence of colored weighted tree-semilattices (Corollary 2);
- every QF-convergent sequence of colored uniform tree-semilattices admits a limit, which is an atomless Borel colored tree-semilattice

(Corollary 3), and conversely: every atomless Borel colored tree-semilattice is the limit of some QF-convergent sequence of colored uniform tree-semilattices (Theorem 1).

The notion of  $m$ -partite cographs has been introduced in [10], based on the generalization of the characterization of cographs by means of cotrees [7]:

A graph  $G$  is an  $m$ -partite cograph if there exists a colored tree-semilattice  $\mathbf{T}$ , such that the vertices of  $G$  are the leaves of  $\mathbf{T}$ , the leaves of  $\mathbf{T}$  are colored with a set of at most  $m$  colors, and the adjacency of any two vertices  $x$  and  $y$  is fully determined by the colors of  $x, y$  and  $x \wedge y$ . (Notice that there is no restriction on the colors used for internal elements of  $\mathbf{T}$ .) In this setting we prove (Theorem 5):

- every left-convergent sequence of  $m$ -partite cographs has a Borel limit, which is the interpretation of an atomless Borel colored tree-semilattice;
- conversely, every interpretation of an atomless Borel colored tree-semilattice is the left-limit of a sequence of  $m$ -partite cographs.

The class of all finite  $m$ -partite cographs can be characterized by means of a finite family  $\mathcal{F}_m$  of excluded induced subgraphs [11, 10]. We prove that this characterization extends to Borel graphs (Theorem 6) in the sense that an atomless Borel graph excludes all graphs in  $\mathcal{F}_m$  as induced subgraphs if and only if it is the interpretation of an atomless colored Borel tree-semilattice.

### 3. QF-CONVERGENCE OF STRUCTURES

The notion of QF-convergence extends several notions of convergence.

**3.1. Left-convergence.** A sequence  $(G_n)_{n \in \mathbb{N}}$  of graphs (or of  $k$ -uniform hypergraphs) with order going to infinity is QF-convergent if and only if it is left-convergent [21]. This is intuitive, as for every finite graph  $F$  with vertex set  $\{1, \dots, p\}$  there is a quantifier-free formula  $\phi_F$  with free variable  $x_1, \dots, x_p$  such that for every graph  $G$  and every  $p$ -tuple  $(v_1, \dots, v_p)$  of vertices of  $G$  it holds  $G \models \phi_F(v_1, \dots, v_p)$  if and only if the map  $i \mapsto v_i$  is a homomorphism from  $F$  to  $G$ .

As mentioned in the introduction the left-limit of a left-convergent sequence of graphs can be represented by a graphon. However it cannot, in general, be represented by a Borel graph — that is a graph having a standard Borel space  $V$  as its vertex set and a Borel subset of  $V \times V$  as its edge set. A graphon  $W$  is *random-free* if it is almost everywhere  $\{0, 1\}$ -valued. Notice that a random-free graphon is essentially the same (up to isomorphism mod 0) as a Borel graph equipped with a non-atomic probability measure on  $V$ . A class of graph  $\mathcal{C}$  is said to be *random-free* if, for every left-convergent sequence of graphs  $(G_n)_{n \in \mathbb{N}}$  with  $G_n \in \mathcal{C}$  (for all  $n$ ) the sequence  $(G_n)_{n \in \mathbb{N}}$  has a random-free limit.

**3.2. Local-convergence.** Local convergence of graphs with bounded degree has been defined by Benjamini and Schramm [4]. In our setting, local convergence can be defined as follows: A sequence  $(G_n)_{n \in \mathbb{N}}$  of graphs with maximum degree  $D$  is *local-convergent* if, for every  $r \in \mathbb{N}$ , the distribution of the isomorphism types of the distance  $r$ -neighborhood of a random vertex of  $G_n$  converges as  $n$  goes to infinity. Local-convergence corresponds immediately to the notion of  $\text{FO}_1^{\text{local}}$ -convergence, where  $\text{FO}_1^{\text{local}}$  denotes the fragment of the local formulas with a single free variable, that is formulas  $\phi$  such that there exists a constant  $r$  with the property that the satisfaction of  $\phi$  only depends on the  $r$ -neighborhood of the free variable.

However, local convergence of bounded degree graphs can also be expressed by means of QF-convergence (in a slightly stronger form): Let  $G_1, \dots, G_n, \dots$  be graphs with maximum degree strictly smaller than  $D$ . By considering a proper edge coloring of  $G_n$  by  $D$  colors, we can represent  $G_n$  as a functional structure  $\mathbf{V}_n$  with signature containing  $D$  unary functions  $f_1, \dots, f_D$ , where  $V_n$  is the vertex set of  $G_n$  and  $f_1, \dots, f_D$  are defined as follows: for every vertex  $v \in V_n$ ,  $f_i(v)$  is either the unique vertex adjacent to  $v$  by an edge of color  $i$ , or  $v$  if no edge of color  $i$  is incident to  $v$ . It is easily checked that if the sequence  $(\mathbf{V}_n)_{n \in \mathbb{N}}$  is QF-convergent if and only if the sequence  $(G_n)_{n \in \mathbb{N}}$  of edge-colored graphs is local-convergent.

If  $(\mathbf{V}_n)_{n \in \mathbb{N}}$  is QF-convergent, then the limit is a *graphing*, that is a functional structure  $\mathbf{V}$  (with same signature as  $\mathbf{V}_n$ ) such that  $V$  is a standard Borel space, and  $f_1, \dots, f_D$  are measure-preserving involutions.

**3.3. QF-convergence of Functional Structures.** In the case above, the property of the functions to be involutions is essential. The case of quantifier free limits of general functional structures is open, even in the case of unary functions. Only the simplest case of a single unary function has been recently settled [24].

The case of QF-limits of functional structures with a single binary function is obviously at least as complicated as the case of graphs, as a graph  $G$  can be encoded by means of a (non-symmetric) function  $f$  defined by  $f(u, v) = u$  if  $u$  and  $v$  are adjacent, and  $f(u, v) = v$  otherwise, with the property that QF-convergence of the encoding is equivalent to left-convergence of the graphs. The natural guess here for a limit object is the following:

**Conjecture 1.** *Let  $\sigma$  be the signature formed by a single binary functional symbol  $f$ .*

*Then the limit of a QF-convergent sequence of finite  $\sigma$ -structures can be represented by means of a measurable function  $w : [0, 1] \times [0, 1] \rightarrow \mathfrak{P}([0, 1])$ , where  $\mathfrak{P}([0, 1])$  stands for the space of probability measures on  $[0, 1]$ .*

As witnessed by the case of local-convergence of graphs with bounded degrees, the “random-free” case, that is the case where the limit object can be represented by a Borel structure with same signature, is of particular interest. In this paper, we will focus on the case of simple structures defined

by a single binary function — the tree semi-lattices — and we will prove that they admit Borel tree semi-lattices for QF-limits. Conversely, we will prove that every Borel tree semi-lattices (with domain equipped with an atomless probability measure) can be arbitrarily well approximated by a finite tree semi-lattices, hence leading to a full characterization of QF-limits of finite tree semi-lattices.

#### 4. PRELIMINARIES

In a finite tree-semilattice, each element  $x$  except the minimum has a unique predecessor, that we call the *father* of  $x$  (as it is the father of  $x$  in the associated tree).

For a tree-semilattice  $\mathbf{T}$  and an element  $v \in T$  we further define

$$\begin{aligned} T_v &= \{u \in T; u \geq v\} \\ F_v &= \{u \text{ minimal in } T_v \setminus \{v\}\}. \end{aligned}$$

Let  $\mathbf{T}$  be a tree-semilattice, and let  $u_1, \dots, u_p \in T$ . The *sub-tree-semilattice*  $\mathbf{T}\langle u_1, \dots, u_p \rangle$  of  $\mathbf{T}$  generated by  $u_1, \dots, u_p$  is the tree-semilattice with elements

$$\left\{ \bigwedge_{i \in I} u_i; \quad \emptyset \neq I \subseteq \{1, \dots, p\} \right\},$$

where  $\wedge$  is defined as the restriction of the  $\wedge$  function of  $\mathbf{T}$  to the domain of  $\mathbf{T}\langle u_1, \dots, u_p \rangle$ .

**Remark 1.** *Condition (2) of the definition of a tree-semilattice can be replaced by condition:*

$$\forall x, y, z \in T \quad |\{x \wedge y, x \wedge z, y \wedge z\}| \leq 2.$$

It follows that for  $u_1, \dots, u_p \in T$ , the sub-tree-semilattice  $\mathbf{T}\langle u_1, \dots, u_p \rangle$  of  $\mathbf{T}$  generated by  $u_1, \dots, u_p$  has domain

$$T \cup \{u_i \wedge u_j : 1 \leq i < j \leq p\}.$$

If  $\phi \in \text{QF}_p$  is any quantifier-free formula with  $p$  free variables (in the language of tree-semilattice) and  $\mathbf{T}$  is a Borel tree-semilattice then  $\phi(\mathbf{T})$  is a Borel subset of  $T^p$ , thus any (Borel) probability measure  $\mu$  on  $\mathbf{T}$  allows to define

$$\langle \phi, (\mathbf{T}, \mu) \rangle = \mu^{\otimes p}(\phi(\mathbf{T})).$$

Let  $\mathbf{T}_1, \mathbf{T}_2$  be Borel tree-semilattices, and let  $\mu_1$  and  $\mu_2$  be probability measures on  $T_1$  and  $T_2$ , respectively. We define the pseudometric

$$\text{dist}((\mathbf{T}_1, \mu_1), (\mathbf{T}_2, \mu_2)) = \sum_{p \geq 1} 2^{-p} \sup_{\phi \in \text{QF}_p} |\langle \phi_p, (\mathbf{T}_1, \mu_1) \rangle - \langle \phi_p, (\mathbf{T}_2, \mu_2) \rangle|.$$

Note that a sequence of Borel tree-semilattices is QF-convergent if and only if it is Cauchy for the above distance.

As mentioned in Section 2, the color of an element  $v$  of a colored tree-semilattice is the set  $c(v)$  of the indices of those unary relations  $v$  belongs to. The order of the relations naturally induces a total order on these colors:

for distinct  $c_1, c_2 \subseteq \{1, \dots, k\}$  it holds  $c_1 < c_2$  if  $\min(c_1 \setminus c_2) < \min(c_2 \setminus c_1)$  (with convention  $\min \emptyset = 0$ ).

### 5. SAMPLING AND APPROXIMATING

Two Borel structures  $\mathbf{A}, \mathbf{B}$  are *QF-equivalent* if  $\langle \phi, \mathbf{A} \rangle = \langle \phi, \mathbf{B} \rangle$  holds for every quantifier free formula  $\phi$ . The following lemma, which is trivial for uniform structures, requires some little work for structures with a probability measure. As it this result is not really needed here, we leave the proof to the reader.

**Lemma 1.** *Two finite structures are QF-equivalent if and only if they are isomorphic.*

**Definition 1.** *Let  $\mathbf{A}$  be a Borel structure, and let  $n, L \in \mathbb{N}$ . The  $(n, L)$ -sampling of  $\mathbf{A}$  is the random structure  $\mathbf{B}_{X_1, \dots, X_n}$  defined as follows:*

- *the domain of  $\mathbf{B}_{X_1, \dots, X_n}$  is the union of sets  $\Omega, \Gamma$ , and  $\{\perp\}$ , where  $\Omega = \{X_1, \dots, X_n\}$  is a set of  $n$  random independent elements of  $\mathbf{A}$  sampled with probability  $\mu_{\mathbf{A}}$ ,  $\Gamma$  is the set of all the elements that can be obtained from  $\Omega$  by at most  $L$  applications of a function, and  $\perp$  is an additional element;*
- *the relations are defined on  $\mathbf{B}_{X_1, \dots, X_n}$  as in  $\mathbf{A}$ , as well as functions when the image belongs to  $\Omega \cup \Gamma$ . When undefined, functions have image  $\perp$ ;*
- *the probability measure on  $\mathbf{B}_{X_1, \dots, X_n}$  assigns probability  $1/n$  to  $X_1, \dots, X_n$ , and probability 0 to the other elements.*

**Lemma 2** ([20]). *Let  $X_1, \dots, X_n$  be independent random variables, with  $X_k$  taking values in a set  $A_k$  for each  $k$ . Suppose that a (measurable) function  $f : \prod A_k \rightarrow \mathbb{R}$  satisfies*

$$|f(\mathbf{x}) - f(\mathbf{x}')| \leq c_k$$

*whenever the vectors  $\mathbf{x}$  and  $\mathbf{x}'$  differ only in the  $k$ th coordinate. Let  $Y$  be the random variable  $f[X_1, \dots, X_n]$ . Then for any  $t > 0$ ,*

$$\Pr(|Y - \mathbb{E}Y| \geq t) \leq 2e^{-2t^2 / \sum c_k^2}.$$

**Lemma 3.** *Let  $\phi$  be a quantifier-free formula  $\phi$  with at most  $p$  free variables and at most  $L$  functional symbols.*

*Then, for every Borel structure  $\mathbf{A}$  and every  $\epsilon > 0$  and  $n > 2/\epsilon$ , it holds*

$$\Pr(|\langle \phi, \mathbf{A} \rangle - \langle \phi, \mathbf{B}_{X_1, \dots, X_n} \rangle| \geq \epsilon) \leq 2e^{-\frac{(\epsilon n - 2)^2}{p^2 n}},$$

*where  $\mathbf{B}_{X_1, \dots, X_n}$  is the  $(n, L)$ -sampling of  $\mathbf{A}$ .*



*Proof.* Let  $n \geq p$  and let  $f$  be the indicator function of  $\phi(\mathbf{A})$ . Then

$$\begin{aligned} \langle \phi, \mathbf{A} \rangle &= \int \cdots \int f(x_1, \dots, x_p) dx_1 \dots dx_p \\ &= \frac{1}{(n)_p} \int \cdots \int \sum_{g \in \text{Inj}([p], [n])} f(x_{g(1)}, \dots, x_{g(p)}) dx_1 \dots dx_n, \end{aligned}$$

where  $(n)_p = n(n-1)\dots(n-p+1)$  is the Pochhammer symbol.

Let

$$h(x_1, \dots, x_n) = \frac{1}{n^p} \sum_{g \in [n]^{[p]}} f(x_{g(1)}, \dots, x_{g(p)}).$$

Then it holds

$$\langle \phi, \mathbf{B}_{X_1, \dots, X_n} \rangle = h(X_1, \dots, X_n).$$

Considering the expectation we get

$$\mathbb{E}h(X_1, \dots, X_n) = \frac{1}{n^p} \int \cdots \int \sum_{g \in [n]^{[p]}} f(x_{g(1)}, \dots, x_{g(p)}) dx_1 \dots dx_n.$$

So we have

$$\begin{aligned} |\langle \phi, \mathbf{A} \rangle - \mathbb{E}h(X_1, \dots, X_n)| &\leq \left(1 - \frac{(n)_p}{n^p}\right) \langle \phi, \mathbf{A} \rangle + \left| \frac{(n)_p}{n^p} \langle \phi, \mathbf{A} \rangle - \mathbb{E}h(X_1, \dots, X_n) \right| \\ &\leq \left(1 - \frac{(n)_p}{n^p}\right) \langle \phi, \mathbf{A} \rangle + \frac{|[n]^{[p]} - \text{Inj}([p], [n])|}{n^p} \\ &< \frac{2}{n} \end{aligned}$$

Now remark that for every  $x_1, \dots, x_n, \hat{x}_i$  it holds

$$|h(x_1, \dots, x_i, \dots, x_n) - h(x_1, \dots, \hat{x}_i, \dots, x_n)| \leq \frac{p}{n},$$

as  $p/n$  bounds the probability that an mapping from  $[p]$  to  $[n]$  will map some value to  $i$ . Thus, according to Lemma 2 it holds for any  $t > 0$ :

$$\Pr(|h(X_1, \dots, X_n) - \mathbb{E}h(X_1, \dots, X_n)| \geq t) \leq 2e^{-\frac{2nt^2}{p^2}}.$$

In particular, for  $t = \epsilon - 2/n$  it holds

$$\begin{aligned} \Pr(|\langle \phi, \mathbf{A} \rangle - \langle \phi, \mathbf{B}_{X_1, \dots, X_n} \rangle| \geq \epsilon) &\leq \Pr(|\mathbb{E}h(X_1, \dots, X_n) - h(X_1, \dots, X_n)| \geq \epsilon - 2/n) \\ &\leq 2e^{-\frac{(\epsilon n - 2)^2}{p^2 n}}. \end{aligned}$$

□

By union bound, we deduce that for sufficiently large  $n$  there exists an  $(n, L)$ -sampling which has  $\epsilon$ -close Stone pairing with any formula with at most  $p$  free variables and at most  $L$  functional symbols. Precisely, we have:

**Corollary 1.** *For every signature  $\sigma$  there exists a function  $R : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$  with the following property:*

*For every Borel  $\sigma$ -structure  $\mathbf{A}$ , every  $\epsilon > 0$  and every  $L \in \mathbb{N}$  there exists, for each  $n \geq R(p, L) \epsilon^{-2}$  an  $(n, L)$ -sampling  $\mathbf{B}$  of  $\mathbf{A}$  such that for every formula  $\phi$  with at most  $p$  free variables and  $L$  functional symbols it holds*

$$|\langle \phi, \mathbf{A} \rangle - \langle \phi, \mathbf{B} \rangle| < \epsilon.$$

Hence we have:

**Corollary 2.** *Every Borel  $\sigma$ -structure is the limit of a sequence of weighted finite  $\sigma$ -structures.*

Note that the finite weighted structures obtained as  $(n, L)$ -sampling of a Borel structure  $\mathbf{A}$  usually have many elements with 0 measure. The problem of determining whether an infinite Borel structure  $\mathbf{A}$  is the limit of a sequence of finite unweighted structures is much more difficult. Note that we have some (easy) necessary conditions on  $\mathbf{A}$ :

- the domain  $A$  is uncountable and the measure  $\mu_{\mathbf{A}}$  is atomless;
- for every definable functions  $f, g : A^r \rightarrow A^r$ , and every definable subset  $X \subseteq A^r$  of the set of fixed points of  $f \circ g$ , the sets  $X$  and  $g(X)$  have the same measure.

The second condition can be seen as a simple generalization of the *intrinsic mass transport principle* of Benjamini and Schramm: a graphing indeed defines a purely functional structure, with  $d$  functional symbols, each interpreted as a measure preserving involution. In this case, the existence for each graphing of a sequence of bounded degree graphs having the given graphing as its limits is the difficult and well-known Aldous-Lyons conjecture [2]. It turns out that one of the main difficulties of this problem concerns the expansion properties of the graphing. This leads naturally to first consider a weakened version we present now (for generalized structures):

Let  $\mathbf{A}$  be a  $\sigma$ -structure and let  $v_1, \dots, v_p \in A$ . The *substructure* of  $\mathbf{A}$  generated by  $v_1, \dots, v_p$  is the  $\sigma$ -structure, denoted  $\mathbf{A}[v_1, \dots, v_p]$ , whose domain is the smallest subset of  $A$  including  $\{v_1, \dots, v_p\}$  closed by all functional symbols of  $\sigma$ , with the same interpretation of the relations and functions symbols as  $\mathbf{A}$ .

We shall now prove that in the case of atomless Borel tree semi-lattices, the sampling techniques can be used to build arbitrarily good finite approximations, thus to build a converging sequence of finite tree semi-lattices with the given Borel semi-lattice as a limit.

**Theorem 1.** *Every atomless Borel tree-semilattice  $(\mathbf{T}, \mu)$  is limit of a sequence of uniform finite Borel tree-semilattices.*

*Proof.* Let  $(\mathbf{T}_N, \mu_N)$  be an  $(N, 1)$ -sampling of  $(\mathbf{T}, \mu)$  (note that in the context of tree-semilattices, and according to Remark 1, taking  $L > 1$  yields the same structures as with  $L = 1$ ). Let  $M = |S| - N$ , that is, the number of vertices of  $\mathbf{T}_N$  that were not directly sampled.

Fix  $C \in \mathbb{N}$  and let  $[C] = \{1, \dots, C\}$ . Let  $(\hat{\mathbf{T}}_N, \hat{\mu}_N)$  be the Borel tree-semilattice with elements set  $\hat{T}_N \subseteq T_N \times [C]$  defined by:

$$\hat{T}_N = \{x \in T_N : \mu_N(\{x\}) = 0\} \times \{1\} \cup \{x \in T_N : \mu_N(\{x\}) > 0\} \times [C]$$

with meet operation defined by

$$(x, i) \wedge (y, j) = \begin{cases} (x \wedge y, 1) & \text{if } x \wedge y \notin \{x, y\} \\ (x, i) & \text{if } x \wedge y = x \text{ and } x \neq y \\ (y, j) & \text{if } x \wedge y = y \text{ and } x \neq y \\ (x, \max(i, j)) & \text{if } x = y \end{cases}$$

and uniform measure  $\hat{\mu}_N$ .

Informally,  $(\hat{\mathbf{T}}_N, \hat{\mu}_N)$  is obtained from  $(\mathbf{T}_N, \mu_N)$  by replacing each of the randomly selected elements used to create  $\mathbf{T}_N$  with a chain on  $c$  vertices, and considering a uniform measure.

Define the map

$$\pi : \hat{T}_N \rightarrow T_N, \quad (x, i) \mapsto x.$$

Note that for every quantifier free formula  $\phi$  with  $p$  free variables, it holds, for every distinct  $u_1, \dots, u_p \in T_N$  and every  $v_1 \in \pi^{-1}(u_1), \dots, v_p \in \pi^{-1}(u_p)$  it holds:

$$\mathbf{T}_N \models \phi(u_1, \dots, u_p) \iff \hat{\mathbf{T}}_N \models \phi(v_1, \dots, v_p).$$

Let  $\mathbf{L}$  be a Borel tree-semilattice,  $p$  an integer, and  $X_1, \dots, X_p$  (resp.  $Y_1, \dots, Y_p$ ) independent random variables in  $T_N$  (resp.  $\hat{T}_N$ ) (with  $X_i(\omega) = \mu_N(\omega)$ , resp.  $Y_i(\omega) = \hat{\mu}_N(\omega)$ ). Let  $\phi$  be a quantifier-free formula. As for any formula  $\psi$  and any structure  $\mathbf{A}$  it holds  $\langle \phi, \mathbf{A} \rangle - \langle \phi \& \psi, \mathbf{A} \rangle \leq \langle \neg\psi, \mathbf{A} \rangle$ , we have

$$\begin{aligned} & \left| \Pr\left(\mathbf{T}_N \models \phi(X_1, \dots, X_p)\right) - \Pr\left(\mathbf{T}_N \models \phi(X_1, \dots, X_p) \& \bigwedge_{i \neq j} (X_i = X_j)\right) \right| \\ & \leq \Pr\left(\bigvee_{1 \leq i < j \leq p} (X_i = X_j)\right) \leq \binom{p}{2} \Pr(X_1 = X_2) \leq \frac{p^2}{2N} \end{aligned}$$

Similarly, denoting  $B$  the event

$$\bigvee_{1 \leq i < j < p} (\pi(Y_i) = \pi(Y_j)) \vee \bigvee_{i=1}^p (\mu_N(\pi(Y_i)) = 0)$$

it holds

$$\begin{aligned} & \left| \Pr\left(\hat{\mathbf{T}}_N \models \phi(Y_1, \dots, Y_p)\right) - \Pr\left(\mathbf{T}_N \models \phi(Y_1, \dots, Y_p) \& \neg B\right) \right| \\ & \leq \Pr(B) \leq \frac{pM}{CN + M} + \frac{p^2}{2N} \leq \frac{p}{N} \left( \frac{M}{C} + \frac{p}{2} \right) \end{aligned}$$

But, denoting by  $I_A$  the indicator function of set  $A$ , it holds:

$$\begin{aligned}
 & \Pr\left(\mathbf{T}_N \models \phi(X_1, \dots, X_p) \ \& \ \neg \bigvee_{i \neq j} X_i = X_j\right) \\
 &= \sum_{f \in \text{Inj}([p], T_N)} \left( \prod_{i=1}^p \mu_N(\{f(i)\}) \right) I_{\phi(\mathbf{T}_n)}(f(1), \dots, f(p)) \\
 &= \frac{1}{N^p} \sum_{f \in \text{Inj}([p], T_N)} I_{\phi(\mathbf{T}_n)}(f(1), \dots, f(p)) \\
 &= \frac{1}{(CN)^p} \sum_{f \in \text{Inj}([p], T_N)} \sum_{g \in [C]^{[p]}} I_{\phi(\mathbf{T}_n)}(f(1), \dots, f(p)) \\
 &= \frac{1}{(CN)^p} \sum_{f \in \text{Inj}([p], T_N)} \sum_{g \in [C]^{[p]}} I_{\phi(\hat{\mathbf{T}}_n)}((f(1), g(1)), \dots, (f(p), g(p))) \\
 &= \frac{(CN + M)^p}{(CN)^p} \Pr\left(\hat{\mathbf{T}}_N \models \phi(Y_1, \dots, Y_p) \ \& \ \neg B\right).
 \end{aligned}$$

Thus

$$\begin{aligned}
 & \left| \Pr\left(\mathbf{T}_N \models \phi(X_1, \dots, X_p) \ \& \ \neg \bigvee_{i \neq j} X_i = X_j\right) - \Pr\left(\hat{\mathbf{T}}_N \models \phi(Y_1, \dots, Y_p) \ \& \ \neg B\right) \right| \\
 & \leq \left( \frac{CN + M}{CN} \right)^p - 1 \leq \frac{pM}{CN}
 \end{aligned}$$

Altogether, we have:

$$\left| \Pr\left(\mathbf{T}_N \models \phi(X_1, \dots, X_p)\right) - \Pr\left(\hat{\mathbf{T}}_N \models \phi(Y_1, \dots, Y_p)\right) \right| \leq \frac{p}{N} \left( p + \frac{2M}{C} \right).$$

In other words, it holds

$$\left| \langle \phi, (\mathbf{T}_N, \mu_N) \rangle - \langle \phi, \hat{\mathbf{T}}_N \rangle \right| \leq \frac{p}{N} \left( p + \frac{2M}{C} \right).$$

Together with Corollary 2, we get that for every atomless Borel tree semi-lattice  $(\mathbf{T}, \mu)$ , every  $\epsilon > 0$  and every  $p \in \mathbb{N}$  there exists a finite (unweighted) tree semi-lattice  $\hat{\mathbf{T}}$  such that for every quantifier free formula  $\phi$  with  $p$  free variables it holds

$$\left| \langle \phi, \mathbf{T} \rangle - \langle \phi, \hat{\mathbf{T}} \rangle \right| < \epsilon,$$

hence if we choose  $p > -\log_2 \epsilon$  it holds  $\text{dist}(\mathbf{T}, \hat{\mathbf{T}}) < \epsilon + 2^{-p} < 2\epsilon$ .  $\square$

## 6. LIMITS OF TREE-SEMILATTICES

In this section, we focus on providing an approximation lemma for finite colored tree-semilattices, which can be seen as an analog of the weak version of Szemerdi's regularity lemma. For the sake of simplicity, in this section, by tree-semilattice we always mean a finite weighted colored tree-semilattice.

### 6.1. Partitions of tree-semilattices.

**Definition 2.** Let  $(\mathbf{T}, \mu)$  be a tree-semilattice and let  $\epsilon > 0$ . Then  $v \in T$  is said:

- $\epsilon$ -light if  $\mu(T_v) < \epsilon$ ;
- $\epsilon$ -singular if  $\mu(T'_v) \geq \epsilon$  where  $T'_v$  is the set  $T_v$  minus the sets  $T_u$  for non- $\epsilon$ -light children  $u$  of  $v$ ;
- $\epsilon$ -chaining if  $v$  is not singular and has exactly 1 non-light child;
- $\epsilon$ -branching if  $v$  is not singular and has at least 2 non-light children.

(One can easily convince themselves that every vertex of a tree-semilattice falls in exactly one of those categories.)

**Definition 3.** Let  $(\mathbf{T}, \mu)$  be a tree-semilattice. A partition  $\mathcal{P}$  of  $T$  is an  $\epsilon$ -partition of  $(\mathbf{T}, \mu)$  if

- each part is of one of the following types (for some  $v \in T$ , see Fig. 1):
  - (1)  $P = \{v\}$ ,
  - (2)  $P = \{v\} \cup \bigcup_{x \in F} T_x$  for some non-empty  $F \subseteq F_v$ ,
  - (3)  $P = \bigcup_{x \in F} T_x$  for some  $F \subseteq F_v$  with  $|F| \geq 2$ ,
  - (4)  $P = T_v \setminus T_w$  for some  $w \in T_v$  distinct from  $v$  ( $w$  is called the cut vertex of  $P$ , and the path from  $v$  to the father of  $w$  is called the spine of  $P$ ),
 where  $v$  (which is easily checked to be the infimum of  $P$ ) is called the attachment vertex of  $P$  and is denoted by  $A_{\mathcal{P}}(P)$ ;
- each attachment vertex of a part of type 3 is also the attachment vertex of some part of type 1 or 2;
- every part which is not a singleton has  $\mu$ -measure at most  $\epsilon$ .

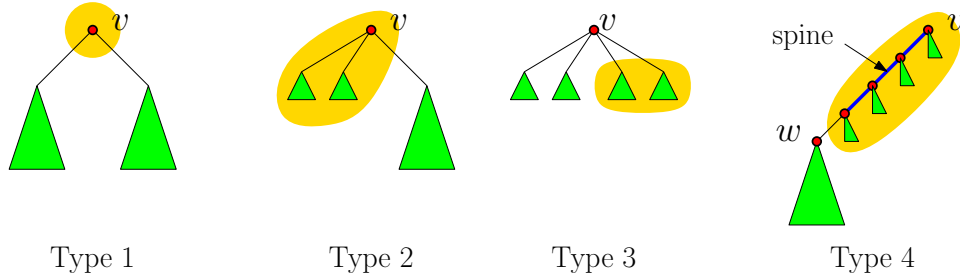


FIGURE 1. Types of part in an  $\epsilon$ -partition

An  $\epsilon$ -partition  $\mathcal{P}$  of a tree-semilattice  $(\mathbf{T}, \mu)$  canonically defines a quotient rooted tree:

**Definition 4.** Let  $\mathcal{P}$  be an  $\epsilon$ -partition of a tree-semilattice  $(\mathbf{T}, \mu)$ . The quotient rooted tree  $\mathbf{T}/\mathcal{P}$  is the rooted tree with node set  $\mathcal{P}$ , the root of which is the unique part of  $\mathcal{P}$  that contains the minimum element of  $\mathbf{T}$ , where the father of any non-root part  $P$  is the part that contain the father

of  $A_{\mathcal{P}}(P)$  if  $A_{\mathcal{P}}(P) \notin P$  (i.e. types 1, 2, or 4) or  $A_{\mathcal{P}}(P)$  itself if  $A_{\mathcal{P}}(P) \in P$  (i.e. type 3). By abuse of notation,  $\mathbf{T}/\mathcal{P}$  will also denote the tree-semilattice defined by the ancestor relation in the rooted tree  $\mathbf{T}/\mathcal{P}$ .

In several circumstances it will be handy to refer to an  $\epsilon$ -partition  $\mathcal{P}$  of a tree-semilattice  $(\mathbf{T}, \mu)$  directly by means of the partition map  $f : T \rightarrow \mathcal{P}$  (where  $\mathcal{P}$  is meant as the vertex set of  $\mathbf{T}/\mathcal{P}$ ). Note that this mapping is a *weak homomorphism* in the sense that for every  $x, y \in T$  it holds  $f(x \wedge y) = f(x) \wedge f(y)$  except (maybe) in the case where  $f(x) = f(y)$ . The definition of  $\epsilon$ -partitions can easily be transposed to provide a characterization of those weak homomorphisms  $f : \mathbf{T} \rightarrow \tilde{\mathbf{T}}$  that define an  $\epsilon$ -partition. Such mappings are called  *$\epsilon$ -partition functions*.

We now prove that every tree-semilattice has a small  $\epsilon$ -partition.

**Lemma 4.** *Let  $(\mathbf{T}, \mu)$  be a tree-semilattice and  $\epsilon > 0$ . Then there exists an  $\epsilon$ -partition  $\mathcal{P}_{\epsilon}(\mathbf{T}, \mu)$  of  $(\mathbf{T}, \mu)$  with at most  $4/\epsilon$  elements.*

*Proof.* For sake of clarity, we construct the desired  $\epsilon$ -partition in two steps. First, let  $\mathcal{P}$  be the partition of  $T$  obtained in the following manner:

- for every  $\epsilon$ -singular vertex  $v$ , keep  $\{v\}$  in its own part, and, for an arbitrary order on the  $\epsilon$ -light children of  $u$ , group the  $\{T_u; u \text{ } \epsilon\text{-light child of } v\}$  greedily such that the  $\mu$ -measure of each part is maximum while remaining less than  $\epsilon$ .
- for every  $\epsilon$ -branching or  $\epsilon$ -chaining vertex  $v$ , group  $v$  with  $\{T_u; u \text{ } \epsilon\text{-light child of } v\}$ ,

It is easily seen that  $\mathcal{P}$  is indeed an  $\epsilon$ -partition of  $(\mathbf{T}, \mu)$  (with only parts of type 1, 2, 3). However, the number of parts is not bounded by  $O(1/\epsilon)$ .

Now, as per the definition, an  $\epsilon$ -chaining vertex has at most one  $\epsilon$ -chaining child, and the previous construction never groups the two together. Hence, we can consider chains of parts rooted at  $\epsilon$ -chaining vertices, the parts of which we merge greedily (starting from the closest to the root, and going from parent to child) so that each part has maximum  $\mu$ -measure while remaining less than  $\epsilon$ , thus yielding a partition  $\mathcal{P}_{\epsilon}(\mathbf{T}, \mu)$ , in which every part is either:

- (1) an  $\epsilon$ -singular vertex alone,
- (2) a subtree of  $T$  rooted at an  $\epsilon$ -branching vertex,
- (3) or  $\bigcup_{x \in F} T_x$  for  $F$  a set of children of some  $\epsilon$ -singular vertex.
- (4)  $T_u \setminus T_v$  for an  $\epsilon$ -chaining vertex  $u$  and a descendent  $v$  of  $u$ ,

Note that these four categories correspond exactly to the 4 types described in Definition 3, and since by definition, all parts are indeed of  $\mu$ -measure at most  $\epsilon$ , we get an  $\epsilon$ -partition of  $(\mathbf{T}, \mu)$ . Now we need to prove that the number of parts is bounded.

First, note that there are at most  $1/\epsilon$  sets of the first kind. Indeed, to each singular vertex  $v$  correspond a subtree ( $T_v$  minus the sets  $T_u$  for non- $\epsilon$ -light children  $u$  of  $v$ ) of measure at least  $\epsilon$ , and each of these subtrees are disjoint.

Sets of type 2 correspond to  $\epsilon$ -branching vertices. Consider the tree obtained by deleting  $\epsilon$ -light vertices. We obtain a rooted tree, in which every chaining vertex has exactly one child, branching vertices have at least two children, and leaves are necessarily singular vertices. Therefore the number of branching vertices is at most the number of singular vertices, hence there are at most  $1/\epsilon$  sets of type 2.

Finally, note that in the greedy construction of sets of both types 3 or 4, we apply a similar principle : we have a collection of disjoint sets, each of measure at most  $\epsilon$ , and of total measure, say  $p$ , and we partition this collection by forming groups of total measure at most  $\epsilon$ . It is an easy observation that this can be done using at most  $2p/\epsilon$  groups : one can sets greedily by decreasing order of measure – this insures that all groups but the last have weight at least  $\epsilon/2$ . Moreover these groups are overall all disjoint (a non-light vertex cannot be the descendent of a light one), so the total number of parts of type 3 and 4 is at most  $2/\epsilon$ . This concludes our proof.  $\square$

A partition  $\mathcal{P}$  is said to be a *refinement* of another one  $\mathcal{P}'$  if each element of  $\mathcal{P}$  is a subset of an element of  $\mathcal{P}'$ .

**Lemma 5.** *Let  $\epsilon' < \epsilon$ ,  $\mathbf{T}$  be a (finite) tree-semilattice, and  $\mathcal{P}$  be an  $\epsilon$ -partition of  $\mathbf{T}$ . Then there exists an  $\epsilon'$ -partition  $\mathcal{P}'$  of  $\mathbf{T}$  with at most  $81/\epsilon'$  elements, that is a refinement of  $\mathcal{P}$ . Moreover,  $\mathcal{P}$  induces an  $\epsilon$ -partition of the tree-semilattice  $T/\mathcal{P}'$ .*

*Proof.* Each part not of type 3 is a tree with total measure at most  $\epsilon$ , so we can apply Lemma 4 independetly on each of these trees. For parts of type 3, we start by putting the attachment vertex back in the part, and then again apply Lemma 4, before removing it. Thus, we obtain an  $\epsilon'$ -partition of  $\mathbf{T}$  with at most  $16/\epsilon'$  elements.  $\square$

**6.2. Approximations.** In the previous subsection, we defined  $\epsilon$ -partitions and described the quotient map and quotient tree associated to an  $\epsilon$ -partition. These are convenient objects to represent the partition (in particular with respects to successive refinement) but they miss some information about the measure and the colors. This is why we introduce her the concept of  $f$ -reduction. We first give the definition and give two easy but essential lemmas before showing how to construct the particular "small" reductions that will be of use to construct approximations in the proof of our main Theorem.

**Definition 5.** *Let  $(\mathbf{T}, \mu)$  be a tree-semilattice and let  $f : \mathbf{T} \rightarrow \tilde{\mathbf{T}}$  be an  $\epsilon$ -partition function of  $(\mathbf{T}, \mu)$ .*

*An  $f$ -reduction of  $\mathbf{T}$  is a color-preserving mapping  $\pi : \mathbf{T} \rightarrow \hat{\mathbf{T}}$ , where  $\hat{\mathbf{T}}$  is a tree-semilattice, such that  $f$  factorizes as  $\hat{f} \circ \pi$ , where  $\hat{f}$  is an  $\epsilon$ -partition function of  $(\hat{\mathbf{T}}, \mu \circ \pi^{-1})$ , and  $\pi$  satisfies the property that for every  $x, y \in T$*

with  $f(x) \neq f(y)$  it holds

$$\begin{aligned}\pi(x \wedge y) &= \pi(x) \wedge \pi(y) \\ x \leq y &\iff \pi(x) \leq \pi(y)\end{aligned}$$

Such a situation we depict by the following commutative diagram:

$$\begin{array}{ccc}\mathbf{T} & \xrightarrow{\pi} & \hat{\mathbf{T}} \\ f \downarrow & \searrow f & \\ \tilde{\mathbf{T}} & & \end{array}$$

**Lemma 6.** *Let  $(\mathbf{T}, \mu)$  be a tree-semilattice, let  $\epsilon > 0$ , let  $f : \mathbf{T} \rightarrow \tilde{\mathbf{T}}$  be an  $\epsilon$ -partition function of  $(\mathbf{T}, \mu)$ , let  $\pi : \mathbf{T} \rightarrow \hat{\mathbf{T}}$  be an  $f$ -reduction of  $\mathbf{T}$ , and let  $v_1, \dots, v_p \in T$  be such that  $f(v_i) \neq f(v_j)$  whenever  $v_i \neq v_j$ .*

*Then  $\pi$  induces an isomorphism of  $\mathbf{T}\langle v_1, \dots, v_p \rangle$  and  $\hat{\mathbf{T}}\langle \pi(v_1), \dots, \pi(v_p) \rangle$*

*Proof.* By assumption, for every distinct  $v_i, v_j$  it holds  $f(v_i) \neq f(v_j)$  thus  $\pi(v_i \wedge v_j) = \pi(v_i) \wedge \pi(v_j)$ . As  $\pi$  is color preserving, in order to prove that  $\pi$  is an homomorphism, we only have to check (as  $\wedge$  is associative) that if  $x \in \mathbf{T}\langle v_1, \dots, v_p \rangle$  and  $1 \leq k \leq p$  then  $\pi(x \wedge v_k) = \pi(x) \wedge \pi(v_k)$ . According to Remark 1, there exist  $i, j$  such that  $x = v_i \wedge v_j$ . Moreover, there exists a permutation  $i', j', k'$  of  $i, j, k$  such that  $v_i \wedge v_j \wedge v_k = v_{i'} \wedge v_{k'} = v_{j'} \wedge v_{k'}$ . Hence  $\pi(v_i \wedge v_j \wedge v_k) = \pi(v_{i'} \wedge v_{k'}) = \pi(v_{i'}) \wedge \pi(v_{k'})$  and similarly  $\pi(v_i \wedge v_j \wedge v_k) = \pi(v_{j'}) \wedge \pi(v_{k'})$ . Thus  $\pi(x \wedge v_k) = \pi(v_i \wedge v_j \wedge v_k) = \pi(v_{i'} \wedge \pi(v_{j'}) \wedge \pi(v_{k'})) = \pi(v_i \wedge \pi(v_j) \wedge \pi(v_k)) = \pi(x) \wedge \pi(v_k)$ .

In order to prove that  $\pi$  is an isomorphism, we have to check that  $\pi$  is injective, that is, according to Remark 1, that  $\pi(v_i \wedge v_j) = \pi(v_k \wedge v_\ell)$  not only implies  $f(v_i \wedge v_j) = f(v_k \wedge v_\ell)$ , but also implies  $v_i \wedge v_j = v_k \wedge v_\ell$ . Let  $v_i, v_j, v_k, v_\ell$  be all distinct. Then  $f(v_i \wedge v_j)$  is of type 1 or 2, thus  $v_i \wedge v_j$  and  $v_k \wedge v_\ell$  coincides with the attachment vertex of  $f(v_i \wedge v_j)$  thus  $v_i \wedge v_j = v_k \wedge v_\ell$ . Also, if  $\pi(v_i \wedge v_j) = \pi(v_k)$  then  $f(v_k)$  has type 1 or 2. As  $\pi(v_k) \leq \pi(v_i)$  it holds  $v_k \leq v_i$  hence  $v_k$  is the attachment vertex of  $f(v_k)$  thus  $v_k = v_i \wedge v_j$ . If  $\pi(v_i \wedge v_k) = \pi(v_j \wedge v_k)$  then either  $f(v_i \wedge v_k)$  has type 1 or 2, in which case  $v_i \wedge v_k = v_j \wedge v_k$ , or  $f(v_i \wedge v_k)$  has type 4 and again  $v_i \wedge v_k = v_j \wedge v_k$ . Last, if  $\pi(v_i \wedge v_j) = \pi(v_j)$  then  $\pi(v_j) \leq \pi(v_i)$  thus  $v_j \leq v_i$ , that is  $v_i \wedge v_j = v_j$ .  $\square$

**Lemma 7.** *Let  $(\mathbf{T}, \mu)$  be a tree-semilattice, let  $\epsilon > 0$ , let  $f : \mathbf{T} \rightarrow \tilde{\mathbf{T}}$  be an  $\epsilon$ -partition function of  $(\mathbf{T}, \mu)$ , let  $\pi : \mathbf{T} \rightarrow \hat{\mathbf{T}}$  be an  $f$ -reduction of  $\mathbf{T}$ , and let  $\hat{\mu} = \mu \circ \pi^{-1}$ .*

*Then for every quantifier free formula  $\phi \in \text{QF}_p$  it holds*

$$|\langle \phi, (\mathbf{T}, \mu) \rangle - \langle \phi, (\hat{\mathbf{T}}, \hat{\mu}) \rangle| < p^2 \epsilon.$$

Thus,

$$\text{dist}((\mathbf{T}, \mu), (\hat{\mathbf{T}}, \hat{\mu})) < 6\epsilon.$$



*Proof.* Let  $\hat{f} : \hat{\mathbf{T}} \rightarrow \tilde{\mathbf{T}}$  be the  $\epsilon$ -partition function associated to the  $f$ -reduction. Define the sets

$$\begin{aligned} A &= \{(v_1, \dots, v_p) \in \phi(\mathbf{T}) : \forall i \neq j (f(v_i) = f(v_j)) \rightarrow (v_i = v_j)\} \\ \hat{A} &= \{(z_1, \dots, z_p) \in \phi(\hat{\mathbf{T}}) : \forall i \neq j (\hat{f}(z_i) = \hat{f}(z_j)) \rightarrow (z_i = z_j)\} \end{aligned}$$

According to Lemma 6, it holds  $\mu^{\otimes p}(A) = \hat{\mu}^{\otimes p}(\hat{A})$ . As

$$\begin{aligned} &|\langle \phi, (\mathbf{T}, \mu) \rangle - \mu^{\otimes p}(A)| \\ &\leq \mu^{\otimes p}(\{(v_1, \dots, v_p) \in T : \exists i \neq j (f(v_i) = f(v_j)) \wedge \neg(v_i = v_j)\}) \\ &\leq \binom{p}{2} \max\{\mu(f^{-1}(p)) : p \in \tilde{\mathbf{T}} \text{ and } |f^{-1}(p)| > 1\} \\ &< p^2 \epsilon / 2 \end{aligned}$$

and as the same holds for  $|\langle \phi, (\hat{\mathbf{T}}, \hat{\mu}) \rangle - \hat{\mu}^{\otimes p}(\hat{A})|$  we deduce that  $|\langle \phi, (\mathbf{T}, \mu) \rangle - \langle \phi, (\hat{\mathbf{T}}, \hat{\mu}) \rangle| < p^2 \epsilon$ . As  $\sum_p p^2 2^{-p} = 6$  the bound on the distance follows.  $\square$

We can now define the particular  $f$ -reductions, called *standard  $f$ -reductions* that we will consider to construct approximations. The fact that this construction yields indeed an  $f$ -reduction is proven in Lemma 8.

**Definition 6.** Let  $(\mathbf{T}, \mu)$  be a tree-semilattice, let  $\epsilon > 0$ , and let  $f : \mathbf{T} \rightarrow \tilde{\mathbf{T}}$  be an  $\epsilon$ -partition function of  $(\mathbf{T}, \mu)$ .

The standard  $f$ -reduction  $\pi : \mathbf{T} \rightarrow \hat{\mathbf{T}}$  of  $(\mathbf{T}, \mu)$  is defined as follows. For each part  $P$  of  $\mathcal{P}$  we associate a rooted tree (or a rooted forest)  $\mathbf{Y}_P$  and we define the projection  $\pi$  from  $T$  to the domain  $\hat{T} = \bigcup Y_P$  of  $\hat{\mathbf{T}}$  as follows:

- If  $P$  is of type 1 ( $P = \{v\}$ ):  
Then  $\mathbf{Y}_P$  is a single-node rooted tree,  $\pi$  maps  $v$  to the only vertex in  $\mathbf{Y}_P$ , which is assigned the same color and weight as  $v$ .
- If  $P$  is of type 2 ( $P = \{v\} \cup \bigcup_{x \in F} T_x$ ):  
Then  $\mathbf{Y}_P$  is rooted at a vertex  $a_P$  (having same color and weight as  $v$ ) that has, for each color  $\gamma$  present in  $P \setminus \{v\}$ , a child  $b_{P,\gamma}$  of color  $\gamma$  whose weight is the sum of the weights of  $\gamma$ -colored vertices of  $P \setminus \{v\}$ ; the projection  $\pi$  maps  $v$  to  $a_P$  and  $u \in P \setminus \{v\}$  to  $b_{P,c(u)}$ .
- If  $P$  is of type 3 ( $P = \bigcup_{x \in F} T_x$ ):  
Then  $\mathbf{Y}_P$  is a set of single node rooted trees, with roots  $b_{P,\gamma_1}, \dots, b_{P,\gamma_\ell}$ , where  $\gamma_1, \dots, \gamma_\ell$  are the colors present in  $P$ . The projection  $\pi$  maps each vertex  $u \in P$  to  $b_{P,c(u)}$ , and the weight of  $b_{P,\gamma_i}$  is defined as the sum of the weights of the  $u \in P$  with color  $\gamma_i$ .
- If  $P$  is of type 4 ( $P = T_v \setminus T_w$ ):  
Then  $\mathbf{Y}_P$  is a rooted caterpillar with spine  $a_{P,\gamma_1}, \dots, a_{P,\gamma_\ell}$  rooted at  $a_{P,\gamma_1}$ , where  $\gamma_1, \dots, \gamma_\ell$  are the colors present in the spine  $S$  of  $P$ ; the projection  $\pi$  maps each vertex  $u \in S$  to  $a_{P,c(u)}$ , and the weight of  $a_{P,\gamma_i}$  is defined as the sum of the weights of the  $u \in S$  with color  $\gamma_i$ . Each vertex  $a_{P,\gamma_i}$  has sons  $b_{P,\gamma_i,\rho_{i,j}}$  ( $1 \leq j \leq s_i$ ) where  $\rho_{i,1}, \dots, \rho_{i,s_i}$

are the (distinct) colors of the vertices  $u \in P \setminus S$  such that  $u \wedge w$  has color  $\gamma_i$ . The projection  $\pi$  maps each  $u \in P \setminus S$  to  $b_{P,c(u \wedge w),c(u)}$ .

The colored tree-semilattice  $\hat{\mathbf{T}}$  is defined by the rooted tree  $\mathbf{T}_{\mathcal{P}}$ , which is constructed from the disjoint union  $\bigcup_{P \in \mathcal{P}} \mathbf{Y}_P$  as follows: For each non root part  $P$  of  $\mathcal{P}$  with father part  $P'$ , the node of  $\mathbf{Y}_{P'}$  that will serve as the father of the root(s) of  $\mathbf{Y}_P$  is defined as follows: if  $P'$  has type 1 or 2, then the father of the root of  $\mathbf{Y}_P$  is the root of  $\mathbf{Y}_{P'}$ ; if  $P$  is of type 4, then the father of the root of  $\mathbf{Y}_P$  is the maximum vertex of the spine of  $\mathbf{Y}_{P'}$ .

Before proving properties of standard  $f$ -reductions, let us illustrate our definitions by an example drawn on Figure 2. The tree  $\mathbf{T}$  has 80 vertices with uniform weight  $1/80$  and suppose  $\epsilon = 1/5$ . Consider for instance the

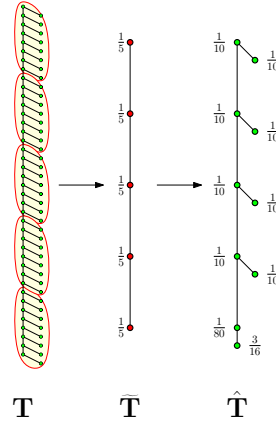


FIGURE 2. Example of  $\mathbf{T}$ ,  $\tilde{\mathbf{T}}$ ,  $\hat{\mathbf{T}}$ .

formula  $\phi : (x_1 \wedge x_2 \neq x_1) \& (x_1 \wedge x_2 \neq x_2)$ . Then  $\langle \phi, \mathbf{T} \rangle = 0.4875$  and  $\langle \phi, \tilde{\mathbf{T}} \rangle = 0.4$  are close, but  $\langle \phi, \hat{\mathbf{T}} \rangle = 0$ .

**Lemma 8.** *Let  $(\mathbf{T}, \mu)$  be a tree-semilattice, let  $\epsilon > 0$ , and let  $f : \mathbf{T} \rightarrow \tilde{\mathbf{T}}$  be an  $\epsilon$ -partition function of  $(\mathbf{T}, \mu)$ .*

*Then the standard  $f$ -reduction  $\pi : \mathbf{T} \rightarrow \hat{\mathbf{T}}$  of  $\mathbf{T}$  is an  $f$ -reduction and*

$$|\hat{\mathbf{T}}| \leq C(C+1)|\tilde{\mathbf{T}}|,$$

*where  $C$  denotes the number of colors for elements of  $T$ .*

*Proof.* Let  $\hat{\mu} = \mu \circ \pi^{-1}$ . As  $\pi(x) = \pi(y)$  imply  $f(x) = f(y)$  there exists a mapping  $\hat{f} : \hat{\mathbf{T}} \rightarrow \tilde{\mathbf{T}}$  such that  $f = \hat{f} \circ \pi$ . Moreover, it is easily checked that  $\hat{f}$  is an  $\epsilon$ -partition function of  $(\hat{\mathbf{T}}, \hat{\mu})$ .

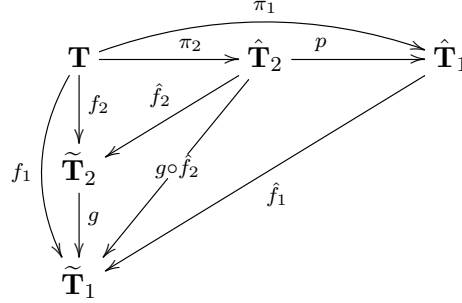
For every  $u, v$  in  $T$  with  $f(u) \neq f(v)$ , the element  $\pi(u) \wedge \pi(v)$  belongs to the part  $Y_{\pi(u \wedge v)}$ , which has the same type as the part  $P$  of  $u \wedge v$ . If  $P$  has type 1 or 2 then  $u \wedge v = A_{\mathcal{P}}(P)$  and  $\pi(u) \wedge \pi(v) = A_{\hat{\mathcal{P}}}(Y_{\pi(u \wedge v)}) = \pi(A_{\mathcal{P}}(P)) = \pi(u \wedge v)$ . If  $P$  has type 3 then one of  $u$  and  $v$  has to belong to  $P$  (as  $P$  has only one child) and the other one has to belong to a part greater than  $P$ . Without loss

of generality assume  $u > v$  and  $v \in P$ . Then  $u \wedge v = w \wedge v$ , where  $w$  denotes the cut-vertex of  $P$ . Similarly, we have  $\pi(u) \wedge \pi(v) = \hat{w} \wedge \pi(v)$ , where  $\hat{w}$  denotes the cut-vertex of  $Y_P$ . By construction, it holds  $\pi(w \wedge v) = \hat{w} \wedge \pi(v)$  hence  $\pi(u \wedge v) = \pi(u) \wedge \pi(v)$ . It follows that  $\pi$  is an  $f$ -reduction.  $\square$

**Lemma 9.** *Let  $(\mathbf{T}, \mu)$  be a tree-semilattice, let  $\epsilon > 0$ , let  $f_1 : \mathbf{T} \rightarrow \tilde{\mathbf{T}}_1$  (resp.  $f_2 : \mathbf{T} \rightarrow \tilde{\mathbf{T}}_2$ ) be an  $\epsilon_1$ -partition function (resp. an  $\epsilon_2$ -partition function) of  $(\mathbf{T}, \mu)$ .*

*Let  $\pi_1 : \mathbf{T} \rightarrow \hat{\mathbf{T}}_1$  (resp.  $\pi_2 : \mathbf{T} \rightarrow \hat{\mathbf{T}}_2$ ) be the standard  $f_1$ -reduction (resp. the standard  $f_2$ -reduction) of  $\mathbf{T}$  of  $\mathbf{T}$ .*

*If there exists a mapping  $g : \tilde{\mathbf{T}}_2 \rightarrow \tilde{\mathbf{T}}_1$  (i.e. if partition  $f_2$  refines partition  $f_1$ ) then the standard  $(g \circ \hat{f}_2)$ -reduction  $p$  of  $\mathbf{T}$  maps  $\hat{\mathbf{T}}_2$  onto  $\hat{\mathbf{T}}_1$  in such a way that the following diagram commutes:*



*Proof.* The tedious but easy check is omitted here.  $\square$

### 6.3. Limits of Tree-Semilattices.

**Theorem 2.** *Every QF-convergent sequence  $(\mathbf{T}_n, \mu_n)$  of tree-semilattices has a Borel tree-semilattice limit  $(\hat{\mathbf{T}}, \hat{\mu})$ .*

*Proof.* We consider a decreasing sequence  $\epsilon_1 > \epsilon_2 > \dots$  of positive reals with limit 0. Using Lemma 4 and Lemma 5, we can construct a sequence  $\mathcal{P}_{n,1}, \mathcal{P}_{n,2}, \dots$  of partitions such that  $\mathcal{P}_{n,i}$  is an  $\epsilon_i$ -partition of  $(\mathbf{T}_n, \mu_n)$  with at most  $16/\epsilon_i$  vertices and  $\mathcal{P}_{n,i+1}$  is a refinement of  $\mathcal{P}_{n,i}$ .

For every fixed  $i$  there are only a finite number of trees on  $16/\epsilon_i$  vertices, so up to extracting subsequences (we do it sequentially starting with  $i = 1$ ), we can assume that all the  $\epsilon_i$ -partition functions corresponding to the  $(\mathcal{P}_{n,i})_{n \in \mathbb{N}}$  have the same target tree  $\tilde{\mathbf{T}}_i$ .

And in fact the same finiteness argument is true also for standard reductions, so we can also assume that all the standard  $f_{n,i}$ -reductions  $\pi_{n,i}$  have the same target  $\hat{\mathbf{T}}_i$  (and partition map  $\hat{f}_i$ ), and by a compactness argument we can also assume that for each  $i$  the measures  $\hat{\mu}_{n,i}$  converge weakly to some measure  $\hat{\mu}_i$ . We further denote by  $g_i$  the map from  $\tilde{\mathbf{T}}_{i+1}$  to  $\tilde{\mathbf{T}}_i$  witnessing that  $\mathcal{P}_{i+1}$  is a refinement of  $\mathcal{P}_i$ , by  $h_i$  the composition  $g_i \circ \hat{f}_{i+1}$ , and by  $p_i : \hat{\mathbf{T}}_{i+1} \rightarrow \hat{\mathbf{T}}_i$  the standard  $h_i$ -reduction of  $\hat{\mathbf{T}}_{i+1}$ .

We define the set  $\hat{T}$  as

$$\hat{T} = \left\{ u \in \prod_{i \in \mathbb{N}} \hat{T}_i : (\forall n \in \mathbb{N}) p_n(u_{n+1}) = u_n \right\},$$

and we denote by  $\zeta_i : \hat{T} \rightarrow \hat{T}_i$  the projection to the  $n$ th coordinate. Then we have the following diagram:

$$\begin{array}{ccccccc}
 \hat{\mathbf{T}} & & & & & & \\
 \vdots & \searrow^{\zeta_{i+1}} & & \searrow^{\zeta_i} & \searrow^{\zeta_{i-1}} & & \\
 & \hat{\mathbf{T}}_{i+1} & \xrightarrow{p_i} & \hat{\mathbf{T}}_i & \xrightarrow{p_{i-1}} & \hat{\mathbf{T}}_{i-1} & \longrightarrow \dots \\
 & \downarrow \hat{f}_{i+1} & \searrow h_i & \downarrow \hat{f}_i & \searrow h_{i-1} & \downarrow \hat{f}_{i-1} & \\
 \dots & & & & & & \dots \\
 & \tilde{\mathbf{T}}_{i+1} & \xrightarrow{g_i} & \tilde{\mathbf{T}}_i & \xrightarrow{g_{i-1}} & \tilde{\mathbf{T}}_{i-1} & \longrightarrow \dots
 \end{array}$$

Note that  $\hat{T}$  (with product topology) is a Cantor space, hence (with Borel  $\sigma$ -algebra) is a standard Borel space. We define the color  $c(u)$  of  $u \in \hat{T}$  as the color of  $u_1$ . As all mappings  $p_n$  are color-preserving, it follows that  $c(u)$  is the color common to all of the  $u_n$ .

For  $u, v \in \hat{T}$ , and  $n \in \mathbb{N}$ , we note  $u_n \sim v_n$  if :

- $\hat{f}_n(u_n) = \hat{f}_n(v_n)$
- either  $u_n$  and  $v_n$  are equal or  $\hat{f}_n(v_n)$  is of type 4 and they are brothers (meaning they both don't belong to the spine and both are children of their infimum  $u_n \wedge v_n$ ).

We then say that  $u$  and  $v$  are *similar* and we note  $u \sim v$  if for every integer  $n$   $u_n \sim v_n$ . Note that  $\sim$  is an equivalence relation on  $\hat{T}$ , and that

$$(u = v) \iff (u \sim v) \wedge (c(u) = c(v))$$

We define the mapping  $\wedge : \hat{T} \times \hat{T} \rightarrow \hat{T}$  as follows: Let  $u, v \in \hat{T}$ . If  $u \sim v$  then  $u \wedge v$  is the one of  $u, v$  with smaller color; otherwise, denoting  $a$  an integer such that either  $\hat{f}_a(u_a) \neq \hat{f}_a(v_a)$  or  $u_a$  and  $v_a$  are neither equal or brothers, we define  $(u \wedge v)_n = u_n \wedge v_n$  if  $n \geq a$ , and (inductively)  $(u \wedge v)_n = p_n((u \wedge v)_{n+1})$  if  $n < a$ .

The mapping  $\wedge$  is clearly symmetric, and it is easily checked to be measurable and associative. Let  $x, y, z \in \hat{T}$ . Assume there exists  $i, j, k \in \mathbb{N}$  such that  $x_i \not\sim y_i$ ,  $y_j \not\sim z_j$ , and  $x_k \not\sim y_k$ . Let  $\ell = \max(i, j, k)$ . Then for every  $n \geq \ell$  it holds  $|\{(x \wedge y)_n, (x \wedge z)_n, (y \wedge z)_n\}| \leq 2$ . Hence either for every  $n \in \mathbb{N}$  it holds  $|\{(x \wedge y)_n, (x \wedge z)_n, (y \wedge z)_n\}| = 1$  (thus  $x \wedge y = x \wedge z = y \wedge z$ ) or  $|\{(x \wedge y)_n, (x \wedge z)_n, (y \wedge z)_n\}| = 2$  for sufficiently large  $n$ , which implies  $|\{x \wedge y, x \wedge z, y \wedge z\}| = 2$ .

Otherwise, assume that there exists  $i \in \mathbb{N}$  such that  $x_i \not\sim y_i$ , but that  $y_n \sim z_n$  holds for every  $n \in \mathbb{N}$ . Then it is easily checked that for  $n > i$  either

it holds  $x_n \wedge y_n = x_n \wedge z_n$ , or that it holds  $x_n \wedge y_n = y_n$  and  $x_n \wedge z_n = z_n$ . In both cases we have  $|\{x \wedge y, x \wedge z, y \wedge z\}| \leq 2$ .

Otherwise,  $x_n \sim y_n$ ,  $y_n \sim z_n$  and  $x_n \sim z_n$  hold for every  $n \in \mathbb{N}$ . Then obviously  $|\{x \wedge y, x \wedge z, y \wedge z\}| \leq 2$ .

Altogether, according to Remark 1, it follows that  $\hat{\mathbf{T}}$  is a tree-semilattice.

For  $i \in \mathbb{N}$  we consider an arbitrary probability measure  $\hat{\lambda}_i$  on  $\hat{T}$  with  $\hat{\lambda}_i(\zeta_i^{-1}(x)) = \hat{\mu}_i(\{x\})$  for every  $x \in \hat{T}_i$ . As all of the  $p_n$  are measure preserving, we have that for every  $n > i$  and every  $x \in \hat{T}_i$  it holds

$$\begin{aligned} \hat{\lambda}_n(\zeta_i^{-1}(x)) &= \hat{\lambda}_n \circ \zeta_n^{-1} \circ p_{n-1}^{-1} \circ \cdots \circ p_i^{-1}(x) \\ &= \hat{\mu}_n \circ p_{n-1}^{-1} \circ \cdots \circ p_i^{-1}(x) \\ &= \hat{\mu}_i(\{x\}) \\ &= \hat{\lambda}_i(\zeta_i^{-1}(x)) \end{aligned}$$

As the sets  $\zeta_i^{-1}(\{x\})$  (for  $i \in \mathbb{N}$  and  $x \in \hat{T}_i$ ) generate the open sets of  $\hat{T}$ , it follows that measures  $\hat{\lambda}_i$  converge weakly to a probability measure  $\hat{\mu}$  as  $i$  grows to infinity.

Let  $\phi$  be a quantifier-free formula with  $p$  free variables. Let  $u^1, \dots, u^p$  be random elements of  $\hat{\mathbf{T}}$  chosen independently with respect to probability measure  $\hat{\mu}$ . With probability at least  $1 - \binom{p}{2}\epsilon_q$  no two distinct  $u^i$ 's have the same  $\zeta_q$  projection and thus  $\hat{\mathbf{T}}\langle u^1, \dots, u^p \rangle$  is isomorphic to  $\hat{\mathbf{T}}_q\langle u_q^1, \dots, u_q^p \rangle$ . As in Lemma 7, we deduce that

$$|\langle \phi, (\hat{\mathbf{T}}, \hat{\mu}) \rangle - \langle \phi, (\hat{\mathbf{T}}_q, \hat{\mu}_q) \rangle| < p^2 \epsilon_q / 2.$$

Moreover, as  $n$  grows to infinity,  $\hat{\mu}_{n,q}$  converges weakly to  $\hat{\mu}_q$ . It follows that there exists  $n_0$  such that for every  $n \geq n_0$  it holds

$$|\langle \phi, (\hat{\mathbf{T}}_q, \hat{\mu}_q) \rangle - \langle \phi, (\hat{\mathbf{T}}_q, \hat{\mu}_{n,q}) \rangle| < \epsilon_q.$$

Also, according to Lemma 7, it holds

$$|\langle \phi, (\mathbf{T}_n, \mu_n) \rangle - \langle \phi, (\hat{\mathbf{T}}_q, \hat{\mu}_{n,q}) \rangle| < p^2 \epsilon_q / 2.$$

Thus it holds

$$|\langle \phi, (\mathbf{T}_n, \mu_n) \rangle - \langle \phi, (\hat{\mathbf{T}}, \hat{\mu}) \rangle| < (p^2 + 1)\epsilon_q.$$

Hence  $(\hat{\mathbf{T}}, \hat{\mu})$  is a Borel tree-semilattice limit of the sequence  $(\mathbf{T}_n, \mu_n)_{n \in \mathbb{N}}$ .  $\square$

**Corollary 3.** *Let  $(\mathbf{T}_n, \mu_n)$  be a QF-converging sequence of uniform (finite) Borel tree-semilattices. Then there exists an atom-less limit  $(\mathbf{T}, \mu)$ .*

## 7. APPLICATIONS: $m$ -PARTITE COGRAPHS

A *cograph*, or *complement-reducible graph*, is a graph that can be generated from  $K_1$  by complementations and disjoint unions. Some generalizations of cographs have been proposed; e.g., bi-cographs [12] or  $k$ -cographs [14].

A well known characterization of cographs is the following [7].

**Theorem 3.** *A graph  $G$  is a cograph if and only if there exists a rooted tree  $T$  (called cotree), whose set of leaves is  $V(G)$ , whose internal nodes are colored 0 or 1, and such that two vertices  $u, v$  are adjacent in  $G$  if and only if the color of their lowest common ancestor is 1.*

The following generalization was proposed in [11, 10]:

**Definition 7** ( *$m$ -partite cograph*). *An  $m$ -partite cograph is a graph that admits an  $m$ -partite cotree representation, that is a rooted tree  $T$  such that*

- *the leaves of  $T$  are the vertices of  $G$ , and are colored by a label from  $\{1, \dots, m\}$ ,*
- *the internal nodes  $v$  of  $T$  are assigned symmetric functions  $f_v : \{1, \dots, m\} \times \{1, \dots, m\} \rightarrow \{0, 1\}$  with the property that two vertices  $x$  and  $y$  of  $G$  with respective colours  $i$  and  $j$  are adjacent iff their least common ancestor  $v$  in  $T$  has  $f_v(i, j) = 1$ .*

(This formal definition is equivalent to the intuitive one given in Section 2.) Note that cographs are exactly 1-partite cographs.

The class of cographs is random-free, as noticed by Janson [15]. The proof is based on the following characterization of random-free hereditary classes of graphs given by Lovász and Szegedy [19]. (Recall that a class  $\mathcal{C}$  is *hereditary* if every induced subgraph of a graph in  $\mathcal{C}$  is in  $\mathcal{C}$ .)

**Theorem 4.** *A hereditary class of graphs  $\mathcal{C}$  is random-free if and only if there exists a bipartite graph  $F$  with bipartition  $(V_1, V_2)$  such that no graph obtained from  $F$  by adding edges within  $V_1$  and  $V_2$  is in  $\mathcal{C}$ .*

It is well known that a cograph is a graph which does not contain the path  $P_4$  on 4 vertices as an induced subgraph. Similarly, the class of  $m$ -partite cographs is a (well quasi-ordered) hereditary class of graphs defined by a finite number of forbidden induced subgraphs  $\mathcal{F}_m$ , one of which is the path  $P_{3(2^m-1)+1}$  of length  $3(2^m - 1)$  [11, 10]. This allows us to extend the result of Janson:

**Lemma 10.** *For every integer  $m$ , the class of all  $m$ -partite cographs is hereditary and random-free.*

*Proof.* That the class of  $m$ -partite cographs is hereditary is immediate from the existence of an  $m$ -partite representation, as deleting leaves in the representation gives a representation of the corresponding induced subgraph.

The path  $P$  of length  $3(2^m - 1)$  is bipartite. Let  $(V_1, V_2)$  be a bipartition of its vertex set. Assume for contradiction that a graph  $H$  obtained from  $P$  by adding edges within  $V_1$  and  $V_2$  is an  $m$ -partite cograph. Let  $Y$  be an  $m$ -partite cotree representation of  $H$ , and assume the leaves are colored  $1, \dots, m$ . Recolor each leaf in  $V_1$  by adding  $m$  to its color ( $c$  becomes  $c+m$ ). For every internal node  $v$ , the color assigned to  $v$  corresponds to a function  $f_v : \{1, \dots, m\} \times \{1, \dots, m\} \rightarrow \{0, 1\}$ . Replace  $f_v$  by the function  $\hat{f}_v :$

$\{1, \dots, 2m\} \times \{1, \dots, 2m\} \rightarrow \{0, 1\}$  defined as follows: for every  $1 \leq i, j \leq m$ , let

$$\begin{aligned} \hat{f}_v(i, j) &= \hat{f}_v(m+i, m+j) = 0 \\ \hat{f}_v(m+i, j) &= \hat{f}_v(i, m+j) = f_v(i, j). \end{aligned}$$

Then it is easily checked that we just constructed a  $2m$ -cotree representation of  $P$ , what contradicts the fact that  $P$  is not a  $2m$ -partite cograph. It follows that no graph  $H$  obtained from  $P$  by adding edges within  $V_1$  and  $V_2$  is an  $m$ -partite cograph. Hence, according to Theorem 4, the class of all  $m$ -partite cographs is random-free.  $\square$

It follows from Lemma 10 and results in [19] that every left-convergent sequence of  $m$ -partite cographs has a limit which is equivalent to a Borel graph excluding (as induced subgraphs) every graph in  $\mathcal{F}_m$ . Thus it holds

**Proposition 1.** *A Borel graph is the QF-limit of a sequence of finite  $m$ -partite cographs if and only if it excludes all the (finite) graphs in the (finite) set  $\mathcal{F}_m$  as induced subgraphs.*

In order to deal more easily with QF-limits, we consider the weighted colored tree-semilattice  $(\mathbf{T}, \mu)$  corresponding to the rooted tree  $T$ , where  $\mu$  is null on internal vertices of  $T$ , and uniform on the leaves of  $T$ . The interpretation  $\mathfrak{l}(\mathbf{T}, \mu)$  of a colored Borel tree-semilattice  $(\mathbf{T}, \mu)$  is the Borel graph  $(G, \nu)$  whose vertex set is the support of  $\mu$ , where  $\nu$  is the restriction of  $\mu$  to its support, where  $x$  and  $y$  are adjacent if  $\Phi(x, y)$  holds, where  $\Phi$  is the quantifier-free formula asserting that for some  $1 \leq i, j \leq m$  and  $f \in \{0, 1\}^{\{1, \dots, m\} \times \{1, \dots, m\}}$  with  $f(i, j) = 1$  it holds that  $x$  has color  $i$ ,  $y$  has color  $j$ , and  $x \wedge y$  has color  $f$ . Note that  $\nu$  is a probability measure on the vertex set of  $G$  as the domain of  $\mathbf{T}$  is a standard Borel space (from what follows that  $\mu(\text{Supp}(\mu)) = 1$ ).

It is easily checked that if  $(\mathbf{T}, \mu)$  is the finite weighted colored tree-semilattice corresponding to the  $m$ -partite cotree representation of a finite  $m$ -partite cograph  $G$  then  $\mathfrak{l}(\mathbf{T}, \mu) = G$ .

It is easily checked that for every quantifier free formula  $\psi$  there exists a quantifier free formula  $\psi^*$  such that for every colored Borel tree-semilattice  $(\mathbf{T}, \mu)$  it holds

$$\langle \psi, \mathfrak{l}(\mathbf{T}, \mu) \rangle = \langle \psi^*, (\mathbf{T}, \mu) \rangle.$$

**Theorem 5.** *Limits of  $m$ -partite cographs are exactly interpretations by  $\mathfrak{l}$  of atomless colored Borel tree-semilattices.*

*Proof.* Let  $(G_n)_{n \in \mathbb{N}}$  be a left-convergent (hence QF-convergent) sequence of  $m$ -partite cographs. Let  $(\mathbf{T}_n, \mu_n)$  be a colored weighted tree-semilattice such that  $G_n = \mathfrak{l}(\mathbf{T}_n, \mu_n)$ . By compactness, there exists an increasing function  $f : \mathbb{N} \rightarrow \mathbb{N}$  such that the sequence  $(\mathbf{T}_{f(n)}, \mu_{f(n)})_{n \in \mathbb{N}}$  is QF-convergent. Let  $(\hat{\mathbf{T}}, \mu)$  be a QF limit Borel tree-semilattice for the sequence  $(\mathbf{T}_{f(n)}, \mu_{f(n)})_{n \in \mathbb{N}}$ .

Note that  $\mu$  has no atom as  $\lim_{n \rightarrow \infty} \sup_{v \in T_n} \mu_n(\{v\}) = 0$ . Define the Borel graph  $(G, \nu)$  as  $G = \mathfrak{l}(\hat{\mathbf{T}}, \mu)$ . For every quantifier free formula  $\psi$ , it holds

$$\begin{aligned} \lim_{n \rightarrow \infty} \langle \psi, G_n \rangle &= \lim_{n \rightarrow \infty} \langle \psi, G_{f(n)} \rangle \\ &= \lim_{n \rightarrow \infty} \langle \psi, \mathfrak{l}(\mathbf{T}_{f(n)}, \mu_{f(n)}) \rangle \\ &= \lim_{n \rightarrow \infty} \langle \psi^*, (\mathbf{T}_{f(n)}, \mu_{f(n)}) \rangle \\ &= \langle \psi^*, (\mathbf{T}, \mu) \rangle \\ &= \langle \psi, \mathfrak{l}(\mathbf{T}, \mu) \rangle \end{aligned}$$

It follows that the Borel graph  $\mathfrak{l}(\mathbf{T}, \mu)$  is the QF-limit of the sequence  $(G_n)_{n \in \mathbb{N}}$ . To this Borel graph corresponds a random-free graphon, which is thus the left limit of the sequence  $(G_n)_{n \in \mathbb{N}}$ .

Conversely, if  $G = \mathfrak{l}(\hat{\mathbf{T}}, \mu)$  is the interpretation of an atomless colored Borel tree-semilattice  $(\mathbf{T}, \mu)$ , then, using a construction analogous to the one use in the proof of Theorem 1, one gets that  $(\mathbf{T}, \mu)$  is the QF-limit of a sequence of finite weighted colored tree-semilattices  $(\mathbf{T}_n, \mu_n)$ , where  $\mu_n$  is uniform on its support. Then for every integer  $n$  the interpretation  $\mathfrak{l}(\mathbf{T}_n, \mu_n)$  is a finite  $m$ -partite cograph (with uniform measure on its vertex set) and, as above,  $G$  is the limit of the sequence  $(G_n)_{n \in \mathbb{N}}$ .  $\square$

Thus, according to Proposition 1, we get the following extension of the characterization of finite  $m$ -partite cographs:

**Theorem 6.** *For an atomless Borel graph  $G$  the following conditions are equivalent:*

- (1)  $G$  is the QF-limit of a sequence of finite  $m$ -partite cographs;
- (2)  $G$  is equivalent to a random-free graphon  $W$  that is the left limit of a sequence of finite  $m$ -partite cographs;
- (3)  $G$  excludes all graphs in  $\mathcal{F}_m$  as induced subgraphs;
- (4)  $G$  is the interpretation by  $\mathfrak{l}$  of an atomless colored Borel tree-semilattice;

#### REFERENCES

- [1] D. Aldous, *Exchangeability and continuum limits of discrete random structures*, Proceedings of the International Congress of Mathematicians (Hindustan Book Agency, ed.), vol. I, 2010, pp. 141–153.
- [2] D. Aldous and R. Lyons, *Processes on unimodular random networks*, Electronic Journal of Probability **12** (2007), 1454–1508.
- [3] T. Austin, *On exchangeable random variables and the statistics of large graphs and hypergraphs*, Probability Surveys **5** (2008), 80–145.
- [4] I. Benjamini and O. Schramm, *Recurrence of distributional limits of finite planar graphs*, Electron. J. Probab. **6** (2001), no. 23, 13pp.
- [5] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, and K. Vesztegombi, *Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing*, Adv. Math. **219** (2008), no. 6, 1801–1851.



- [6] C. Borgs, J.T. Chayes, L Lovász, V.T. Sós, and K. Vesztergombi, *Convergent sequences of dense graphs II: Multiway cuts and statistical physics*, Annals of Mathematics **176** (2012), 151–219.
- [7] D.G Corneil, H. Lerchs, and L. Stewart Burlingham, *Complement reducible graphs*, Discrete Applied Mathematics **3** (1981), no. 3, 163–174.
- [8] P. Diaconis and S. Janson, *Graph limits and exchangeable random graphs.*, Rend. Mat. Appl., VII. Ser. **28** (2008), no. 1, 33–61 (English).
- [9] G. Elek and B. Szegedy, *Limits of hypergraphs, removal and regularity lemmas. A non-standard approach*, arXiv:0705.2179v1 [math.CO], 2007.
- [10] R. Ganian, P. Hliněný, J. Nešetřil, J. Obdržálek, and P. Ossona de Mendez, *When trees grow low: Shrub-depth and  $m$ -partite cographs*, (2013), submitted.
- [11] R. Ganian, P. Hliněný, J. Nešetřil, J. Obdržálek, P. Ossona de Mendez, and R. Ramadurai, *When trees grow low: Shrubs and fast  $\text{MSO}_1$ , MFCS 2012*, Lecture Notes in Computer Science, vol. 7464, Springer-Verlag, 2012, pp. 419–430.
- [12] V. Giakoumakis and J.-M. Vanherpe, *Bi-complement reducible graphs*, Adv. Appl. Math. **18** (1997), 389–402.
- [13] D. Hoover, *Relations on probability spaces and arrays of random variables*, Tech. report, Institute for Advanced Study, Princeton, NJ, 1979.
- [14] L.-J. Hung and T. Kloks,  *$k$ -cographs are Kruskalian*, Chicago Journal of Theoretical Computer Science **2011** (2011), 1–11, Article 2.
- [15] S. Janson, *Graph limits and hereditary properties*, arXiv:1102.3571, Mar 2013.
- [16] O. Kallenberg, *Probabilistic symmetries and invariance principles*, Probability and Its Applications, Springer, 2005.
- [17] L Lovász, *Large networks and graph limits*, Colloquium Publications, vol. 60, American Mathematical Society, 2012.
- [18] L. Lovász and B. Szegedy, *Limits of dense graph sequences*, J. Combin. Theory Ser. B **96** (2006), 933–957.
- [19] ———, *Regularity partitions and the topology of graphons*, An irregular mind (Szemerédi is 70) (I. Bárány and J. Solymosi, eds.), Bolyai Society Mathematical Studies, vol. 21, Springer, 2010, pp. 415–446.
- [20] C. McDiarmid, *On the method of bounded differences*, Surveys in combinatorics **141** (1989), no. 1, 148–188.
- [21] J. Nešetřil and P. Ossona de Mendez, *A model theory approach to structural limits*, Commentationes Mathematicae Universitatis Carolinae **53** (2012), no. 4, 581–603.
- [22] ———, *Modeling limits in hereditary classes: Reduction and application to trees*, arXiv:1312.0441 [math.CO], December 2013.
- [23] ———, *A unified approach to structural limits (with application to the study of limits of graphs with bounded tree-depth)*, arXiv:1303.6471v2 [math.CO], October 2013.
- [24] ———, *Limits of mappings*, in preparation.

- [25] Y. Zhao, *Hypergraph limits: a regularity approach*, arXiv:1302.1634v3 [math.CO], March 2014.

PIERRE CHARBIT, LIAFA - UNIVERSITÉ PARIS DIDEROT - PARIS 7, CASE 7014, 75205 PARIS CEDEX 13, FRANCE

*E-mail address:* pierre.charbit@liafa.univ-paris-diderot.fr

LUCAS HOSSEINI, CENTRE D'ANALYSE ET DE MATHÉMATIQUES SOCIALES (CNRS, UMR 8557), 190-198 AVENUE DE FRANCE, 75013 PARIS, FRANCE

*E-mail address:* lucas.hosseini@gmail.com

PATRICE OSSONA DE MENDEZ, CENTRE D'ANALYSE ET DE MATHÉMATIQUES SOCIALES (CNRS, UMR 8557), 190-198 AVENUE DE FRANCE, 75013 PARIS, FRANCE — AND — COMPUTER SCIENCE INSTITUTE OF CHARLES UNIVERSITY (IUUK), MALOSTRANSKÉ NÁM.25, 11800 PRAHA 1, CZECH REPUBLIC

*E-mail address:* pom@ehess.fr