

LIMITS, REGULARITY AND REMOVAL FOR FINITE STRUCTURES

ASHWINI AROSKAR AND JAMES CUMMINGS

ABSTRACT. Our work builds on known results for k -uniform hypergraphs including the existence of limits, a Regularity Lemma and a Removal Lemma. Our main tool here is a theory of measures on ultraproduct spaces which establishes a correspondence between ultraproduct spaces and Euclidean spaces. First we show the existence of a limit object for convergent sequences of relational structures and as a special case, we retrieve the known limits for graphs and digraphs. Then we extend this notion to finite models of a fixed universal theory. We also state and prove a Regularity Lemma and a Removal Lemma. We will discuss connections between our work and Razborov’s flag algebras as well.

1. INTRODUCTION

The theory of *graph limits* was initiated by Lovász and Szegedy [5]. The theory associates to each sequence of graphs (G_n) , which is increasing and convergent in an appropriate sense, a “limit object” in the form of a Lebesgue-measurable function called a *graphon*. There are close connections between the theory of graph limits and some well-known structural facts about large finite graphs, such as, the Regularity Lemma and the Removal Lemma. The monograph by Lovász [4] gives a very detailed treatment of graph limits.

Elek and Szegedy [1] developed a theory of measures on ultraproduct spaces and investigated limits, regularity and removal in the context of k -uniform hypergraphs. The limit objects are significantly more complicated in this case: a graphon is a function of two variables, but the limit of a sequence of k -uniform hypergraphs is a function of $2^k - 2$ variables. In this paper we build on the work of Elek and Szegedy to construct and apply limit objects for sequences of finite relational structures, and then extend our results to the setting of finite models of a fixed universal first-order theory T . Razborov’s work on flag algebras [8] also gives a notion of a limit object for sequences of models of T , in the form of a certain kind of \mathbb{R} -algebra homomorphism; we will discuss

the connection between our analytic limit objects and Razborov’s more algebraic limits. The paper is organised as follows:

- In section 2 we cover some necessary background material. The impatient reader should probably skip this section and refer back as necessary. In subsection 2.1 we discuss relational languages and structures for such languages, define various kinds of structure-preserving maps between structures, and then define the important notion of a *convergent sequence of structures*.

In subsection 2.2 we define a scheme for coding an arbitrary n -ary relation by a family of directed hypergraphs, which we will use in section 3. In subsection 2.3 we discuss the elementary theory of ultrafilters, ultraproducts and ultrapowers. In subsection 2.4 we give the basic facts about the “Loeb measure” [3] on an ultraproduct of finite sets, and finally in subsection 2.5 we describe some results by Elek and Szegedy about the Loeb measure.

- Section 3 contains our main results. In subsection 3.1 we show how to associate a limit object to a convergent sequence of relational structures, and prove that we can use the limit object as a “template” to construct a random sequence of structures which resembles (in a precise sense) the original sequence. Our proofs use some of Razborov’s flag algebra machinery, and we briefly discuss the relationship between our work and the theory of flag algebras.

In subsection 3.2 we discuss a version of the technical notion of *hyperpartition*, due originally to Elek and Szegedy [1]. In subsection 3.3 we prove a form of Strong Removal for finite relational structures, and finally in subsection 3.4 we sketch a version of the Regularity Lemma for such structures.

- In section 4 we extend the theory from section 3 to the more general setting of structures which are required to be models of a fixed universal theory T . In particular we prove a version of Strong Removal in this setting.
- The appendix has two parts. In part A we work out our limit theory in a very simple case (one binary relation), and then discuss the relationship between our work and the theories of digraph limits (Offner and Pikhurko [7]) and poset limits (Janson [2]). In part B we prove a technical result needed for the results of section 4.

Our notation is mostly standard; here is a brief review. If X and I are sets then X^I is the set of functions from I to X , which we usually

think of as I -indexed sequences. A typical element of X^I is generally denoted as $\vec{x} = (x_i : i \in I)$, or just as (x_i) when I is clear from the context. We write X^t for the set of t -tuples from X . We denote the set of t -tuples in X^t with no repeated entry as $X^{\underline{t}}$. Note that this is in line with one notation used for the “descending factorial” $n^{\underline{t}} = n!/(n-t)!$, which we will adopt.

As usual $[m] = \{1, \dots, m\}$. Note that in particular $[m]^t$ is the set of t -tuples of elements of $[m]$.

If X is a set and R is an n -ary relation on X , then we use the equivalent notations $R(\vec{x})$ or $\vec{x} \in R$ as convenient.

In an effort to improve readability, we have adopted some typographical conventions:

- Objects related to logic (languages, structures, sets of structures) are denoted by calligraphic letters (\mathcal{L} for languages, \mathcal{M} and \mathcal{N} for structures, \mathcal{F} for sets of structures).
- Indexed families of Lebesgue measurable sets (which play the role of limit objects in our theory) are denoted by gothic letters (e.g. \mathfrak{E} , \mathfrak{F}).
- Following Razborov [8], random objects are denoted by boldface symbols, in particular \mathcal{N} is always a random structure.
- Some families of partitions known as *hyperpartitions* (see subsection 3.2) are denoted by variations of the fancy calligraphic letter \mathcal{H} .

2. PRELIMINARIES

2.1. Structures. We work in a general setting which covers many combinatorially interesting structures. We will fix a *finite relational language* \mathcal{L} , which is specified by giving finitely many relation symbols R_1, \dots, R_n together with a natural number $r_i > 0$ for each symbol R_i ; r_i is called the *arity* of R_i . An \mathcal{L} -*structure* \mathcal{M} is a set X together with r_i -ary relations $R_i^{\mathcal{M}} \subseteq X^{r_i}$, for $1 \leq i \leq n$. We call X the *underlying set* of the structure \mathcal{M} and write $X = |\mathcal{M}|$, so that the cardinality of the underlying set for \mathcal{M} is denoted by $\|\mathcal{M}\|$. When \mathcal{M} is a structure and $A \subseteq |\mathcal{M}|$, we can define a new structure with underlying set A by interpreting each relation symbol R_i as $R_i^{\mathcal{M}} \cap A^{r_i}$, obtaining a substructure $\mathcal{M}|_A$.

The formulae of the relational language are built up from the given relation symbols together with a special binary relation symbol $=$, variable symbols, connectives and quantifiers in the usual way. When we are interpreting formulae of our language, the symbol $=$ will always

be interpreted as equality. Given a formula ϕ , a structure \mathcal{M} and elements $m_1, \dots, m_s \in |\mathcal{M}|$, we follow standard logical usage and write “ $\mathcal{M} \models \phi(m_1, \dots, m_s)$ ” for “ ϕ is true of the elements m_1, \dots, m_s in the structure \mathcal{M} ”.

A *sentence* is a formula in which every variable is in the scope of some quantifier, and which therefore has a definite truth value in every structure. A *theory* is a set of sentences. A *model* of a theory T is a structure in which every sentence from T is true. A formula ϕ is *universal* if ϕ has the form $\forall x_1 \dots \forall x_s \psi$ where ψ contains no quantifiers, and a *universal theory* is a theory consisting of universal sentences.

We will work in the context of a countable universal theory T , and we will assume that T has arbitrarily large finite models. By the well-known compactness theorem from first order logic, this is equivalent to the assertion that T has an infinite model. Since T is universal, whenever \mathcal{M} is a model of T and $A \subseteq |\mathcal{M}|$ we have that the “induced” substructure $\mathcal{M}|_A$ is also a model of T .

In the following sections we will first develop a version of the theory in the setting when $T = \emptyset$, that is a theory of “pure relational structures”. We will then extend the theory to models of T using an analysis of models of T in terms of “forbidden induced substructures”, which we describe in section 4.

Given structures \mathcal{M} and \mathcal{N} :

- A *homomorphism from \mathcal{M} to \mathcal{N}* is a function $f : |\mathcal{M}| \rightarrow |\mathcal{N}|$ such that

$$R_i^{\mathcal{M}}(m_1, \dots, m_{r_i}) \implies R_i^{\mathcal{N}}(f(m_1), \dots, f(m_{r_i}))$$

for all i with $1 \leq i \leq n$ and all r_i -tuples (m_1, \dots, m_{r_i}) of elements of $|\mathcal{M}|$.

- An *embedding of \mathcal{M} into \mathcal{N}* is an injective function $f : |\mathcal{M}| \rightarrow |\mathcal{N}|$ such that

$$R_i^{\mathcal{M}}(m_1, \dots, m_{r_i}) \iff R_i^{\mathcal{N}}(f(m_1), \dots, f(m_{r_i}))$$

for all i with $1 \leq i \leq n$ and all r_i -tuples (m_1, \dots, m_{r_i}) of elements of $|\mathcal{M}|$.

- An *isomorphism between \mathcal{M} and \mathcal{N}* is a bijective embedding of \mathcal{M} into \mathcal{N} . The structures \mathcal{M} and \mathcal{N} are *isomorphic* if and only if there is an isomorphism between them, and in this case we write $\mathcal{M} \simeq \mathcal{N}$.

We now define various quantities which measure (in slightly different senses) the “density of \mathcal{M} in \mathcal{N} ” for finite \mathcal{L} -structures \mathcal{M} and \mathcal{N} . The important quantities for us are p and t_{ind} ; we include the quantities t

and t_0 , which play more of a role in the classical theory of graph and hypergraph limits, for the sake of completeness.

Definition 1. *Let \mathcal{M} and \mathcal{N} be finite \mathcal{L} -structures.*

- (1) *The induced substructure density $p(\mathcal{M}, \mathcal{N})$ is the probability that $\mathcal{N}|_{\mathbf{A}} \simeq \mathcal{M}$, where \mathbf{A} is a subset of $|\mathcal{N}|$ of cardinality $\|\mathcal{M}\|$ chosen uniformly at random.*
- (2) *The homomorphism density $t(\mathcal{M}, \mathcal{N})$ is the probability that a function $\mathbf{f} : |\mathcal{M}| \rightarrow |\mathcal{N}|$ chosen uniformly at random is a homomorphism.*
- (3) *The injective homomorphism density $t_0(\mathcal{M}, \mathcal{N})$ is the probability that an injective function $\mathbf{f} : |\mathcal{M}| \rightarrow |\mathcal{N}|$ chosen uniformly at random is a homomorphism.*
- (4) *The embedding density or induced homomorphism density $t_{\text{ind}}(\mathcal{M}, \mathcal{N})$ is the probability that an injective function $\mathbf{f} : |\mathcal{M}| \rightarrow |\mathcal{N}|$ chosen uniformly at random is an embedding.*

By convention the quantities $p(\mathcal{M}, \mathcal{N})$, $t_0(\mathcal{M}, \mathcal{N})$ and $t_{\text{ind}}(\mathcal{M}, \mathcal{N})$ are zero when $\|\mathcal{N}\| < \|\mathcal{M}\|$.

Definition 2. *A sequence (\mathcal{N}_k) of structures is increasing if and only if $\|\mathcal{N}_k\| \rightarrow \infty$, and convergent if and only if it is increasing and additionally the sequence of induced substructure densities $(p(\mathcal{M}, \mathcal{N}_k))$ converges for every finite structure \mathcal{M} .*

Since there are only countably many finite \mathcal{L} -structures, an easy diagonal argument along the lines of the Bolzano-Weierstrass theorem shows that every increasing sequence has a convergent subsequence.

It is easy to see that if (\mathcal{N}_k) is convergent then each of the sequences $(t(\mathcal{M}, \mathcal{N}_k))$, $(t_0(\mathcal{M}, \mathcal{N}_k))$, and $(t_{\text{ind}}(\mathcal{M}, \mathcal{N}_k))$ converges for all \mathcal{M} , and their limiting values may be computed from the limiting values for $(p(\mathcal{M}, \mathcal{N}_k))$. Given a convergent sequence (\mathcal{N}_k) , we will define a function Φ_p by setting $\Phi_p(\mathcal{M}) = \lim_{k \rightarrow \infty} p(\mathcal{M}, \mathcal{N}_k)$: of course the definition depends on the convergent sequence (\mathcal{N}_k) , but this should always be clear from the context. The functions Φ_t , Φ_{t_0} and $\Phi_{t_{\text{ind}}}$ are defined similarly. It is easy to see that $|t(\mathcal{M}, \mathcal{N}) - t_0(\mathcal{M}, \mathcal{N})|$ is $O(\|\mathcal{N}\|^{-1})$, so that $\Phi_t = \Phi_{t_0}$.

Fact 1. *Let \mathcal{M} and \mathcal{N} be finite \mathcal{L} -structures. Then $t_{\text{ind}}(\mathcal{M}, \mathcal{N}) = p(\mathcal{M}, \mathcal{N})t_{\text{ind}}(\mathcal{M}, \mathcal{M})$.*

Proof. Consider the process of choosing a random injection \mathbf{f} from $|\mathcal{M}|$ to $|\mathcal{N}|$ as a two-step process, in which we first choose the range of \mathbf{f} and then choose its values. For this process to yield an embedding, the first step must yield a set inducing a substructure isomorphic to \mathcal{M} ,

and the second step must yield an isomorphism between \mathcal{M} and this isomorphic copy of \mathcal{M} . \square

It follows that $\Phi_{t_{\text{ind}}}(\mathcal{M}) = \Phi_p(\mathcal{M})t_{\text{ind}}(\mathcal{M}, \mathcal{M})$. The quantity $t_{\text{ind}}(\mathcal{M}, \mathcal{M})$ is the probability that a random permutation of $|\mathcal{M}|$ is an automorphism of \mathcal{M} .

2.2. Decomposing t -ary relations into uniform directed hypergraphs. For technical reasons we will find it helpful to break up t -ary relations into simpler pieces. To see why this is so, consider the language with one binary relation R and let \mathcal{M}_n be a structure with n elements such that $R(x, y) \iff x = y$. Clearly (\mathcal{M}_n) is a convergent sequence, and the probability that two randomly chosen elements are related tends to zero as $n \rightarrow \infty$; but this “two dimensional” information is not enough to determine the limiting behaviour of $(p(\mathcal{M}, \mathcal{M}_n))$ for all finite structures \mathcal{M} . We also need to know the “one dimensional information” that the probability that a single randomly chosen element is related to itself tends to one.

Recall that an r -uniform directed hypergraph is a pair (X, E) where X is a set and $E \subseteq X^r$. Given a t -ary relation R on X , we will associate to R a family of directed hypergraphs which code it. Let Part_t be the set of partitions of $[t]$, and for each partition $p \in \text{Part}_t$ let $\|p\|$ be the number of classes in p . If $\|p\| = r$ then we will enumerate the r classes of p as C_1^p, \dots, C_r^p where $\min(C_i^p)$ increases with i .

Given a t -tuple $\vec{x} \in X^t$, we let $p(\vec{x}) \in \text{Part}_{[t]}$ be the partition in which j and j' lie in the same class if and only if $x_j = x_{j'}$. If $\|p(\vec{x})\| = r$, then we let $C(\vec{x})$ be the r -tuple obtained from \vec{x} by deleting elements from \vec{x} as follows: we retain the first occurrence of each element of $\{x_i : 1 \leq i \leq t\}$, but delete all subsequent occurrences. Given a t -ary relation $R \subseteq X^t$, we let $R_p = \{\vec{x} \in R : p(\vec{x}) = p\}$ and let $\text{DH}_p^R = \{C(\vec{x}) : \vec{x} \in R_p\}$ for each $p \in \text{Part}_t$. For example if R is a 3-ary relation then R will be coded by the directed hypergraphs with edge sets:

- $\text{DH}_{\{\{1,2,3\}\}}^R = \{x \in X : R(x, x, x)\}$.
- $\text{DH}_{\{\{1,2\},\{3\}\}}^R = \{(x, y) \in X^2 : R(x, x, y)\}$.
- $\text{DH}_{\{\{1,3\},\{2\}\}}^R = \{(x, y) \in X^2 : R(x, y, x)\}$.
- $\text{DH}_{\{\{1\},\{2,3\}\}}^R = \{(x, y) \in X^2 : R(x, y, y)\}$.
- $\text{DH}_{\{\{1\},\{2\},\{3\}\}}^R = \{(x, y, z) \in X^3 : R(x, y, z)\}$.

When \mathcal{N} is an \mathcal{L} -structure, we write $\text{DH}_p^{\mathcal{N},i}$ for $\text{DH}_p^{R_i^{\mathcal{N}}}$. The structure \mathcal{N} with underlying set X can be coded by the family $\text{Code}(\mathcal{N}) = (\text{DH}_p^{\mathcal{N},i} : i \in [n], p \in \text{Part}_{[r_i]})$. For notational convenience later we

define Index to be the set of pairs (i, p) with $i \in [n]$ and $p \in \text{Part}_{[r_i]}$, which forms the index set for $\text{Code}(\mathcal{M})$. If $\text{DH} = (\text{DH}_p^i : (i, p) \in \text{Index})$ is a family of directed hypergraphs of the appropriate types on some set X , then we write $\text{Decode}(\text{DH})$ for the unique structure on X coded by DH .

2.3. Ultrafilters and ultraproducts. A *filter* on \mathbb{N} is a family F of subsets of \mathbb{N} such that $\emptyset \notin F$, $\mathbb{N} \in F$, F is upwards closed (in the sense that $A \in F$ and $A \subseteq B \subseteq \mathbb{N}$ implies $B \in F$), and finally F is closed under finite intersections. An *ultrafilter* on \mathbb{N} is a filter U that is maximal under inclusion, equivalently a filter such that for every set A , exactly one of the sets A and $\mathbb{N} \setminus A$ lies in U . It is easy to see that U is an ultrafilter if and only if U is of the form $\{A : \mu(A) = 1\}$, for some finitely additive probability measure μ that is defined on all subsets of \mathbb{N} and takes only the values 0 and 1. If an ultrafilter U contains a finite set then it has the form $\{A : n \in A\}$ for some n and is said to be *principal*: an easy application of Zorn's lemma shows that any infinite set lies in a non-principal ultrafilter. For the rest of this paper we assume that we have fixed a non-principal ultrafilter U on \mathbb{N} . Given a sequence of \mathcal{L} -structures $(\mathcal{M}_k)_{k \in \mathbb{N}}$, a standard construction in logic is the formation of a new \mathcal{L} -structure called the *ultraproduct*. The standard notation for this structure is $\prod_k \mathcal{M}_k / U$; in this discussion we will call it \mathcal{M}_∞ . In the special case when $\mathcal{M}_k = \mathcal{M}$ for all k , we write $\mathcal{M}^\mathbb{N} / U$ for the corresponding ultraproduct which is called the *ultrapower of \mathcal{M} by U* . To define \mathcal{M}_∞ we consider sequences $(x_k) \in \prod_k |\mathcal{M}_k|$, introduce an equivalence relation

$$(x_k) =_U (y_k) \iff \{k : x_k = y_k\} \in U,$$

and let the underlying set $|\mathcal{M}_\infty|$ be the set of equivalence classes. We denote the equivalence class of (x_k) by $[x_k]$ and interpret the language as follows: for each relation symbol R , if R has arity r then

$$R^{\mathcal{M}_\infty}([x_k^1], \dots, [x_k^r]) \iff \{k : R^{\mathcal{M}_k}(x_k^1, \dots, x_k^r)\} \in U.$$

The crucial fact about ultraproducts is Łoś' Theorem: for *any* first-order formula ϕ ,

$$\mathcal{M}_\infty \models \phi([x_k^1], \dots, [x_k^r]) \iff \{k : \mathcal{M}_k \models \phi(x_k^1, \dots, x_k^r)\} \in U.$$

This is proved by a straightforward induction on the structure of the formula ϕ , with the interpretation of the relation symbols as the base case. We note the easy corollary that for any theory T , the class of models of T is closed under ultraproducts. For our purposes it will be very convenient to use ideas from set theory. We recall that the *language of set theory* has one binary relation symbol \in denoting

membership. ZFC set theory is a first-order theory written in this language, whose intended model is the class V of all sets; we can view almost all mathematical objects as sets, and can justify almost all mathematical constructions and proofs on the basis of ZFC. We will form the ultrapower ¹ $V^* = V^{\mathbb{N}}/U$.

By the preceding discussion, V^* is itself a model of ZFC and contains versions of almost all standard mathematical objects. If \in^* is V^* 's version of the membership relation, then we may view each element b of V^* as a set in its own right whose elements are those $a \in V^*$ such that $V^* \models a \in^* b$; with this in mind we will refer to such objects a as the *elements of b* and write “ $a \in b$ ” rather than “ $V^* \models a \in^* b$ ”. In particular we may view the ultraproduct \mathcal{M}_∞ as $[\mathcal{M}_i]$, a structure lying in V^* .

Given any set $x \in V$, we let c_x be the sequence which is constant with value x and then x^* be the element of V^* represented by c_x ; it follows immediately from the definitions and Loś' Theorem that $x \mapsto x^*$ is an injective map and

$$V \models \phi(x_1, \dots, x_n) \iff V^* \models \phi(x_1^*, \dots, x_n^*)$$

for all ϕ . In the terminology of logic we have an “elementary embedding from V to V^* ”.

If X is an infinite set then X^* contains elements which are not of the form x^* for any $x \in X$. What is more, if Y is the power set of X then by elementarity, Y^* is the set of subsets of X^* which lie in V^* ; when X is infinite it turns out that Y^* has elements which are not of the form A^* for any $A \in Y$, and in addition there exist subsets of X^* which are not in Y^* . To clarify these points, consider the example $X = \mathbb{N}$. It is not hard to see that $\{n^* : n \in \mathbb{N}\}$ is an initial segment of \mathbb{N}^* which is isomorphic to \mathbb{N} , and that setting $x_k = k$ we obtain an element $[x_k]$ of \mathbb{N}^* which is greater than n^* for all n . If we now let $y_k = \{0, 1, \dots, k\}$ then by Loś' theorem $V^* \models “[y_k] \text{ is finite}”$, and it follows that $[y_k]$ is a subset of \mathbb{N}^* which lies in V^* but is not of the form A^* for any $A \subseteq \mathbb{N}$. Additionally we claim that the set $\{n^* : n \in \mathbb{N}\}$ is not in V^* ; for if it were, then by elementarity its complement would have a minimal element y say, but by elementarity again y would have a predecessor n^* leading to an immediate contradiction.

¹Formally speaking we should be careful here, because each equivalence class (mod U) of functions is a proper class and it would be improper to form the collection of equivalence classes. This issue can be dealt with by an argument due to Dana Scott [9] in which each equivalence class is replaced by the set of elements within it of minimal rank.

The structure $V^* = V^{\mathbb{N}}/U$ has a compactness property known as \aleph_1 -saturation. This states that if $\{\phi_i(x) : i \in \mathbb{N}\}$ is a set of formulae, and for every j there is $a_j \in V^*$ such that $\phi_i(a_j)$ holds in V^* for all $i \leq j$, then there is $a \in V^*$ such that $\phi_i(a)$ holds in V^* for all i . The argument is a routine diagonalisation using Loś' theorem.

Let X be a compact Hausdorff space and let $y \in X^*$. A routine argument shows that there is a unique $x \in X$ such that $y \in O^*$ for every open $O \ni x$; we call this point x the *standard part* of y and write $x = \text{std}(y)$. A more topological view of this construction can be obtained by observing that if $y = [y_k]$ then x is the unique point such that $\lim_{k \rightarrow U} y_k = x$, that is for all open $O \ni x$ we have $\{k : y_k \in O\} \in U$.

2.4. Measure theory on ultraproducts: basic facts. Measures on ultraproducts of finite sets are a special case of *Loeb measures* [3], which are well-known objects in non-standard analysis. We review some basic facts. Let (X_k) be a sequence of finite sets with $|X_k| \rightarrow \infty$, and let $X_\infty = [X_k]$, so that

$$V^* \models \text{“}X_\infty \text{ is finite”}.$$

In actuality X_∞ is easily seen to be an uncountable set. Our goal is to define a measure on X_∞ . We will start by assigning a measure to those subsets of X_∞ which lie in V^* . Let $A_\infty \in V^*$ be a subset of X_∞ . We begin by working in V^* and computing the measure of A_∞ with respect to the “normalised counting measure” on X_∞ , that is $\frac{|A_\infty|}{|X_\infty|} \in [0, 1]^*$. Since $[0, 1]$ is compact this number has a standard part which we call $\mu(A_\infty)$. For a more concrete definition, we may choose $A_k \subseteq X_k$ such that $A_\infty = [A_k]$ and then check that $\mu(A_\infty) = \lim_{k \rightarrow U} \frac{|A_k|}{|X_k|}$.

The subsets of X_∞ which lie in V^* form a Boolean algebra. To extend μ to a wider class of sets, we say that an arbitrary set $A \subseteq X_\infty$ is *null* if and only if for every $\epsilon > 0$ there is $B \in V^*$ such that $A \subseteq B \subseteq X_\infty$ and $\mu(B) < \epsilon$. We let σ be the class of subsets $A \subseteq X_\infty$ such that $A \Delta B$ is null for some $B \in V^*$, and extend μ onto σ by defining $\mu(A) = \mu(B)$ for some (any) $B \in V^*$ such that $A \Delta B$ is null. The following fact is a consequence of the \aleph_1 -saturation of V^* .

Fact 2. σ is a σ -algebra and μ is a σ -additive probability measure.

2.5. Measure theory on ultraproducts: results of Elek and Szegedy. Elek and Szegedy [1] proved a family of theorems about measure theory on ultraproducts which will play a central role in our results. We give a brief account of the results we need.

Let $A \subseteq [t]$ with $|A| = s$, let A be enumerated in increasing order as i_1, \dots, i_s , and let π_A^t denote the map on the class of t -tuples which

returns for each t -tuple $(a_i : i \in [t])$ the s -tuple $(a_{i_j} : j \in [s])$. Then $\sigma_A^t = \{\pi_A^{t-1}[B] : B \in \sigma_{[s]}\}$, that is to say elements of σ_A^t are measurable sets which depend only on coordinates in A . It is easy to see that π_A^t sets up an isomorphism between $(X_\infty^t, \sigma_A^t, \mu_{[t]} \upharpoonright \sigma_A^t)$ and $(X_\infty^s, \sigma_{[s]}, \mu_{[s]})$. To lighten the notation, we will often drop the superscript t when it is clear from the context. The following result is a version of a well-known theorem of Maharam [6].

Fact 3. *Let (X, τ, ν) be a separable atomless complete measure space with $\nu(X) = 1$. Then the associated measure algebra is isomorphic to the measure algebra of the unit interval with Lebesgue measure. Moreover there is a map $f : X \rightarrow [0, 1]$ defining an isomorphism between these measure algebras, that is:*

- For every measurable $B \subseteq [0, 1]$, $f^{-1}[B] \in \tau$ and $\lambda(B) = \mu(f^{-1}[B])$.
- For every $A \in \tau$ there is a measurable set $B \subseteq [0, 1]$ such that A and $f^{-1}[B]$ differ by a null set.

For a non-empty set $A \subseteq [t]$, we define σ_A^* to be the subalgebra of σ_A generated by the union of the algebras σ_B for $B \subseteq A$, $|B| = |A| - 1$. It is a surprising fact that σ_A is much larger than σ_A^* ; this plays a crucial role in everything that follows.

Let $A \subseteq [t]$ be non-empty and let $0 < s \leq |A|$, then we define $r(A, s)$ to be the set of subsets $B \subseteq A$ with $0 < |B| \leq s$. As a special case $r(A) = r(A, |A|)$, that is the set of non-empty subsets of A . The group $S_{[t]}$ of permutations of $[t]$ acts in the natural way on subsets of $[t]$ by permuting their entries, and on t -tuples by permuting their coordinates, and we write z^σ for the image of z by σ ; we get an induced action on sets Z of t -tuples by defining $Z^\sigma = \{z^\sigma : z \in Z\}$, and so on inductively for sets of sets of t -tuples etcetera. When $A \subseteq [t]$ we write S_A for the group of permutations in $S_{[t]}$ which fix points outside A ; we will often identify this with the permutation group of A in the obvious way.

Definition 3. *A separable system on X_∞^t is a family $\{l_A : A \in r([t])\}$ such that*

- (1) l_A is an atomless separable subalgebra of σ_A , which is independent of σ_A^* (that is to say sets in l_A are independent of σ_A^*).
- (2) $(l_A)^\pi = l_{A^\pi}$ for all A .
- (3) $Y^\pi = Y$ for all $Y \in l_A$ and $\pi \in S_A$.

To illustrate the symmetry and independence properties consider the case when $t = 2$, enumerating $r([2])$ as $\{1\}$, $\{2\}$, $\{1, 2\}$; the separable system consists of algebras $l_{\{1\}}, l_{\{2\}}, l_{\{1, 2\}}$ where exchanging coordinates

1 and 2 exchanges the algebras $l_{\{1\}}$ and $l_{\{2\}}$, while every set in $l_{\{1,2\}}$ is symmetric under this exchange and is independent of rectangles.

It is a key fact that arbitrary measurable sets can be well-approximated by separable systems:

Fact 4. *Let $\langle X_n^B : n \in \mathbb{N}, B \in r([t]) \rangle$ be such that $X_n^B \in \sigma_B$ for all n and B . Then there exists a separable system $\{l_A : A \in r([t])\}$ such that, for every n , X_n^B differs on a set of measure zero from a set in the σ -algebra generated by $\bigcup_{A \subseteq B} l_A$.*

In this situation we will say that the separable system is *suitable* for the sets X_n^B .

Given a separable system on X_∞^t , we can “realise” it in a symmetry-preserving way on the space $[0, 1]^{r([t])}$ (considered as a measure space equipped with Lebesgue measure).

Definition 4. *If $\{l_A : A \in r([t])\}$ is a separable system, then a separable realisation of the system is a measure-preserving map $\phi : X_\infty^t \rightarrow [0, 1]^{r([t])}$ such that (writing $\phi_A(x)$ for the A -component of $\phi(x)$)*

- (1) ϕ_A only depends on coordinates in A .
- (2) For every measurable set $B \subseteq [0, 1]$, $\phi_A^{-1}(B) \in l_A$.
- (3) For every permutation $\sigma \in S_{[t]}$, $\phi_{A^\sigma}(x^\sigma) = \phi_A(x)$.

It follows easily from the properties of such a map ϕ that there exist maps $\phi^j : X_\infty^j \rightarrow [0, 1]$ for $j \in [t]$ such that:

- (1) ϕ^j is invariant under permutation of its arguments.
- (2) For all $A = \{a_1, \dots, a_n\} \in r([t])$, $\phi_A(x_1, \dots, x_t) = \phi^n(x_{a_1}, \dots, x_{a_t})$.

We will refer to the maps ϕ^j as the *levels* of the realisation map ϕ . For example when $t = 2$, we may write $\phi(x_1, x_2) = (\phi^1(x_1), \phi^1(x_2), \phi^2(x_1, x_2))$ where ϕ^2 is a symmetric function. Since we are interested in structures of varying sizes with relations of varying arities, we need to vary the arities of the functions in a realisation.

Fact 5. *There exists $\phi : X_\infty^t \rightarrow [0, 1]^{r([t])}$, a separable realisation of $\{l_A : A \in r([t])\}$ with levels ϕ^j for $j \in [t]$.*

- (Restriction) If $u \leq t$, then ϕ induces a realisation $\psi : X_\infty^u \rightarrow [0, 1]^{r([u])}$ defined by setting $\psi_{\{b_1, \dots, b_m\}}(x_1, \dots, x_u) = \phi^m(x_{b_1}, \dots, x_{b_m})$.
- (Lifting) If $u \geq t$, then ϕ induces a measure preserving map $\psi : X_\infty^u \rightarrow [0, 1]^{r([u], t)}$ defined by setting $\psi_{\{b_1, \dots, b_m\}}(x_1, \dots, x_u) = \phi^m(x_{b_1}, \dots, x_{b_m})$.

In a mild abuse of notation we will not distinguish carefully between the original realisation ϕ and its various restrictions and liftings.

3. LIMITS, REGULARITY AND REMOVAL FOR PURE RELATIONAL STRUCTURES

In this section we develop some technical machinery and then prove versions of the existence of limits, Strong Removal and Regularity for finite relational structures. Many of our arguments run parallel to those given by Elek and Szegedy [1] for the case of uniform hypergraphs.

In the first part of this section we define limit objects for convergent sequences (\mathcal{N}_k) of finite structures. We proceed as follows:

- Form the ultraproduct \mathcal{N}_∞ .
- Use separable realisations to associate to \mathcal{N}_∞ a family $\mathfrak{E} = (E_p^i : (i, p) \in \text{Index})$ of Lebesgue measurable sets where $E_p^i \subseteq [0, 1]^{r[\|p\|]}$.
- Associate to \mathfrak{E} a random sequence (\mathcal{N}_k) of finite structures, and prove that almost surely (\mathcal{N}_k) is convergent and $\lim_{k \rightarrow \infty} p(\mathcal{M}, \mathcal{N}_k) = \lim_{k \rightarrow \infty} p(\mathcal{M}, \mathcal{N}_k)$ for all finite structures \mathcal{M} .

In the remainder of the section we use the existence of limits to prove general versions of Strong Removal and Regularity.

3.1. Limits. We are given a convergent sequence (\mathcal{N}_k) . Recall that we defined a function Φ_p on the class of finite \mathcal{L} -structures by setting $\Phi_p(\mathcal{M}) = \lim_{k \rightarrow \infty} p(\mathcal{M}, \mathcal{N}_k)$.

We will begin by forming the ultraproduct $\mathcal{N}_\infty = \prod_k \mathcal{N}_k / U$. Let $X_k = |\mathcal{N}_k|$ and $X_\infty = \prod_k X_k / U$, so that \mathcal{N}_∞ is an \mathcal{L} -structure with underlying set X_∞ . The first key point is that the structure \mathcal{N}_∞ captures the function Φ_p .

Claim 1. *For every finite \mathcal{L} -structure \mathcal{M} , $\Phi_p(\mathcal{M})$ is the standard part of $p(\mathcal{M}^*, \mathcal{N}_\infty)$, and similar statements hold for the density functions t , t_0 and t_{ind} .*

Proof. The standard part of $p(\mathcal{M}^*, \mathcal{N}_\infty)$ is $\lim_{k \rightarrow U} p(\mathcal{M}, \mathcal{N}_k)$, that is the unique real number r such that for every $\epsilon > 0$, $\{k : p(\mathcal{M}, \mathcal{N}_k) \in (r - \epsilon, r + \epsilon)\} \in U$. Since $p(\mathcal{M}, \mathcal{N}_k) \rightarrow \Phi_p(\mathcal{M})$ we see that $r = \Phi_p(\mathcal{M})$. The proofs for the other density functions are exactly similar. \square

We now recall the coding scheme from subsection 2.2, in which an arbitrary \mathcal{L} -structure \mathcal{N} is coded by a family $\text{Code}(\mathcal{N}) = (\text{DH}_p^{\mathcal{N}, i} : (i, p) \in \text{Index})$. Since the coding scheme is defined in a uniform way, it is easy to see that the family of hypergraphs $\text{Code}(\mathcal{N}_\infty)$ is represented in the ultrapower by the sequence $(\text{Code}(\mathcal{N}_k) : k \in \mathbb{N})$. It is immediate from the definitions that if $p \in \text{Part}_{[r_i]}$ is a partition with $\|p\| = t$, then the relation $\text{DH}_p^{\mathcal{N}_\infty, i}$ is represented by $(\text{DH}_p^{\mathcal{N}_k, i} : k \in \mathbb{N})$, in particular it is a subset of X_∞^t which lies in the σ -algebra $\sigma_{[t]}$.

Let $r_{\max} = \max_i r_i$. Appealing to Facts 4 and 5 we may find a separable system $\{l_A : A \in r([r_{\max}])\}$ which is suitable for all the sets appearing in $\text{Code}(\mathcal{N}_\infty)$, and then a realisation $\phi : X_\infty^{r_{\max}} \rightarrow [0, 1]^{r([r_{\max}])}$ for this system. In line with the discussion from section 2.5, for all $t \leq r_{\max}$ we obtain restricted separable realisations $\phi : X_\infty^t \rightarrow [0, 1]^{r([t])}$.

We associate a measurable set in a suitable power of the unit interval to each set of the form $\text{DH}_p^{\mathcal{N}_\infty, i}$. Let $\|p\| = t$, and choose $E_p^i \subseteq [0, 1]^{r([t])}$ such that $\text{DH}_p^{\mathcal{N}_\infty, i} \Delta \phi^{-1}[E_p^i]$ is a null set for the measure $\mu_{[t]}$. We note that the definition of E_p^i involves using restrictions of the original map ϕ , and is legitimate because $\|p\| \leq r_i \leq r_{\max}$ for all $(i, p) \in \text{Index}$.

Let $\mathfrak{E} = (E_p^i : (i, p) \in \text{Index})$; systems of measurable sets of this form will serve as limit objects for convergent sequences of \mathcal{L} -structures.

Definition 5. *An \mathcal{L} -limit is a family of sets $\mathfrak{F} = (F_p^i : (i, p) \in \text{Index})$ such that F_p^i is a measurable subset of $[0, 1]^{r(\|p\|)}$ for each p and i .*

Let \mathfrak{F} be an \mathcal{L} -limit. We will view \mathfrak{F} as a template for the construction, for each integer m , of a random \mathcal{L} -structure $\mathcal{N}(\mathfrak{F}, m)$ with underlying set $[m]$. To define this random structure we will choose uniformly at random a tuple of real numbers $\vec{y} \in [0, 1]^{r([m], r_{\max})}$, and compute from this tuple a structure $\mathcal{N}(\mathfrak{F}, m, \vec{y})$ with underlying set $[m]$.

To define the structure $\mathcal{N}(\mathfrak{F}, m, \vec{y})$, we describe its coding by directed hypergraphs. For each $(i, p) \in \text{Index}$ and each t -tuple $(b_1, \dots, b_t) \in [m]^t$, let $\text{DH}_p^{\mathcal{N}(\mathfrak{F}, m, \vec{y}), i}(b_1, \dots, b_t)$ if and only if $\vec{z} \in F_p^i$, where $\vec{z} = (z_B : B \in r([t]))$ is given by $z_B = y_{\{b_l : l \in B\}}$.

For the record we also give an uncoded version. Let $\vec{a} = (a_1, \dots, a_{r_i})$ be an r -tuple of elements of $[m]$. Let $p = p(\vec{a}) \in \text{Part}_{[r_i]}$ be the induced partition of $[r_i]$, in which j and j' are in the same class if and only if $a_j = a_{j'}$, and let $t = \|p\|$.

Let $\vec{b} = C(\vec{a})$ be the t -tuple obtained from \vec{a} by deleting repetitions, so that $\vec{b} \in [m]^t$, and let $\vec{z} = (z_B : B \in r([t]))$ be given by $z_B = x_{\{b_l : l \in B\}}$. Then $R_i^{\mathcal{N}(\mathfrak{F}, m, \vec{y}), i}(a_1, \dots, a_{r_i})$ if and only if $\vec{z} \in F_p^i$.

We will now form the random sequence $(\mathcal{N}(\mathfrak{E}, j^2) : j \in \mathbb{N})$, which (formally speaking) is a random variable defined on the product over $j \in \mathbb{N}$ of the probability spaces associated with the random structures $\mathcal{N}(\mathfrak{E}, j^2)$. We will prove that almost surely the random sequence $(\mathcal{N}(\mathfrak{E}, j^2))$ is convergent and resembles the original sequence (\mathcal{N}_k) in a certain precise sense. Our proof uses ideas from work by Razborov [8]. Razborov associates to each convergent sequence of finite \mathcal{L} -structures²

²Actually Razborov's theory is more general, and applies to the class of models of a universal theory containing a labelled copy of a given finite structure.

a limit object of a rather different kind, namely a homomorphism from a certain \mathbb{R} -algebra (which depends only on \mathcal{L}) to \mathbb{R} . The following fact appears (in a more general form) as Theorem 3.18 in [8].

Fact 6. *Let $(\mathcal{N}_k : k \in \mathbb{N})$ be a convergent sequence of finite \mathcal{L} -structures. For each j , make the set of isomorphism classes of structures \mathcal{M} with $\|\mathcal{M}\| = j^2$ into a finite probability space in which each \mathcal{M} is assigned probability $\Phi_p(\mathcal{M})$. Let $(\mathcal{N}_j^* : j \in \mathbb{N})$ be the associated random sequence of structures (that is it's the random variable on the product of the finite probability spaces given by the identity function on this product). Then almost surely $(\mathcal{N}_j^* : j \in \mathbb{N})$ is convergent and $\lim_{j \rightarrow \infty} p(\mathcal{M}, \mathcal{N}_j^*) = \Phi_p(\mathcal{M})$ for all finite \mathcal{L} -structures \mathcal{M} .*

Theorem 1. *Almost surely $(\mathcal{N}(\mathfrak{E}, j^2))$ is a convergent sequence and*

$$\lim_{j \rightarrow \infty} p(\mathcal{M}, \mathcal{N}(\mathfrak{E}, j^2)) = \Phi_p(\mathcal{M})$$

for all finite structures \mathcal{M} .

Before starting the proof, we define some auxiliary sets relating to the density function t_{ind} .

Definition 6. *Let \mathcal{M} be a finite \mathcal{L} -structure with underlying set X , enumerated as x_1, \dots, x_m .*

- (1) *Let \mathcal{N} be an arbitrary \mathcal{L} -structure. Then $T_{\text{ind}}(\mathcal{M}, \mathcal{N})$ is the set of m -tuples $(x'_1, \dots, x'_m) \in |\mathcal{N}|^k$ such that the map which takes x_i to x'_i for each i is an embedding from \mathcal{M} to \mathcal{N} .*
- (2) *Let \mathfrak{F} be an \mathcal{L} -limit. Then $T_{\text{ind}}(\mathcal{M}, \mathfrak{F})$ is the set of tuples $\vec{y} \in [0, 1]^{r(m), r_{\text{max}}}$ such that the map which takes x_i to i for each i is an embedding from \mathcal{M} to $\mathcal{N}(\mathfrak{F}, m, \vec{y})$.*

Of course the definitions of these sets depend on the choice of an enumeration for $|\mathcal{M}|$, but this is irrelevant in applications. We note that when \mathcal{N} is finite, $t_{\text{ind}}(\mathcal{M}, \mathcal{N})$ is just the normalised counting measure of $T_{\text{ind}}(\mathcal{M}, \mathcal{N})$.

Proof. By Fact 6, it will suffice to show that for each m and each structure \mathcal{M} with $\|\mathcal{M}\| = m$, $P(\mathcal{N}(\mathfrak{E}, m) \simeq \mathcal{M}) = \Phi_p(\mathcal{M})$. We will actually show that the probability p that a random bijection from $|\mathcal{M}|$ to $|\mathcal{N}(\mathfrak{E}, m)|$ is an isomorphism is $\Phi_{t_{\text{ind}}}(\mathcal{M})$, from which the desired conclusion follows immediately using Fact 1. To show that $p = \Phi_{t_{\text{ind}}}(\mathcal{M})$, we will define measurable sets F_1 and F_2 such that the measure of F_1 is $\Phi_{t_{\text{ind}}}(\mathcal{M})$ and the measure of F_2 is p , and then use the ideas of subsection 2.5 to show that F_1 and F_2 have the same measure.

Let $F_1 = T_{\text{ind}}(\mathcal{M}, \mathcal{N}_\infty)$; by Claim 1 and the definition of $\mu_{[m]}$ as the standard part of the normalised counting measure on X_∞^m , it is immediate that $\Phi_{t_{\text{ind}}}(\mathcal{M}) = \mu_{[m]}(F_1)$.

Let $F_2 = T_{\text{ind}}(\mathcal{M}, \mathfrak{E})$; it is easy to see that the probability p that a bijection between $|\mathcal{M}|$ and $|\mathcal{N}(\mathfrak{E}, m)|$ selected uniformly at random is an isomorphism is $\lambda(F_2)$, where λ is Lebesgue measure on $[0, 1]^{r([m], r_{\text{max}})}$.

Let $|\mathcal{M}| = \{x_1, \dots, x_m\}$. By Fact 5, the separable system $\{l_A : A \in r([r_{\text{max}}])\}$ and its realisation $\phi : X_\infty^{r_{\text{max}}} \rightarrow [0, 1]^{r([r_{\text{max}}])}$ can be lifted, to obtain a family of algebras $\{l_A : A \in r([m], r_{\text{max}})\}$ and a map $\phi : X_\infty^m \rightarrow [0, 1]^{r([m], r_{\text{max}})}$.

By definition, F_1 is the set of tuples $(x'_1, \dots, x'_m) \in X_\infty^m$ such that for every $i \in [n]$, and every r_i -tuple (a_1, \dots, a_{r_i}) of elements of $[m]$, $R_i^{\mathcal{M}}(x_{a_1}, \dots, x_{a_{r_i}})$ if and only if $R_i^{\mathcal{N}_\infty}(x'_{a_1}, \dots, x'_{a_{r_i}})$. Recalling how we coded relations into families of directed hypergraphs, it is routine to check that F_1 is the set of tuples $(x'_1, \dots, x'_m) \in X_\infty^m$ such that for every $(i, p) \in \text{Index}$ (say with $\|p\| = t$) and every t -tuple $(b_1, \dots, b_t) \in [m]^t$, $\text{DH}_p^{\mathcal{M}, i}(x_{b_1}, \dots, x_{b_t})$ if and only if $\text{DH}_p^{\mathcal{N}_\infty, i}(x'_{b_1}, \dots, x'_{b_t})$.

By definition, F_2 is the set of tuples \vec{y} of reals such that for every $i \in [n]$, and every r_i -tuple (a_1, \dots, a_{r_i}) of elements of $[m]$, $R_i^{\mathcal{M}}(x_{a_1}, \dots, x_{a_{r_i}})$ if and only if $R_i^{\mathcal{N}(\mathfrak{E}, m, \vec{y})}(a_1, \dots, a_{r_i})$. In terms of the coding by directed hypergraphs, this condition states that for every $(i, p) \in \text{Index}$ with $\|p\| = t$ say, and every t -tuple $(b_1, \dots, b_t) \in [m]^t$, $\text{DH}_p^{\mathcal{M}, i}(x_{b_1}, \dots, x_{b_t})$ if and only if $\text{DH}_p^{\mathcal{N}(\mathfrak{E}, m, \vec{y}), i}(b_1, \dots, b_t)$.

By the definition of the structure $\mathcal{N}(\mathfrak{E}, m, \vec{y})$, $\text{DH}_p^{\mathcal{N}(\mathfrak{E}, m, \vec{y}), i}(b_1, \dots, b_t)$ if and only if $\vec{z} \in E_p^i$, where $\vec{z} = (z_B : B \in r([t]))$ is given by $z_B = y_{\{b_l : l \in B\}}$. Recalling that we chose the sets E_p^i such that $\text{DH}_p^{\mathcal{N}_\infty, i}$ and $\phi^{-1}[E_p^i]$ differ by a null set, routine calculation now shows that F_1 and $\phi^{-1}[F_2]$ differ by a null set, so that (since the lifted version of ϕ is measure preserving) $\Phi_{t_{\text{ind}}}(\mathcal{M}) = \mu_{[k]}(F_1) = \lambda(F_2) = p$ as required. \square

We make a few concluding remarks:

- Adapting proofs by Lovász and Szegedy [5] for graph limits, one can show that in fact for any \mathcal{L} -limit \mathfrak{F} the sequence $(\mathcal{N}(\mathfrak{F}, j^2))$ is convergent.
- From the results above, the reader who is familiar with flag algebras will see that our notion of \mathcal{L} -limit and Razborov's notion of flag algebra homomorphism (in the special case of the zero flag and the empty theory) are intertranslatable. One could see them as two ways of describing a notion of “random sequence of \mathcal{L} -structures”.

- It is easy to see that any quantity of the form $\Phi_{p/t/t_{\text{ind}}}(\mathcal{M})$ can be expressed as a suitable integral. In general these integral expressions are quite complicated, we will give a few simple examples in the Appendix.

3.2. Hyperpartitions. We sketch some ideas of Elek and Szegedy from their work [1] on hypergraph limits. To motivate the definitions in this section, we consider a separable realisation $\phi : X_\infty^t \rightarrow [0, 1]^{r([t])}$. Given $l > 0$, we divide the unit interval into l subintervals of equal measure by defining $I_l^a = [\frac{a-1}{l}, \frac{a}{l}]$ for $1 \leq a < l$, $I_l^l = [\frac{l-1}{l}, 1]$.

Given a tuple $\vec{e} = (e_i : i \in I) \in [l]^I$ for some index set I , we let $\text{Cube}(\vec{e})$ be the ‘‘hypercube’’ $\prod_{i \in I} I_l^{e_i} \subseteq [0, 1]^I$. One key point here is that any measurable subset of $[0, 1]^I$ can be approximated arbitrarily well by sets which are unions of hypercubes: this plays an important role in the proofs of Removal and Regularity. Of course formally the definition of $\text{Cube}(\vec{e})$ depends on the value l , but in practice this should always be clear from the context. For every tuple $\vec{x} \in X_\infty^t$ there is a unique tuple $e(\vec{x}) \in [l]^{r([t])}$ such that $\phi(\vec{x}) \in \text{Cube}(e(\vec{x}))$. In particular, there is a natural equivalence relation according to which tuples $\vec{x}, \vec{x}' \in X_\infty^t$ are equivalent if and only if $e(\vec{x}) = e(\vec{x}')$.

Recall now the description of ϕ in terms of the ‘‘levels’’ ϕ^j from subsection 2.5. For each j and e , let $H_e^j = \{\{x_1, \dots, x_j\} \in X_\infty^j : \phi^j(x_1, \dots, x_j) \in I_l^e\}$, where we note that H_e^j is a family of sets rather than tuples, and the definition makes sense because ϕ^j is invariant under permutation of its arguments.

It is now immediate from the definitions that:

- $e(\vec{x})_A$ is the unique $e \in [l]$ such that $\{x_j : j \in A\} \in H_e^{|A|}$.
- If \vec{x} and \vec{x}' are t -tuples in X_∞^t , \vec{x} and \vec{x}' are equivalent (in the sense that $e(\vec{x}) = e(\vec{x}')$) if and only if for all $A \in r([t])$, $\{x_j : j \in A\}$ and $\{x'_j : j \in A\}$ lie in the same cell of the partition of $X_\infty^{|A|}$ given by $\{H_e^{|A|} : e \in [l]\}$.

These considerations motivate the abstract definition of a *hyperpartition*, and the related notions of *equitability* and δ -*equitability*.

Definition 7. *Let X be a set.*

- A (t, l) -hyperpartition for X is a family $\mathcal{H} = (H_e^j : j \in [t], e \in [l])$ such that for each $j \in [t]$, $(H_e^j : e \in [l])$ forms a partition of X^j .
- Given such a hyperpartition, we let $h_e^j = \{(x_1, \dots, x_j) : \{x_1, \dots, x_j\} \in H_e^j\} \in H_e^j$.
- If X is equipped with a measure ν and for each j and e the set h_e^j is ν^j -measurable (where ν^j is product measure on X^j), then

\mathcal{H} is equitable if and only if $\nu^j(h_e^j) = \nu^j(h_{e'}^j)$ for all e, e' , and δ -equitable if and only if $|\nu^j(h_e^j) - \nu^j(h_{e'}^j)| < \delta$ for all e, e' .

It should be clear that the family of partitions defined above from a separable realisation ϕ constitutes an equitable hyperpartition for X_∞ . In what follows we will denote this hyperpartition by $\mathcal{H}_\infty^\phi = (H_{\infty,e}^{\phi,j} : j \in [t], e \in [l])$.

Let \mathcal{H} be a (t, l) -hyperpartition for some set X , let $u \leq t$ and let $\vec{x} = (x_1, \dots, x_u) \in X^u$. Then we define $e^{\mathcal{H}}(\vec{x}) \in [l]^{r([u])}$ by setting $e^{\mathcal{H}}(\vec{x})_A = b$ for the unique b such that $\{x_i : i \in A\} \in H_b^u$. We note that $(H_a^j : j \in [u], a \in [l])$ is a (u, l) -hyperpartition for X , and it suffices to determine the values of $e^{\mathcal{H}}$ on u -tuples; in the case of the hyperpartition \mathcal{H}_∞^ϕ , this (u, l) hyperpartition coincides with the one defined from the lowered version of ϕ . Given $\vec{e} \in [l]^{r([u])}$, we further define $\text{Cell}^{\mathcal{H}}(\vec{e})$ to be the set of u -tuples \vec{x} such that $e^{\mathcal{H}}(\vec{x}) = \vec{e}$. Note that by definition, $\text{Cell}^{\mathcal{H}_\infty^\phi}(\vec{e}) = \phi^{-1}[\text{Cube}(\vec{e})]$, and that $\text{Cell}^{\mathcal{H}}(\vec{e})$ is a u -uniform directed hypergraph.

For use in the proofs of Removal and Regularity, we elaborate our discussion of \mathcal{H}_∞^ϕ by relating it to the ultrapower construction. Each of the sets $h_{\infty,a}^j$ is measurable and so it differs on a set of measure zero from a set in V^* ; we may therefore choose a new equitable hyperpartition $\mathcal{H}_\infty \in V^*$ such that $h_{\infty,e}^j \Delta h_{\infty,e}^{\phi,j}$ is null for all e and j . We now choose (\mathcal{H}_k) so that $\mathcal{H}_\infty = [\mathcal{H}_k]$, and \mathcal{H}_k is a (t, l) -hyperpartition on $X_k = |\mathcal{N}_k|$ for each k . For use in the proof of Regularity, we note that we may find numbers δ_k such that \mathcal{H}_k is δ_k -equitable and $\lim_{k \rightarrow U} \delta_k = 0$. To lighten the notation we write e_k for the function $e^{\mathcal{H}_k}$, and similarly Cell_k for $\text{Cell}^{\mathcal{H}_k}$. It is routine to verify that $[\text{Cell}_k(\vec{e})] = \text{Cell}^{\mathcal{H}_\infty}(\vec{e})$, which differs on a null set from $\text{Cell}^{\mathcal{H}_\infty^\phi}(\vec{e}) = \phi^{-1}[\text{Cube}(\vec{e})]$.

3.3. Removal. Given a finite set X , we introduce a metric d on the set of \mathcal{L} -structures with underlying set X . The key point is that $d(\mathcal{M}, \mathcal{N})$ measures the difference between the edge sets of the various directed hypergraphs comprising $\text{Code}(\mathcal{M})$ and $\text{Code}(\mathcal{N})$. We start by defining $d_p^i(\mathcal{M}, \mathcal{N})$ to be $|\text{DH}_p^i(\mathcal{M}) \Delta \text{DH}_p^i(\mathcal{N})| / |X|^{\|p\|}$ for all $(i, p) \in \text{Index}$, then let $d(\mathcal{M}, \mathcal{N}) = \sum_{(i,p) \in \text{Index}} d_p^i(\mathcal{M}, \mathcal{N})$. We can now state and prove a version of the Strong Removal Theorem.

Theorem 2. *Let \mathcal{F} be a (possibly infinite) set of finite \mathcal{L} -structures. For every $\epsilon > 0$ there exist $\delta > 0$ and m such that for all sufficiently large finite \mathcal{L} -structures \mathcal{N} , if $p(\mathcal{M}, \mathcal{N}) < \delta$ for all $\mathcal{M} \in \mathcal{F}$ with $\|\mathcal{M}\| \leq m$, then there is \mathcal{N}' such that $|\mathcal{N}| = |\mathcal{N}'|$, $d(\mathcal{N}, \mathcal{N}') < \epsilon$ and $p(\mathcal{M}, \mathcal{N}') = 0$ for all $\mathcal{M} \in \mathcal{F}$.*

Proof. If not, then we fix an $\epsilon > 0$ for which the claim of the theorem fails. We use this to construct an increasing sequence (\mathcal{N}_k) of finite structures such that for all k :

- $p(\mathcal{M}, \mathcal{N}_k) < 1/k$ for all $\mathcal{M} \in \mathcal{F}$ with $\|\mathcal{M}\| \leq k$.
- There is no \mathcal{N}' such that $|\mathcal{N}_k| = |\mathcal{N}'|$, $d(\mathcal{N}_k, \mathcal{N}') < \epsilon$ and $p(\mathcal{M}, \mathcal{N}') = 0$ for all $\mathcal{M} \in \mathcal{F}$.

The strategy of the proof is as follows: we will produce for some large k , structures \mathcal{N}_k^* and \mathcal{N}_k^\dagger such that $d(\mathcal{N}_k, \mathcal{N}_k^*) < \epsilon/2$, $d(\mathcal{N}_k^*, \mathcal{N}_k^\dagger) < \epsilon/2$, and $p(\mathcal{M}, \mathcal{N}_k^\dagger) = 0$ for all $\mathcal{M} \in \mathcal{F}$. This will give an immediate contradiction. Thinning it out if necessary, we may assume that the sequence (\mathcal{N}_k) is convergent and may form the associated function Φ_p . By construction, $\Phi_p(\mathcal{M}) = 0$ for all $\mathcal{M} \in F$. We now form the ultrapower \mathcal{N}_∞ , and define a separable realisation ϕ and \mathcal{L} -limit \mathfrak{E} as in subsection 3.1. Since $\Phi_p(\mathcal{M}) = 0$ for all $\mathcal{M} \in F$, it follows (as in the proof of Theorem 1) that $T_{\text{ind}}(\mathcal{M}, \mathfrak{E})$ is a null set for all $\mathcal{M} \in F$. We recall that \mathfrak{E} is a family of Lebesgue measurable sets E_p^i , where p is a partition of $[r_i]$ into some number $t = \|p\|$ of pieces and $E_p^i \subseteq [0, 1]^t$. We choose $\epsilon' > 0$ small enough that $|\text{Index}| \epsilon' < \epsilon/2$, and then choose an integer l so large that every set E_p^i can be approximated with error ϵ' by a union of hypercubes which are products of intervals of the form I_l^e : that is for every i and p there is a set $X_p^i \subseteq [l]^{r(\|p\|)}$ such that if we set $E_p^{*,i} = \bigcup_{\vec{e} \in X_p^i} \text{Cube}(\vec{e})$ then $\lambda(E_p^i \Delta E_p^{*,i}) < \epsilon'$.

With these definitions in place, we can now define (r_{\max}, l) -hyperpartitions \mathcal{H}_∞^ϕ , \mathcal{H}_∞ and \mathcal{H}_k exactly as in subsection 3.2. Let $l_k = \|\mathcal{N}_k\|$, so that we may assume $|\mathcal{N}_k| = [l_k]$. We use the hyperpartition \mathcal{H}_k and the sets X_p^i to define a structure \mathcal{N}_k^* with $|\mathcal{N}_k^*| = |\mathcal{N}_k|$.

For each $i \in [n]$ and each partition p of $[r_i]$, we use \mathcal{H}_k to partition $\|\mathcal{N}_k\|^{\|p\|}$ into cells of the form $\text{Cell}_k(\vec{e})$ for $e \in [l]^{r[t]}$. Then we form sets $\text{DH}_{k,p}^{*,i} = \bigcup_{\vec{e} \in X_p^i} \text{Cell}_k(\vec{e})$, and finally we decode the resulting system of directed hypergraphs by setting $\mathcal{N}_k^* = \text{Decode}(\text{DH}_{k,p}^{*,i} : (i, p) \in \text{Index})$.

By a routine application of Loś' theorem, the ultrapower \mathcal{N}_∞^* is the result of decoding the hypergraphs obtained in a similar way from \mathcal{H}_∞ . Chasing through the definitions, this means that for each i and p , the directed hypergraph $\text{DH}_p^{\mathcal{N}_\infty^*,i}$ is $\bigcup_{\vec{e} \in X_p^i} \text{Cell}^{\mathcal{H}_\infty}(\vec{e})$. Recalling that ϕ is measure preserving, and $\lambda(E_p^i \Delta E_p^{*,i}) < \epsilon'$, we see that $\mu(\phi^{-1}[E_p^i] \Delta \phi^{-1}[E_p^{*,i}]) < \epsilon'$. Since $\phi^{-1}(E_p^i) \Delta \text{DH}_p^{\mathcal{N}_\infty^*,i}$ is null, $\phi^{-1}[E_p^{*,i}] = \bigcup_{\vec{e} \in X_p^i} \text{Cell}^{\mathcal{H}_\infty^\phi}(\vec{e})$, and $\text{Cell}^{\mathcal{H}_\infty}(\vec{e}) \Delta \text{Cell}^{\mathcal{H}_\infty^\phi}(\vec{e})$ is null, we see that

$$\mu(\text{DH}_p^{\mathcal{N}_\infty^*,i} \Delta \text{DH}_p^{\mathcal{N}_\infty^*,i}) < \epsilon'$$

for all i and p , so that easily $d(\mathcal{N}_k^*, \mathcal{N}_k) < \epsilon/2$ for U -almost every k .

Both \mathfrak{E} and \mathfrak{E}^* are \mathcal{L} -limits, but (following an idea of Elek and Szegedy which they dub ‘‘hyperpartition sampling’’) we will use them in a slightly different way from that described in subsection 3.1 to define random structures \mathcal{N}_k and \mathcal{N}_k^* . Before making the definitions we warn the reader that \mathcal{N}_k^* is a rather trivial random structure in that (although we will shortly give a more complicated definition) it is always equal to \mathcal{N}_k^* ; the point in defining \mathcal{N}_k^* in this seemingly obtuse way is to facilitate comparison with \mathcal{N}_k . We define \mathcal{N}_k using the same function $\vec{y} \mapsto \mathcal{N}(\mathfrak{E}, l_k, \vec{y})$ that was used in the definition of $\mathcal{N}(\mathfrak{E}, l_k)$ from subsection 3.1; the key difference is that we choose values of \vec{y} uniformly from the hypercube $\text{Cube}(\vec{c}) \subseteq [0, 1]^{r([l_k], r_{\max})}$, where $\vec{c} = (c_A : A \in r([l_k], r_{\max}))$ and c_A is defined to be the unique $j \in [l]$ such that $A \in H_{k,j}^{|A|}$. Similarly we define \mathcal{N}_k^* to be the value of $\mathcal{N}(\mathfrak{E}^*, l_k, \vec{y})$ where again \vec{y} is chosen uniformly at random from $\text{Cube}(\vec{c})$. Since $T_{\text{ind}}(\mathcal{M}, \mathfrak{E})$ is a null set for all $\mathcal{M} \in \mathcal{F}$ and \mathcal{F} is countable, it is easy to see that almost surely $p(\mathcal{M}, \mathcal{N}_k) = 0$ for all $\mathcal{M} \in \mathcal{F}$. We claim that $\mathcal{N}(\mathfrak{E}^*, l_k, \vec{y}) = \mathcal{N}_k^*$ for all $\vec{y} \in \text{Cube}(\vec{c})$, which we will verify by inspecting the computation of the associated directed hypergraphs. Let us fix $\vec{y} \in \text{Cube}(\vec{c})$, $(i, p) \in \text{Index}$ with $\|p\| = t$ say, and $(a_1, \dots, a_t) \in [l_k]^t$: then by definition $(a_1, \dots, a_t) \in \text{DH}_p^{\mathcal{N}(\mathfrak{E}^*, l_k, \vec{y}), i}$ if and only if $(y_{\{a_i : i \in B\}} : B \in r([t])) \in E_i^{*,p}$ if and only if $e_k(a_1, \dots, a_t) = (c_{\{a_i : i \in B\}} : B \in r([t])) \in X_i^p$ if and only if $(a_1, \dots, a_t) \in \text{DH}_p^{\mathcal{N}_k^*, i}$. Next we claim that $\lim_{k \rightarrow U} \mathbb{E}(d(\mathcal{N}_k, \mathcal{N}_k^*)) < \epsilon/2$. Fix k , and $(i, p) \in \text{Index}$ and let $t = \|p\|$. For each $\vec{a} = (a_1, \dots, a_t) \in [l_k]^t$, let $\mathbf{Y}(\vec{a})$ be the indicator function of the event ‘‘ $\vec{a} \in \text{DH}_p^{\mathcal{N}_k, i} \Delta \text{DH}_p^{\mathcal{N}_k^*, i}$ ’’. The expected value of $\mathbf{Y}(\vec{a})$ is the probability that the tuple $(y_{\{a_i : i \in B\}} : B \in r([t]))$ lies in $E_i^p \Delta E_i^{*,p}$; since $y_{\{a_i : i \in B\}}$ is chosen uniformly from the interval $I_l^{y_{\{a_i : i \in B\}}}$, we see easily that

$$\mathbb{E}(\mathbf{Y}(\vec{a})) = l^{2t-1} \lambda(E_i^p \Delta E_i^{*,p} \cap B(\vec{a})),$$

where $B(\vec{a}) = \text{Cube}(c_{\{a_i : i \in A\}} : A \in r([t]))$. So

$$\begin{aligned} \mathbb{E}(|\text{DH}_p^{\mathcal{N}_k, i} \Delta \text{DH}_p^{\mathcal{N}_k^*, i}| / l_k^{2t-1}) &= \sum_{\vec{a} \in [l_k]^t} \lambda(E_i^p \Delta E_i^{*,p} \cap B(\vec{a})) \\ &= \sum_{\vec{c} \in [l]^{r([t])}} |\text{Cell}_k(\vec{c})| (l/l_k)^{2t-1} \lambda(E_i^p \Delta E_i^{*,p} \cap \text{Cube}(\vec{c})). \end{aligned}$$

Since the sequence $(\text{Cell}_k(\vec{e}))$ represents a set of measure l^{1-2^t} , it follows that $\lim_{k \rightarrow U} |\text{Cell}_k(\vec{e})| (l/l_k)^{2^t-1} = 1$. Hence

$$\begin{aligned} \lim_{k \rightarrow U} \mathbb{E}(|\text{DH}_p^{\mathcal{N}_{k,i}} \Delta \text{DH}_p^{\mathcal{N}_{k,i}^*}| / l_k^{2^t-1}) &= \sum_{\vec{b} \in [l]^{r[t]}} \lambda(E_i^p \Delta E_i^{*,p} \cap \text{Cube}(\vec{b})) \\ &= \lambda(E_i^p \Delta E_i^{*,p}) < \epsilon'. \end{aligned}$$

By the choice of ϵ' , it follows that $\lim_{k \rightarrow U} \mathbb{E}(d(\mathcal{N}_k, \mathcal{N}_k^*)) < \epsilon/2$.

Since almost surely $p(M, \mathcal{N}_k) = 0$ for all $M \in F$ and all k , it follows from the claim of the last paragraph that for U -many k there is a structure \mathcal{N}_k^\dagger such that $d(\mathcal{N}_k^\dagger, \mathcal{N}_k^*) < \epsilon/2$ and $p(M, \mathcal{N}_k^\dagger) = 0$ for all $M \in F$. Since $d(\mathcal{N}_k^*, \mathcal{N}_k) < \epsilon/2$ for U -many k , and d is a metric, this implies that there is some k such that $d(\mathcal{N}_k^\dagger, \mathcal{N}_k) < \epsilon$ and $p(M, \mathcal{N}_k^\dagger) = 0$ for all $M \in F$. This is a contradiction. \square

It is easy to see that Theorem 2 implies the form of removal with one omitted structure and homomorphic images:

Theorem 3. *Let \mathcal{M} be a finite \mathcal{L} -structure. For every $\epsilon > 0$ there exists $\delta > 0$ such that for all large \mathcal{L} -structures \mathcal{N} , if $t(\mathcal{M}, \mathcal{N}) < \delta$ then there is \mathcal{N}' such that $|\mathcal{N}| = |\mathcal{N}'|$, $d(\mathcal{N}, \mathcal{N}') < \epsilon$ and $p(\mathcal{M}, \mathcal{N}') = 0$.*

Theorem 2 can also be applied to prove testability for hereditary properties of structures, using standard arguments.

3.4. Regularity.

Definition 8. *A cylindric intersection set L in X^A is a set $L = \bigcap_{\{B: B \subseteq A\}} \pi_B^{-1}(Y_B)$, where each $Y_B \subseteq X^B$ is measurable.*

If L is a cylindric intersection in the ultraproduct $X_\infty^{[r]}$ then $L \in \sigma_{[r]}^*$.

In the case that $r = 2$, every cylindric intersection in $X^{[2]}$ is a measurable rectangle, whether X is a finite set or the ultraproduct.

Definition 9. *A directed r -uniform hypergraph H on X is ϵ -regular if for any cylindric intersection set L in $X^{[r]}$ with $\mu(L) \geq \epsilon$, $|\mu(H \cap L) - \mu(H)\mu(L)| < \epsilon$.*

If H is ϵ -regular for all $\epsilon > 0$, then $\mu(H \cap L) = \mu(H)\mu(L)$ for any cylindric intersection L .

We can now state and prove a version of the Regularity Lemma.

Theorem 4. *Given $\epsilon > 0$ and $d \in \mathbb{N}$, there exist $D, N \in \mathbb{N}$ such that for any \mathcal{L} -structure \mathcal{M} with $\|\mathcal{M}\| \geq N$, there exists an ϵ -equitable (r_{\max}, l) -hyperpartition \mathcal{H} on $|\mathcal{M}|$ such that*

- $d \leq l \leq D$
- Each h_e^j is ϵ -regular.
- For each $(i, p) \in \text{Index}$, there exist m and tuples $\vec{e}_1, \dots, \vec{e}_m$ such that $\mu(\text{DH}_p^{\mathcal{M}, i} \Delta \bigcup_{i=1}^m \text{Cell}^{\mathcal{H}}(\vec{e}_i)) < \epsilon$.

Proof. If not, then we fix an $\epsilon > 0$ for which the claim of the theorem fails. We use this to construct an increasing sequence (\mathcal{N}_k) of finite structures such that for all k , there is no ϵ -equitable (r_{\max}, l) -hyperpartition, for $d \leq \ell \leq k$ satisfying above conditions for any \mathcal{N}_k .

We now form the ultrapower \mathcal{N}_∞ and define (r_{\max}, l) -hyperpartitions $\mathcal{H}_\infty^\phi, \mathcal{H}_\infty$ and \mathcal{H}_k as in the proof of the Removal Lemma.

Each $h_{\infty, e}^j$ is independent of every set $L \in \sigma_{[j]}^*$, that is, $\mu(h_{\infty, e}^j \cap L) = \mu(h_{\infty, e}^j) \mu(L)$. Then \mathcal{H}_∞^ϕ and \mathcal{H}_∞ are ϵ -regular as well as equitable for all $\epsilon > 0$. Also \mathcal{H}_∞ has the property that for each $(i, p) \in \text{Index}$, $\text{DH}_p^{\mathcal{N}, i} = \bigcup_{i=1}^m \text{Cell}^{\mathcal{H}_\infty}(\vec{e}_i)$ for some m and tuples $\vec{e}_1, \dots, \vec{e}_m$. For U -almost every k , \mathcal{H}_k is an ϵ -equitable (r_{\max}, l) -hyperpartition and for each $(i, p) \in \text{Index}$, $\text{DH}_p^{\mathcal{N}_k, i} = \bigcup_{i=1}^m \text{Cell}_k(\vec{e}_i)$ for some m and tuples $\vec{e}_1, \dots, \vec{e}_m$. By our assumption there must exist e and j such that for U -almost every k , $h_{e, k}^j$ is not ϵ -regular. Let (L_k) be the sequence of cylindrical intersection sets that witness this fact. Let $L = [L_k] \in \sigma_{[j]} C^*$. Our assumption implies $h_{e, \infty}^j$ and L are not independent, which is a contradiction. \square

4. MODELS OF A GENERAL UNIVERSAL THEORY

Let T be a universal theory in the language \mathcal{L} . We need a technical result describing the models of T .

Lemma 1. *There is a set Forbid^T of isomorphism types of finite \mathcal{L} -structures such that for any \mathcal{L} -structure \mathcal{M} , $\mathcal{M} \models T$ if and only if no finite induced substructure of \mathcal{M} has isomorphism type in Forbid^T .*

Since the proof is easy and tedious, we have relegated it to the appendix. It follows immediately that the limit of a convergent sequence of models of T describes a random model of T . To be more precise we have :

Theorem 5. *Let (\mathcal{N}_k) be a convergent sequence of finite models of T , and let \mathfrak{E} be the limit as constructed in subsection 3.1. Then almost surely $\mathcal{N}(\mathfrak{E}, m)$ is a model of T for all m .*

Proof. For each type \mathcal{M} in Forbid^T , $p(\mathcal{M}, \mathcal{N}_\infty) = 0$ and so $T_{\text{ind}}(\mathcal{M}, \mathfrak{E})$ is a null set. It follows that almost surely $p(\mathcal{M}, \mathcal{N}(\mathfrak{E}, m)) = 0$ for each \mathcal{M} and m , so we are done by σ -additivity. \square

The integrals representing $\Phi_{p/t/t_{\text{ind}}}(\mathcal{M})$ in terms of the limit object \mathfrak{E} sometimes become simpler when \mathfrak{E} is a limit of models of T . Some examples appear in Section A of the Appendix.

Theorem 6. *Let T be a universal theory in the finite language \mathcal{L} .*

Let F be a (possibly infinite) set of finite models of T . For every $\epsilon > 0$ there exist $\delta > 0$ and m such that for all sufficiently large finite models \mathcal{N} of T , if $p(\mathcal{M}, \mathcal{N}) < \delta$ for all $\mathcal{M} \in F$ with $\|\mathcal{M}\| \leq m$, then there is a model \mathcal{N}' of T such that $|\mathcal{N}| = |\mathcal{N}'|$, $d(\mathcal{N}, \mathcal{N}') < \epsilon$ and $p(\mathcal{M}, \mathcal{N}') = 0$ for all $\mathcal{M} \in F$.

Proof. Let $F^+ = F \cup \text{Forbid}^T$ and apply Theorem 2 to F^+ . \square

Note that the proof gives a bit more: there exist m and a finite subset T_0 of T such that any large model of T_0 with the property that $p(\mathcal{M}, \mathcal{N}) < \delta$ for all $\mathcal{M} \in F$ with $\|\mathcal{M}\| \leq m$ is within ϵ of some model of T .

5. FINAL REMARKS

We note that our Strong Removal Lemma for models of universal theories allows for the extension of property testing results to such structures. We have also successfully investigated the existence of limits, regularity and removal lemmas for general weighted structures. This work, which also uses the theory of measures on ultraproducts, will appear in a subsequent paper. If (G_n) is a sequence of graphs where the density of edges tends to zero (for example, a sequence of graphs of bounded degree) then the limit graphon is identically zero. There is a cultivated theory of limits for sequences of graphs of bounded degree, where the limit object (a *graphing*) is defined in a different way. It is not clear whether this theory can be extended in a meaningful way to more general classes of structures.

APPENDIX A. SOME EXAMPLES

A.1. Structures with one binary relation. Let \mathcal{L} only contain one binary relation symbol R . If \mathcal{M} is an \mathcal{L} -structure, then $\text{Code}(\mathcal{M}) = (\text{DH}_{\{\{1\}\{2\}\}}^{\mathcal{M}}, \text{DH}_{\{\{1,2\}\}}^{\mathcal{M}})$, which amounts to saying that we can view \mathcal{M} as a digraph with loops and we code it by separately recording the set of directed edges between distinct points and the set of vertices with a loop. An \mathcal{L} -limit \mathfrak{F} is of the form $(F_{\{\{1\}\{2\}\}}, F_{\{\{1,2\}\}})$ where $F_{\{\{1\}\{2\}\}}$ is a measurable subset of $[0, 1]^{r[2]}$ and $F_{\{\{1,2\}\}}$ is a measurable subset of $[0, 1]^{r[1]}$.

Let \mathcal{M} be a binary relation with $\|\mathcal{M}\| = n$, where we assume for convenience that $|\mathcal{M}| = [n]$. As in subsection 3.1 we can compute

$\Phi_t(\mathcal{M})$ from \mathfrak{F} by finding the measure of a certain subset of $[0, 1]^{r([n], 2)}$. To give a formula for the measure of this set in a palatable form, let f_p be the characteristic function of the measurable set F_p in the limit \mathcal{F} . Then $\Phi_t(\mathcal{M})$ is the integral of a certain product $F(\vec{x})G(\vec{x})$ over $[0, 1]^{r([n], 2)}$:

- $F(\vec{x})$ is the product of terms of the form $f_{\{\{1\}\{2\}\}}(x_{\{i\}}, x_{\{j\}}, x_{\{i,j\}})$, taken over all pairs (i, j) where $i \neq j$ and there is an edge from i to j .
- $G(\vec{x})$ is the product of terms of the form $f_{\{\{1,2\}\}}(x_{\{i\}})$, taken over all i such that there is a loop at i .

The formula for $\Phi_{t_{\text{ind}}}(\mathcal{M})$ is very similar in form, but the functions $F(\vec{x})$ and $G(\vec{x})$ are more complex in this case:

- $F(\vec{x})$ is the product over pairs (i, j) with $i \neq j$ of terms which have the form $f_{\{\{1\}\{2\}\}}(x_{\{i\}}, x_{\{j\}}, x_{\{i,j\}})$ if there is an edge from i to j , and the form $(1 - f_{\{\{1\}\{2\}\}}(x_{\{i\}}, x_{\{j\}}, x_{\{i,j\}}))$ if there is no edge from i to j .
- $G(\vec{x})$ is the product over i of terms which have the form $f_{\{\{1,2\}\}}(x_{\{i\}})$ if there is a loop at i , and the form $(1 - f_{\{\{1,2\}\}}(x_{\{i\}}))$ if there is no loop at i .

A.2. Digraphs. From our point of view a digraph is just a binary relation R in which vRv is false for all v . In syntactic terms digraphs are \mathcal{L} -structures which are models of the theory $T_{\text{digraphs}} = \{\forall v \neg vRv\}$, while in the language of forbidden substructures they are structures that forbid the substructure which has a single vertex with a loop. In terms of our coding, digraphs are \mathcal{L} -structures \mathcal{M} such that $\text{DH}_{\{\{1,2\}\}}^{\mathcal{M}} = \emptyset$. It follows that a limit of digraphs is a limit with $F_{\{\{1,2\}\}} = \emptyset$. In this case, if \mathcal{M} is an \mathcal{L} -structure with some loops then (as we would expect) the integral formula gives the value zero. If \mathcal{M} is a digraph, then the integral formula for $\Phi_t(\mathcal{M})$ only includes the first product $F(\vec{x})$. This integral can be simplified in a suggestive way, by collecting all the terms which involve each pair of variables. Define four functions $F_i(x, y)$ by setting:

- $F_0(x, y) = \lambda(\{z : (x, y, z) \notin F_{\{\{1\}\{2\}\}} \text{ and } (y, x, z) \notin F_{\{\{1\}\{2\}\}}\})$,
- $F_1(x, y) = \lambda(\{z : (x, y, z) \in F_{\{\{1\}\{2\}\}} \text{ and } (y, x, z) \notin F_{\{\{1\}\{2\}\}}\})$,
- $F_2(x, y) = \lambda(\{z : (x, y, z) \notin F_{\{\{1\}\{2\}\}} \text{ and } (y, x, z) \in F_{\{\{1\}\{2\}\}}\})$,
- $F_3(x, y) = \lambda(\{z : (x, y, z) \in F_{\{\{1\}\{2\}\}} \text{ and } (y, x, z) \in F_{\{\{1\}\{2\}\}}\})$.

It is immediate from the definitions that $F_0 + F_1 + F_2 + F_3 = 1$, F_0 and F_3 are symmetric, and $F_1(x, y) = F_2(y, x)$. The integral formula for $\Phi_t(\mathcal{M})$ can now be written as the integral over $[0, 1]^2$ of a product (taken over all pairs (i, j) with $i < j$) of terms of the following form:

- $(F_1 + F_3)(x_{\{i\}}, x_{\{j\}})$ if there is only an edge from i to j ,
- $(F_2 + F_3)(x_{\{i\}}, x_{\{j\}})$ if there is only an edge from j to i ,
- $F_3(x_{\{i\}}, x_{\{j\}})$ if there are edges in both directions between i and j .

This calculation has recovered a version of the notion of “digraph limit” due to Offner and Pikhurko [7].

The integral formula for $\Phi_{t_{\text{ind}}}(\mathcal{M})$ can be similarly written as the integral over $[0, 1]^2$ of a product (taken over all pairs $\{i, j\}$ with $i < j$) of terms of the following form:

- $F_0(x_{\{i\}}, x_{\{j\}})$ if there is no edge in either direction between i and j ,
- $F_1(x_{\{i\}}, x_{\{j\}})$ if there is only an edge from i to j ,
- $F_2(x_{\{i\}}, x_{\{j\}})$ if there is only an edge from j to i ,
- $F_3(x_{\{i\}}, x_{\{j\}})$ if there are edges in both directions between i and j .

A.3. Graphs and hypergraphs. A graph is just a symmetric digraph, or a model of $T_{\text{graphs}} = \{\forall v \neg vRv, \forall v \forall w (vRw \implies wRv)\}$. In the language of forbidden substructures, graphs are structures which forbid the single vertex with a loop, and a pair of vertices with an edge going only one way. It is easy to see that (in our formulation) the limit of a sequence of graphs will have $F_{\{\{1,2\}\}} = \emptyset$, while $F_{\{\{1\}\{2\}\}}$ is (without loss of generality) a set of triples (x, y, z) which is symmetric in the sense that $(x, y, z) \in F_{\{\{1\}\{2\}\}} \iff (y, x, z) \in F_{\{\{1\}\{2\}\}}$.

The integral formula for $\Phi_t(M)$ now simplifies further: retaining our notation for digraph limits we have $F_1 = F_2 = 0$, $F_0 + F_3 = 1$ and it is easy to see that we recovered the classical graph limit called graphon in the form of the symmetric measurable function F_3 .

A.4. Posets. A (*strict*) *partial order* on X is an irreflexive, transitive binary relation. The theory of partially ordered sets, or *posets* is $T_{\text{posets}} = \{\forall v \neg vRv, \forall u \forall v \forall w (uRv \wedge vRw \implies uRw)\}$. Forbid^T = $\{\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3, \mathcal{M}_4\}$ where

- $|\mathcal{M}_1| = [1]$ with $R^{\mathcal{M}_1} = \{(1, 1)\}$,
- $|\mathcal{M}_2| = [2]$ with $R^{\mathcal{M}_2} = \{(1, 2), (2, 1)\}$,
- $|\mathcal{M}_3| = [3]$ with $R^{\mathcal{M}_3} = \{(1, 2), (2, 3)\}$,
- $|\mathcal{M}_4| = [3]$ with $R^{\mathcal{M}_4} = \{(1, 2), (2, 3), (3, 1)\}$.

An \mathcal{L} -limit \mathcal{F} has the property that $\Phi_{t_{\text{ind}}}(\mathcal{M}_i) = 0$ for $i \in [4]$. A poset is also a digraph. Using the simplified notation for digraph limits, we can say that a poset limit has the additional properties that $F_3 = 0$ and if $F_1(x, y) > 0$ and $F_1(y, z) > 0$, then $F_1(x, z) = 1$.

Janson [2] defined a notion of poset limit. Such a limit consists of a partial ordering \prec on $[0, 1]$ and a measurable function $W : [0, 1]^2 \rightarrow [0, 1]$ with the properties that

- $W(x, x) = 0$,
- $W(x, y) > 0 \implies W(y, x) = 0$,
- $W(x, y) > 0$ and $W(y, z) > 0 \implies W(x, z) = 1$.

The corresponding construction of a random partial ordering of $[n]$ involves randomly choosing x_i and $z_{i,j}$ in $[0, 1]$, and then putting i below j if and only if $x_i \prec x_j$ and $z_{i,j} < W(x_i, x_j)$. Given a poset limit in Janson's sense, we can easily retrieve an equivalent poset limit in our sense. In the other direction, while we can show that a Janson limit (\prec, W) equivalent to any poset limit \mathfrak{F} in our sense exists, we do not have an easy algebraic way to find such a Janson limit.

APPENDIX B. PROOF OF LEMMA 1

Let the variable symbols of \mathcal{L} be listed as x_1, x_2, \dots . For each $m > 0$ let X_m^1 be the set of formulae of the form $x_i = x_j$ for $1 \leq i < j \leq m$, and let Y_m^1 be the set of conjunctions which contain exactly one of $\psi, \neg\psi$ for each $\psi \in X_m^1$ and are logically consistent (we note that such conjunctions correspond in the obvious way to partitions of the set $[m]$). Let X_m^2 be the set of formulae of the form $R_i(z_1, \dots, z_{n_i})$ where $z_j \in \{x_1, \dots, x_m\}$, and let Y_m^2 be the set of conjunctions which contain exactly one of $\psi, \neg\psi$ for each $\psi \in X_m^2$. Intuitively formulae in Y_m^1 are complete descriptions of the equality relation on $\{x_1, \dots, x_m\}$, while formulae in Y_m^2 are complete descriptions of the other relations in the language on the same set. If ϕ is a quantifier-free formula which only mentions variables among x_1, \dots, x_m , then by elementary propositional logic it is equivalent to some disjunction of formulae of the form $\psi^1 \wedge \psi^2$ where $\psi^i \in Y_m^i$. Call this the ‘‘normal form’’ for ϕ .

Each axiom Ψ of T is equivalent to a sentence of the form $\forall x_1 \dots \forall x_m \phi$ where ϕ is quantifier-free and is in normal form, say $\phi = \bigvee_i (\psi_i^1 \wedge \psi_i^2)$. Clearly Ψ is equivalent to the set of formulae of the form $\forall x_1 \dots \forall x_m (\psi \implies \phi)$ with $\psi \in Y_m^1$, where breaking up the formula $\forall x_1 \dots \forall x_m \phi$ in this way corresponds to a case analysis depending on the equality relation on $\{x_1, \dots, x_m\}$.

For each $\psi \in Y_m^1$, it is easy to see that $\psi \implies \phi$ is equivalent to $\psi \implies \bigvee_{i, \psi_i^1 = \psi} \psi_i^2$. Finally, if ψ corresponds to a partition into t classes, then $\forall x_1 \dots \forall x_m \psi \rightarrow \phi$ is equivalent (replacing variables in the j^{th} class by x_j and eliminating redundancies) to a formula of the form $\forall x_1 \dots \forall x_t (\chi \implies \Phi)$, where $\chi = \bigvee_{1 \leq i < j \leq t} x_i \neq x_j$ and Φ is a disjunction of formulae in Y_t^2 . The point of all this is that axioms

of the form $\forall x_1 \dots \forall x_t (\chi \implies \Phi)$ have a very natural combinatorial interpretation: such an axiom is true in \mathcal{M} if and only if for every $A \subseteq |\mathcal{M}|$ with $|A| = t$, the isomorphism type of $\mathcal{M}|_A$ is among those specified by the elements of Y_t^2 appearing in Φ . In keeping with the usual usage in graph theory, we can view axioms of this kind as saying that certain induced substructures are *forbidden*: what we have shown is that for any universal theory T there is a sequence $\langle \text{Forbid}_t^T : t > 0 \rangle$ such that Forbid_t^T is a set of isomorphism types for \mathcal{L} -structures of size t , and the models of T are exactly those \mathcal{M} such that the isomorphism type of $\mathcal{M}|_A$ is not in Forbid_t^T for every t and every $A \subseteq |\mathcal{M}|$ with $|A| = t$.

REFERENCES

- [1] G. Elek and B. Szegedy. A Measure-theory Approach to the Theory of Dense Hypergraphs. *Advances in Mathematics*, 231(3-4):1731–1772, 2012.
- [2] S. Janson. Poset Limits and Exchangeable Random Posets. *Combinatorica*, 31:529–563, 2011.
- [3] P. Loeb. Conversion from Nonstandard to Standard Measure Spaces and Applications in Probability Theory. *Transactions of the American Mathematical Society*, 211:113–122, 1975.
- [4] L. Lovász. *Large Networks and Graph Limits*, volume 60 of *Colloquium Publications*. American Mathematical Society, 2012.
- [5] L. Lovász and B. Szegedy. Limits of Dense Graph Sequences. *Journal of Combinatorial Theory Ser. B*, 96:933–957, 2006.
- [6] D. Maharam. On Homogeneous Measure Algebras. *Proceedings of the National Academy of Science USA*, 28:108–111, 1942.
- [7] D. Offner. *Extremal Problems on the Hypercube*. PhD thesis, Carnegie Mellon University, 2009.
- [8] A. Razborov. Flag Algebras. *Journal of Symbolic Logic*, 72(4):1239–1282, 2007.
- [9] D. Scott. Measurable cardinals and constructible sets. *Bulletin de l'Académie Polonaise des Sciences (Série des sciences mathématiques, astronomiques et physiques)*, 9:521–524, 1961.