Logical Methods in Computer Science
Volume 15, Issue 3, 2019, pp. 4:1–4:12
https://lmcs.episciences.org/

Submitted Jul. 10, 2017
Published Jul. 25, 2019

LOGICAL PROPERTIES OF RANDOM GRAPHS FROM SMALL ADDABLE CLASSES

ANUJ DAWAR [a] AND ERYK KOPCZYŃSKI [b]

[a] Department of Computer Science and Technology, University of Cambridge
e-mail address: anuj.dawar@cl.cam.ac.uk

[b] Institute of Informatics, University of Warsaw
e-mail address: erykk@mimuw.edu.pl

Abstract. We establish zero-one laws and convergence laws for monadic second-order logic (MSO) (and, a fortiori, first-order logic) on a number of interesting graph classes. In particular, we show that MSO obeys a zero-one law on the class of connected planar graphs, the class of connected graphs of tree-width at most k and the class of connected graphs excluding the k-clique as a minor. In each of these cases, dropping the connectivity requirement leads to a class where the zero-one law fails but a convergence law for MSO still holds.

1. Introduction

The zero-one law for first-order logic [GKLT69, Fag76] established that every first-order sentence ϕ, when evaluated over a random n-element finite structure, has a probability of being true that converges to either 0 or 1 as n goes to infinity. This prompted much further investigation into the asymptotic behaviour of classes of structures definable in logic. Zero-one laws have been established for fragments of second-order logic [KV90]; extensions of first-order logic such as the infinitary logic $L^{\omega}_{\infty\omega}$ [KV92] which subsumes various fixed-point logics; and logics with generalized quantifiers [DG10], among many others.

Another widely studied extension of first-order logic is *monadic second-order logic* (MSO). This does not have a zero-one law but its asymptotic behaviour has been studied on restricted classes of structures. For many interesting classes it does admit a zero-one law, such as on free labelled trees [McC02]. On rooted labelled trees, MSO does not have a zero-one law, but still admits a convergence law [McC02]. This means that the probability of any given sentence ϕ being true in an n-element structure does converge to a limit, though that limit is not necessarily 0 or 1. Zero-one and convergence laws for MSO on a number of other classes are shown in [Com89].

Key words and phrases: zero-one laws, limit law, first-order logic, random planar graphs, smooth addable classes.

The research reported here was carried out during a visit by the second author to the University of Cambridge in 2011, supported by the ESF Research Networking Programme GAMES. Eryk Kopczyński was also partially supported by the Polish National Science research grant DEC-2012/07/D/ST6/02435.

DOI:10.23638/LMCS-15(3:4)2019

In this paper, we are concerned with the asymptotic behaviour of first-order logic (FO) and MSO on restricted classes of finite structures, and more specifically, restricted classes of graphs. We look particularly at *tame* classes of graphs in the sense of [Daw07]. These include planar graphs, graphs of bounded treewidth and some other proper minor-closed classes. A flavour of our results is given by the following examples. There is a constant $c \leq 0.036747$ such that the asymptotic probability of any MSO sentence ϕ on the class of planar graphs converges to a real number in the range $[0, c) \cup (1 - c, 1]$. On the class of *connected*, planar graphs, MSO admits a zero-one law. These results can be strengthened (modulo changing the constant c) from planar graphs to other minor-closed classes of graphs that are *smooth* and *addable*, which we define formally later. Here we note that examples of smooth, addable, minor-closed graph classes include not just the planar graphs but also the class of graphs of treewidth at most k, for any k.

Technically, we rely on the combinatorial results on random planar graphs proved in [MSW05] and the extensions to minor-closed classes established in [McD09]. We combine these with logical techniques from [Com89] and [McC02].

Since the first submission, it has come to our attention that some of the results reported here have also been obtained independently in [HMNT18].

2. Graph Classes

Let R_n denote a random graph drawn from the uniform distribution on graphs on the vertex set $[n] = \{1, \ldots, n\}$. Equivalently, R_n is obtained by putting, for each i, j with $1 \leq i < j \leq n$ an edge between i and j with probability $1/2$. If $\mathcal{C}$ is a class of graphs closed under isomorphisms, we write $\mathcal{C}_n$ for the graphs in $\mathcal{C}$ on the vertex set $[n]$ and $R_n(\mathcal{C})$ to denote the random graph drawn from the uniform distribution on $\mathcal{C}_n$. Unlike with R_n, it is not immediately clear how to effectively sample from this distribution. When $\mathcal{C}$ is the class of planar graphs, a first partially successful attempt was in [DVW96], using a Markov chain whose only stable distribution is the uniform distribution on $\mathcal{C}_n$. However, this is not quite practical as the mixing rate of this Markov chain is unknown. Nevertheless, this formulation did enable experimental validation of some conjectures about random planar graphs. Finally, in [BGK07], it is shown that there is a polynomial-time algorithm that can generate a random planar graph on $[n]$.

An excellent analysis of random planar graphs, resolving some conjectures of [DVW96] was given in [MSW05]. Some of this analysis was predicated on a conjecture, termed the "isolated vertices conjecture" to the effect that the number of isolated vertices in a random planar graph on $[n]$ tends to a limit as n increases. This conjecture was proved in [GN05] and later shown [McD09] to be an instance of the *smoothness* of *addable* minor-closed classes. Much of the analysis of random planar graphs in [MSW05] can be extended to other graph classes that are smooth, addable and *small* [MSW06]. We begin by defining these central notions.

Definition 2.1. Say that a graph class $\mathcal{C}$ is *decomposable* if $G \in \mathcal{C}$ if, and only if, every connected component of G is in $\mathcal{C}$.

Equivalently, $\mathcal{C}$ is decomposable if, and only if, $\mathcal{C}$ is both closed under disjoint unions and closed under taking connected components.

Definition 2.2. Say that graph class $\mathcal{C}$ is *bridge-addable* if, for every $G \in \mathcal{C}$, if u and v are vertices in distinct connected components of G, then the graph obtained by adding the edge $\{u, v\}$ to G is also in $\mathcal{C}$.

Definition 2.3. A graph class $\mathcal{C}$ is *addable* if it is both decomposable and bridge-addable.

Note that the class of planar graphs is clearly addable. As noted in [MSW06], the following classes of graphs are all addable: forests; the class of graphs of treewidth at most k; the class of graphs with no cycle of length greater than k; the class of graphs that exclude K_k as a minor. On the other hand, the class of graphs embeddable in a torus is not addable. It is bridge-addable but not decomposable since it contains K_5 but not the graph that is the disjoint union of two copies of K_5. In general, the class of graphs of genus at most k for positive k is not addable.

We write $Q_n(\mathcal{C})$ to denote the number of graphs in $\mathcal{C}_n$ (i.e. graphs in $\mathcal{C}$ on the vertex set $[n]$).

Definition 2.4. A graph class $\mathcal{C}$ is *small* if $Q_n(\mathcal{C}) \leq d^n n!$ for some $d \in \mathbb{R}$.

With any class $\mathcal{C}$, we can associate a *growth constant* $\gamma_{\mathcal{C}}$ defined by

$$\gamma_{\mathcal{C}} = \limsup_{n\to\infty} (Q_n(\mathcal{C})/n!)^{1/n}.$$

Then, $\mathcal{C}$ is small just in case $\gamma_{\mathcal{C}}$ is finite.

Definition 2.5 [MSW06]**.** A graph class $\mathcal{C}$ is *smooth* if $\frac{Q_n}{nQ_{n-1}}$ tends to a finite limit as $n \to \infty$.

It is known (see [MSW06]) that if $\mathcal{C}$ is smooth, then the limit of $\frac{Q_n}{nQ_{n-1}}$ is the growth constant $\gamma_{\mathcal{C}}$. Thus, every smooth class is small. However, there are small graph classes that are not smooth. When $\mathcal{C}$ is the class of forests, $\gamma_{\mathcal{C}} = e$. We can now state the two results on smoothness that we need.

Theorem 2.6 [GN05]**.** *The class of planar graphs is smooth, and the growth constant is* $\gamma \approx 27.22679$ *and is given by an explicit analytic expression.*

Theorem 2.7 [McD09]**.** *Each addable proper minor-closed class of graphs is smooth.*

3. Logics

We assume the reader is familiar with the syntax and semantics of first-order logic (FO) and monadic second-order logic (MSO) interpreted on finite structures, as defined, for instance in [Lib04]. We give a brief review of definitions, especially where our notation deviates from the standard.

3.1. **Basics.** We assume we have two disjoint countable sets of symbols Var^1 (the first-order variables) and Var^2 (the second-order variables). Given a graph $G = (V, E)$, a *valuation* σ on G is a pair (σ^1, σ^2) of *partial* functions $\sigma^1 : \mathrm{Var}^1 \rightharpoonup V$ and $\sigma^2 : \mathrm{Var}^2 \rightharpoonup \mathscr{P}(V)$. For $x \in \mathrm{Var}^1 \cup \mathrm{Var}^2$, we use $\sigma(x)$ to denote $\sigma^1(x)$ or $\sigma^2(x)$ as appropriate.

A *structure* is a pair $\mathfrak{G} = (G, \sigma)$ where G is a graph and σ is a valuation on G. If $\mathfrak{G} = (G = (V, E), \sigma)$ is a structure, we sometimes write $G(\mathfrak{G})$, $V(\mathfrak{G})$, $E(\mathfrak{G})$, $\sigma(\mathfrak{G})$, $x(\mathfrak{G})$ to denote $G, V, E, \sigma, \sigma(x)$, respectively. Moreover, we write $\mathrm{Var}^1(\mathfrak{G})$ and $\mathrm{Var}^2(\mathfrak{G})$ to denote

the subsets of Var^1 and Var^2 respectively on which $\sigma^1(\mathfrak{G})$ and $\sigma^2(\mathfrak{G})$ are defined. For even greater brevity, when referring to a structure $\mathfrak{G}_1$, we may write G_1, V_1, etc. for $G(\mathfrak{G}_1)$, $V(\mathfrak{G}_1)$, etc. when no ambiguity arises. We write $\sigma[v/x]$ to denote the valuation that agrees with σ at all values other than x and that maps x to v. We also write $\mathfrak{G}[v/x]$ to denote $(G(\mathfrak{G}), \sigma(\mathfrak{G})[v/x])$ and $\mathfrak{G}[/x]$ to denote the structure $(G(\mathfrak{G}), \sigma'(\mathfrak{G}))$ where the valuation σ' is the same as σ except that it is undefined at x.

Note that, by including the interpretation of variables in our definition of structure, we can uniformly talk of graphs, coloured graphs and graphs with distinguished constants as structures, all over a fixed vocabulary.

The formulas of MSO are built-up as usual according to the following grammar, where ϕ_1 and ϕ_2 are formulas, $x, y \in \mathrm{Var}^1$ and $X, Y \in \mathrm{Var}^2$.

$$\phi \ ::= \ x \in X \mid E(x,y) \mid \phi_1 \wedge \phi_2 \mid \phi_1 \vee \phi_2 \mid \neg\phi_1 \mid \forall x\phi \mid \exists x\phi \mid \forall_2 X\phi \mid \exists_2 X\phi$$

The formulas of first-order logic are those that involve no occurrence of a variable from Var^2.

The definition of satisfaction $\mathfrak{G} \models \phi$ is standard. We define the *quantifier rank* of a formula ϕ, denoted $\mathrm{qr}(\phi)$, to be the maximum depth of nesting of quantifiers in ϕ counting both first- and second-order quantifiers (as in [Lib04, Definition 7.4]).

3.2. **Ehrenfeucht-Fraïssé games.** We write $\mathfrak{G}_1 \equiv_m \mathfrak{G}_2$ to denote that the two structures $\mathfrak{G}_1$ and $\mathfrak{G}_2$ cannot be distinguished by any formula with quantifier rank at most m. Formally, this is defined by induction on m as follows. We say that $\mathfrak{G}_1 \equiv_0 \mathfrak{G}_2$ if, for any quantifier-free formula ϕ such that σ_1 and σ_2 are defined on all the free variables of ϕ, we have that $\mathfrak{G}_1 \models \phi$ if, and only if, $\mathfrak{G}_2 \models \phi$. Inductively, we say that $\mathfrak{G}_1 \equiv_{m+1} \mathfrak{G}_2$ if all the following conditions are satisfied.

- For $x \in \mathrm{Var}^1$, $\forall v_1 \in V_1 \exists v_2 \in V_2 \mathfrak{G}_1[v_1/x] \equiv_m \mathfrak{G}_2[v_2/x]$.
- For $x \in \mathrm{Var}^1$, $\forall v_2 \in V_2 \exists v_1 \in V_1 \mathfrak{G}_1[v_1/x] \equiv_m \mathfrak{G}_2[v_2/x]$.
- For $X \in \mathrm{Var}^2$, $\forall U_1 \subseteq V_1 \exists U_2 \subseteq V_2 \mathfrak{G}_1[U_1/X] \equiv_m \mathfrak{G}_2[U_2/X]$.
- For $X \in \mathrm{Var}^2$, $\forall U_2 \subseteq V_2 \exists U_1 \subseteq V_1 \mathfrak{G}_1[U_1/X] \equiv_m \mathfrak{G}_2[U_2/X]$.

As usual, this definition $\mathfrak{G}_1 \equiv_m \mathfrak{G}_2$ can be understood as a game between two players conventionally called Spoiler and Duplicator (known as the *m-round Ehrenfeucht-Fraïssé game*). The game is played on a "board" consisting of two structures $\mathfrak{G}_1$ and $\mathfrak{G}_2$. For $m = 0$, Duplicator wins iff $\mathfrak{G}_1 \equiv_0 \mathfrak{G}_2$. Otherwise, Spoiler chooses either a variable $x \in \mathrm{Var}^1$ and a vertex v_i in one of the graphs G_i or a variable $X \in \mathrm{Var}^2$ and a set of vertices $U_i \subseteq V_i$ in the graph G_i. In the first case, Duplicator responds by choosing an element $v_{3-i} \in G_{3-i}$ and the play then proceeds for $m-1$ rounds starting with the board $\mathfrak{G}_1[v_1/x]$ and $\mathfrak{G}_2[v_2/x]$. In the second case, Duplicator responds by choosing a set $U_{3-i} \subseteq V_{3-i}$ and again the play then proceeds for $m-1$ rounds starting with the board $\mathfrak{G}_1[U_1/X]$ and $\mathfrak{G}_2[U_2/X]$. Then, it is clear that Duplicator has a strategy for winning the m round game on the board $\mathfrak{G}_1$ and $\mathfrak{G}_2$ just in case $\mathfrak{G}_1 \equiv_m \mathfrak{G}_2$. The connection with MSO comes from the following standard theorem (see [Lib04, Theorem 7.7])

Theorem 3.1. *The following conditions are equivalent:*

- $\mathfrak{G}_1 \equiv_m \mathfrak{G}_2$
- $\mathfrak{G}_1$ *and* $\mathfrak{G}_2$ *satisfy exactly the same MSO formulas of quantifier rank at most* m.

Finally, we note that, if we have only a fixed finite set of variables, then the equivalence relation $\equiv_m$ is of finite index for each m. That is, fix a finite set $\Xi \subseteq \mathrm{Var}$ and let $\mathcal{S}[\Xi]$ denote the class of structures $\mathfrak{G}$ such that $\mathrm{Var}^1(\mathfrak{G}), \mathrm{Var}^2(\mathfrak{G}) \subseteq \Xi$.

Proposition 3.2. *The relation $\equiv_m$ restricted to $\mathcal{S}[\Xi]$ has finite index.*

The proof is a proof by induction that there are, up to equivalence, only finitely many formulas of MSO of quantifier rank m with free variables among Ξ (see [Lib04, Prop. 7.5]). We write $t_m(\Xi)$ to denote the index of $\equiv_m$ restricted to $\mathcal{S}[\Xi]$. Note that the value of $t_m(\Xi)$ is completely determined by the number of elements in $\Xi \cap \mathrm{Var}^1$ and $\Xi \cap \mathrm{Var}^2$.

4. Adding graphs

For structures $\mathfrak{G}_1$ and $\mathfrak{G}_2$, the disjoint union $\mathfrak{G}_1 \oplus \mathfrak{G}_2$ is defined if $\mathrm{Var}_1^1 \cap \mathrm{Var}_2^1 = \emptyset$. In this case, the set of vertices of $\mathfrak{G}_3 = \mathfrak{G}_1 \oplus \mathfrak{G}_2$ is the disjoint union of V_1 and V_2, $E_3 = E_1 \cup E_2$, $\sigma_3^1(x) = \sigma_i^1(x)$ for $i \in \{1,2\}$ whenever $x \in \mathrm{Var}_i^1$ and $\sigma_3^2(X) = \sigma_1^2(X) \cup \sigma_2^2(X)$.

It is well-known that the equivalence relation $\equiv_m$ is a congruence with respect to the disjoint union of structures. Moreover, for every m and finite set $\Xi \subseteq \mathrm{Var}^1 \cup \mathrm{Var}^2$ there is a threshold $r_m(\Xi)$ such that for any $p, q > r_m(\Xi)$, the disjoint union of p copies of a structure is $\equiv_m$-equivalent to the disjoint union of q copies. We formally state this for later use.

Lemma 4.1. *Fix a positive integer m and a finite set $\Xi \subseteq \mathrm{Var}^1 \cup \mathrm{Var}^2$. There is a positive integer $r_m(\Xi)$ such that if $\{\mathfrak{G}_i \mid i \in I\}$ is a collection of structures such that $\mathfrak{G}_i \equiv \mathfrak{G}_j$ for all $i, j \in I$ and $P, Q \subseteq I$ are sets such that either $|P| = |Q|$ or $|P|, |Q| \geq r_m(\Xi)$ then*

$$\bigoplus_{i \in P} \mathfrak{G}_i \equiv_m \bigoplus_{j \in Q} \mathfrak{G}_j.$$

Indeed, similar facts can be established for many operations other than disjoint union (see, for example [Mak04]). Here, we are also interested in a particular operation of taking the disjoint union of two structures while adding an edge between distinguished vertices (as in the definition of bridge addable graphs: Definition 2.2). We give the formal definition below and prove the properties we need.

We say that a structure $\mathfrak{G}$ is *rooted* if $\mathrm{Root} \in \mathrm{Var}^1(\mathfrak{G})$. Let $\mathfrak{G}_1$ and $\mathfrak{G}_2$ be two rooted structures such that $\mathrm{Var}_1^1 \cap \mathrm{Var}_2^1 = \{\mathrm{Root}\}$ and $V_1 \cap V_2 = \emptyset$. We write $\mathfrak{G}_1 + \mathfrak{G}_2$ to denote the structure $\mathfrak{G}$ such that:

- $V = V_1 \cup V_2$, $E = E_1 \cup E_2 \cup \{\sigma_1(\mathrm{Root}), \sigma_2(\mathrm{Root})\}$
- $\sigma(X) = \sigma_1(X) \cup \sigma_2(X)$ for $X \in \mathrm{Var}_1^2 \cap \mathrm{Var}_2^2$
- $\sigma(X) = \sigma_i(X)$ for $X \in \mathrm{Var}_i^2 - \mathrm{Var}_{3-i}^2$
- $\sigma(\mathrm{Root}) = \sigma_1(\mathrm{Root})$
- $\sigma(x) = \sigma_i(x)$ for $x \in \mathrm{Var}_i^1$ and $x \neq \mathrm{Root}$

That is to say that $\mathfrak{G}_1 + \mathfrak{G}_2$ is obtained by taking the disjoint union of the structures $\mathfrak{G}_1$ and $\mathfrak{G}_2$ seen as coloured graphs, putting an edge between the two roots, and making the root of $\mathfrak{G}_1$ the root of the combined structure. This asymmetry in the choice of root means that, in general $\mathfrak{G}_1 + \mathfrak{G}_2 \neq \mathfrak{G}_2 + \mathfrak{G}_1$. However, it is still the case that $(\mathfrak{G}_1 + \mathfrak{G}_2) + \mathfrak{G}_3 = (\mathfrak{G}_1 + \mathfrak{G}_3) + \mathfrak{G}_2$.

A simple argument using Ehrenfeucht-Fraïssé games establishes the following proposition.

Proposition 4.2. *If $\mathfrak{G}_1 \equiv_m \mathfrak{G}_2$, then $\mathfrak{G}_0 + \mathfrak{G}_1 \equiv_m \mathfrak{G}_0 + \mathfrak{G}_2$.*

Proof. Suppose Duplicator has a winning strategy in the m-move Ehrenfeucht-Fraïssé game played on $\mathfrak{G}_1$ and $\mathfrak{G}_2$. We show that she also has a winning strategy in the m-move game on $\mathfrak{G}_0 + \mathfrak{G}_1$ and $\mathfrak{G}_0 + \mathfrak{G}_2$. The strategy is described as follows.

- if Spoiler chooses $w \in V_0$, then Duplicator responds with the same w;

- if Spoiler chooses $w \in V_i$ for $i \in \{1,2\}$, then Duplicator responds with $w' \in V_{3-i}$ given by her winning strategy in the game on $\mathfrak{G}_1$ and $\mathfrak{G}_2$; and
- if Spoiler chooses $W \subseteq V_0 \cup V_i$ for $i \in \{1,2\}$, then Duplicator responds with $(W \cap V_0) \cup W_2$, where W_2 is her response to $W \cap V_i$ in the game on $\mathfrak{G}_1$ and $\mathfrak{G}_2$.

It is easy to show that this does indeed describe a winning strategy. □

If $\mathfrak{G}_2$ is a structure with $\mathrm{Var}_2^1 = \{\mathrm{Root}\}$ and c is a positive integer, we write $\mathfrak{G}_1 + c\mathfrak{G}_2$ to denote the structure obtained by adding $\mathfrak{G}_2$ to $\mathfrak{G}_1$ c times. More formally, this is defined by induction on c:

- $\mathfrak{G}_1 + 0\mathfrak{G}_2 = \mathfrak{G}_1$,
- $\mathfrak{G}_1 + (c+1)\mathfrak{G}_2 = (\mathfrak{G}_1 + c\mathfrak{G}_2) + \mathfrak{G}_2$

Analogously to Lemma 4.1, we have the following proposition.

Proposition 4.3. *Let $\Xi \subseteq \mathrm{Var}$ be a finite set. For each m, there is number $q_m(\Xi)$ such that for any $\mathfrak{G} \in \mathcal{S}[\Xi]$ and any structure $\mathfrak{G}_0$ we have*

$$\mathfrak{G}_0 + p\mathfrak{G} \equiv_m \mathfrak{G}_0 + q\mathfrak{G}$$

whenever $p, q \geq q_m(\Xi)$.

Proof. We define the value of $q_m(\Xi)$ by induction on m, simultaneously for all finite sets $\Xi \subseteq \mathrm{Var}$.

- $q_0(\Xi) = 0$ for all Ξ; and
- $q_{m+1}(\Xi) = \max(q_m(\Xi \cup \{x\}) + 1,\ t_m(\Xi \cup \{X\}) \cdot q_m(\Xi \cup \{X\}) + m)$,

where $x \in \mathrm{Var}^1$, $X \in \mathrm{Var}^2$ are any variables not in Ξ. Recall that $t_m(\Xi)$ denotes the index of the relation $\equiv_m$ in $\mathcal{S}[\Xi]$.

Now, we argue by induction on m that if $p, q \geq q_m(\Xi)$ then $\mathfrak{G}_0 + p\mathfrak{G} \equiv_m \mathfrak{G}_0 + q\mathfrak{G}$. In the following, we use the sets $[p]$ and $[q]$ to index the copies of $\mathfrak{G}$. Thus, we may refer to the graph G_i for $i \in [p]$. The base case, $m = 0$ is immediate. Suppose now that the claim is true for some m and all finite Ξ, and let $\mathfrak{G}_A = \mathfrak{G}_0 + p\mathfrak{G}$ and $\mathfrak{G}_B = \mathfrak{G}_0 + q\mathfrak{G}$ for some $p, q \geq q_{m+1}(\Xi)$. We consider a number of cases corresponding to different moves that Spoiler might make in the Ehrenfeucht-Fraïssé game.

- Suppose $x \in \mathrm{Var}^1$ and $v \in V_0$. Then,

$$\begin{array}{rcl} (\mathfrak{G}_0 + p\mathfrak{G})[v/x] & = & \mathfrak{G}_0[v/x] + p\mathfrak{G} \\ & \equiv_m & \mathfrak{G}_0[v/x] + q\mathfrak{G} \\ & = & (\mathfrak{G}_0 + q\mathfrak{G})[v/x], \end{array}$$

  where the central $\equiv_m$ is true by induction hypothesis, since $p, q \geq q_{m+1}(\Xi) > q_m(\Xi \cup \{x\})$.
- Suppose $x \in \mathrm{Var}^1$ and $v \in V_i$ for $i \in [p]$. Then

$$\begin{array}{rcl} (\mathfrak{G}_0 + p\mathfrak{G})[v/x] & \cong & (\mathfrak{G}_0 + \mathfrak{G})[v/x] + (p-1)\mathfrak{G} \\ & \equiv_m & (\mathfrak{G}_0 + \mathfrak{G})[v/x] + (q-1)\mathfrak{G} \\ & \cong & (\mathfrak{G}_0 + q\mathfrak{G})[v/x], \end{array}$$

  where the central $\equiv_m$ is again true by induction hypothesis, since $p-1, q-1 \geq q_m(\Xi \cup \{x\})$. The situation where $v \in V_i$ for $i \in [q]$ is entirely symmetric.
- Suppose now that $X \in \mathrm{Var}^2$ and $U \subseteq V_A$. We partition $[p]$ into sets $P_1, \ldots, P_t$ where $t \leq t_m(\Xi \cup \{X\})$ such that i and j are in the same part if, and only if, $\mathfrak{G}_i[U \cap V_i/X] \equiv_m \mathfrak{G}_j[U \cap V_j/X]$. Since $p, q \geq q_{m+1}(\Xi) \geq t_m(\Xi \cup \{X\}) \cdot q_m(\Xi \cup \{X\})$ we can partition $[q]$ into parts $Q_1, \ldots, Q_t$ such that for all k, either P_k and Q_k have the same size, or they

both have more than $q_m(\Xi \cup \{X\})$ elements. For each l choose an $i_l \in P_l$ and define, for $j \in [q]$, U'_j to be $U \cap V_{i_l}$ whenever $j \in Q_l$. Let $U' = \bigcup_{j\in[q]} U'_j$. Then, by the induction hypothesis and Proposition 4.2, it follows that $\mathfrak{G}_A[U/X] \equiv_m \mathfrak{G}_B[U/X]$. Again, the case when $U \subseteq V_B$ is entirely symmetric. □

4.1. **Universal Connected Structure.** There is a sentence of MSO (with quantifier rank 5) which is true in a structure $\mathfrak{G}$ if, and only if, $\mathfrak{G}$ is connected. It follows that if $m \geq 5$ then each $\equiv_m$ class of structures either contains only connected structures or disconnected ones.

For the rest of this subsection, we fix a value m with $m \geq 5$. Also, let $\mathcal{C}$ be a class of graphs closed under the operation $+$. Let $\mathfrak{G}_1, \ldots, \mathfrak{G}_t$ (with $t < t_m(\emptyset)$) be a set of representatives from $\mathcal{C}$ of all $\equiv_m$ classes of connected graphs which have elements in $\mathcal{C}$. Let $\mathfrak{G}_R$ denote the rooted structure with one element, no edges, and interpreting no variables other than Root. We define the *m-universal connected rooted graph in* $\mathcal{C}$ to be the structure $\mathfrak{G}_{U(m)} = \mathfrak{G}_R + \sum_{1\leq i\leq t} q_k(\emptyset)\mathfrak{G}_i$. That is, $\mathfrak{G}_{U(m)}$ is obtained by adding $q_k(\emptyset)$ copies of a representative of each $\equiv_m$ class of graphs to $\mathfrak{G}_R$. Changing the order in which these graphs are added does not change the isomorphism type of $\mathfrak{G}_{U(m)}$ and changing the choice of representatives $\mathfrak{G}_1, \ldots, \mathfrak{G}_t$ does not affect the $\equiv_m$ class of $\mathfrak{G}_{U(m)}$.

Definition 4.4. Say that a rooted structure $\mathfrak{G}$ *appears* in a graph $\mathfrak{G}_1$ if there is an induced substructure $\mathfrak{G}_2$ of $\mathfrak{G}_1$ and a vertex $r \in V_2$ such that:

- $\mathfrak{G}$ is isomorphic to $\mathfrak{G}_2[r/\text{Root}]$; and
- there is only one edge between V_2 and $V_1 \setminus V_2$ and this edge is incident on r.

Proposition 4.5. *If the universal structure $\mathfrak{G}_{U(m)}$ appears in a connected graph $\mathfrak{G}$, then $\mathfrak{G} \equiv_m \mathfrak{G}_{U(m)}[/\text{Root}]$.*

Proof. By the definition of appearance, there is a vertex r in $\mathfrak{G}$ such that $\mathfrak{G}[r/\text{Root}]$ is isomorphic to $\mathfrak{G}_{U(m)} + \mathfrak{G}'$ for some $\mathfrak{G}'$. Let i be such that $\mathfrak{G}_i$ with $\mathfrak{G}_i \equiv_m \mathfrak{G}'$ is the representative of the $\equiv_m$ equivalence class of $\mathfrak{G}'$ in the definition of $\mathfrak{G}_{U(m)}$. Then, by Prop. 4.2, $\mathfrak{G}[r/\text{Root}] \equiv_m \mathfrak{G}_{U(m)} + \mathfrak{G}_i$. Since, $\mathfrak{G}_{U(m)}$ contains more than $q_m(\emptyset)$ copies of $\mathfrak{G}_i$, Prop, 4.3 gives us that $\mathfrak{G}_{U(m)} \equiv_m \mathfrak{G}_{U(m)} + \mathfrak{G}_i$ and the result follows. □

5. MSO zero-one law for random connected graphs

Let $\mathcal{C}$ be a class of graphs. Recall that $\mathcal{C}_n$ is the class of graphs in $\mathcal{C}$ on the vertex set $[n]$ and $R_n(\mathcal{C})$ denotes the random graph drawn from the uniform distribution on $\mathcal{C}_n$. The following is a direct consequence of Theorem 1.1 in [CP16].

Theorem 5.1. *Let $\mathcal{C}$ be a small addable class of graphs. Then $R_n(\mathcal{C})$ is connected with probability at least $1/\sqrt{e} - o(1)$.*

Indeed, Theorem 1.1 of [CP16] establishes this more generally for *bridge addable* classes. The bound $1/\sqrt{e}$ is tight in that this is the limiting probability for connectedness among forests.

We also rely on the following result which is obtained as a consequence of Theorem 5.1 in [MSW06].

Theorem 5.2. *Let $\mathcal{C}$ be a small addable class of graphs, and $\mathfrak{G}$ be a rooted graph in $\mathcal{C}$. Then the probability that $\mathfrak{G}$ appears in $R_n(\mathcal{C})$ tends to 1 as $n \to \infty$.*

In fact, Theorem 5.1 of [MSW06] establishes the stronger result that the number of appearances of $\mathfrak{G}$ in $R_n(\mathcal{C})$ grows linearly with n (more precisely, there is an $\alpha > 0$ such that the probability that $\mathfrak{G}$ appears αn times tends to 1).

Together, these enable us to establish the zero-one law for the class of connected graphs in any small addable class $\mathcal{C}$. In the following we write Conn for the class of *connected* graphs.

Theorem 5.3. *Let ϕ be a sentence of MSO, and $\mathcal{C}$ a small addable class of graphs. Let p_n denote the probability that $R_n(\mathcal{C} \cap \mathrm{Conn})$ satisfies ϕ. Then $\lim_{n\to\infty} p_n$ is defined and equal to either 0 or 1.*

Proof. Let m be the quantifier rank of ϕ. By Theorem 5.2 we know that the probability that $\mathfrak{G}_{U(m)}$ appears in $R_n(\mathcal{C})$ tends to 1. Moreover, since the probability that $R_n(\mathcal{C})$ is connected is non-zero (by Theorem 5.1) it follows that the probability that $\mathfrak{G}_{U(m)}$, the m-universal connected rooted graph in $\mathcal{C}$, appears in $R_n(\mathcal{C} \cap \mathrm{Conn})$ also tends to 1. Thus, by Proposition 4.5, with probability tending to 1, we have $R_n(\mathcal{C} \cap \mathrm{Conn}) \equiv_m \mathfrak{G}_{U(m)}$. Hence, if $\mathfrak{G}_{U(m)} \models \phi$, then p_n tends to 1, otherwise p_n tends to 0. □

As an immediate consequence, we have a zero-one law for MSO for a number of interesting classes of graphs.

Corollary 5.4. *MSO admits a zero-one law on each of the following classes of graphs.*

- *The class of connected planar graphs.*
- *For each k, the class of connected graphs of tree-width at most k.*
- *For each $k > 2$, the class of connected graphs excluding K_k as a minor.*

6. MSO limit law for random graphs

We are now ready to establish our general result on the existence of a limit law (also known as a convergence law) for MSO on smooth addable classes of graphs. Note that, while Theorem 5.3 was stated for small addable classes, from now on we will restrict ourselves further to *smooth* classes. Recall that every smooth class is also small.

To establish the limit law we need two specific results from [MSW06]. The first is a direct consequence of Theorem 6.4 in that paper.

Theorem 6.1. *Let $\mathcal{C}$ be a smooth addable class of graphs. For any $\epsilon > 0$ there exist a constant $g(\epsilon)$ such that, for sufficiently large values of n, with probability at least $1-\epsilon$, $R_n(\mathcal{C})$ has a connected component which contains at least $n - g(\epsilon)$ vertices.*

In general, we refer to the *giant component* of $R_n(\mathcal{C})$.

The second result we need is a consequence of Theorem 9.2 of [MSW06].

Theorem 6.2. *Let $\mathcal{C}$ be a smooth addable class of graphs, H be a connected graph, and $k \in \mathbb{N}$. Then there is a number $p_k(H)$ such that the probability that $R_n(\mathcal{C})$ has exactly k components isomorphic to H tends to $p_k(H)$. Moreover, these events are asymptotically independent for non-isomorphic graphs H.*

Actually, Theorem 9.2 of [MSW06] does not state that the events are asymptotically independent, but it is easily seen to be the case. Indeed, what is stated there is that the distribution of the number of components isomorphic to H tends to Poisson distribution

as $n \to \infty$, with parameter $\lambda = 1/(\gamma(\mathcal{C})^{|H|}|\mathrm{Aut}(H)|)$, where $|\mathrm{Aut}(H)|$ is the number of automorphisms of H.

As an example, for large values of n, the random planar graph on n vertices has on average $1/\gamma(\mathcal{C}) \approx 0.03673$ isolated vertices, $1/2\gamma(\mathcal{C})^2 < 0.0007$ pairs of vertices with an edge between them (and no other edges incident on them), less than 0.00004 isolated connected subgraphs with 3 vertices, and so on. Summing over all of these, we can show that the random planar graph has $\lambda \approx 0.037439$ connected components other than the giant component on average. From the fact that the distribution is Poisson we get that the probability that the graph is not connected is $1 - e^{-\lambda} \approx 0.036746$.

We are interested in the frequency of occurrence of connected components, not just up to isomorphism, but up to $\equiv_m$ for suitable values of m. Specifically, we are interested in whether the number of components from a fixed $\equiv_m$ equivalence class is greater than the threshold $r_m(\emptyset)$ from Lemma 4.1 and the exact number if it is not. In the following result, we use Theorems 6.1 and 6.2 to show that the relevant probabilities converge.

Theorem 6.3. *Let $\mathcal{C}$ be a smooth addable class of graphs and $m \in \mathbb{N}$. If $\mathcal{A}$ is an $\equiv_m$-equivalence class of connected graphs then for each $k \leq r_m(\emptyset)$ there is a $p_k(\mathcal{A}) \in \mathbb{R}$ such that the probability that $R_n(\mathcal{C})$ has exactly k components from $\mathcal{A}$ tends to $p_k(\mathcal{A})$ as n goes to infinity.*

Proof. Let p_n denote the probability that $R_n(\mathcal{C})$ contains exactly k components in $\mathcal{A}$. We show that for any $\epsilon > 0$ there is a p such that $|p_n - p| < \epsilon$ for large enough n. Thus, the sequence $(p_n)_{n\in\mathbb{N}}$ is a Cauchy sequence and so converges to a limit.

Let $g = g(\epsilon/3)$ be the value given by Theorem 6.1 such that, for sufficiently large n, with probability at least $1 - \epsilon/3$, $R_n(\mathcal{C})$ has a connected component which contains at least $n - g$ vertices. Moreover, let $n_0 \in \mathbb{N}$ be such that for all $n > n_0$, this probability is indeed at least $1 - \epsilon/3$.

By exactly the same argument as in the proof of Theorem 5.3, we can show that the conditional probability, given that $R_n(\mathcal{C})$ has a connected component which contains at least $n - g$ vertices, that this giant component is $\equiv_m$-equivalent to $\mathfrak{G}_{U(m)}$, the m-universal connected rooted graph in $\mathcal{C}$, tends to 1 as n goes to infinity. Thus, we can fix a value $n_1 \in \mathbb{N}$ such that this probability is at least $1 - \epsilon/3$ for all $n > n_1$.

Let $H_1, \ldots, H_M$ enumerate (up to isomorphism) all graphs in $\mathcal{A}$ with at most g vertices. Let K denote the collection of all functions $f : [M] \to \mathbb{N}$ such that $\sum_{i\in[M]} f(i) = k$ and note that this is a finite set. Let $p = \sum_{f\in K} \prod_{i\in[M]} p_{f(i)}(H_i)$, where $p_{f(i)}(H_i)$ is the limiting probability, given by Theorem 6.2, that $R_n(\mathcal{C})$ contains exactly $f(i)$ components isomorphic to H_i. If p'_n denotes the probability that $R_n(\mathcal{C})$ contains exactly k components that are isomorphic to one of $H_1, \ldots, H_M$, then clearly the sequence $(p'_n)_{n\in\mathbb{N}}$ tends to the limit p. Let $n_2 \in \mathbb{N}$ be such that $|p'_n - p| < \epsilon/3$ for all $n > n_2$.

First, consider the case that $\mathfrak{G}_{U(m)} \notin \mathcal{A}$, i.e. $\mathcal{A}$ is an $\equiv_m$-equivalence class distinct from that of $\mathfrak{G}_{U(m)}$. In this case, our aim is to show that for all sufficiently large n, in particular for all $n > \max(n_0, n_1, n_2)$, we have $p - \epsilon < p_n < p + \epsilon$, establishing the result. We return to the case where $\mathfrak{G}_{U(m)} \in \mathcal{A}$ later.

Fix n with $n > \max(n_0, n_1, n_2)$ and let p_0 denote the probability that $R_n(\mathcal{C})$ contains *no* component from $\mathcal{A}$ except those that are isomorphic to one of $H_1, \ldots, H_M$. Then clearly $p_n > p'_n \cdot p_0$ since the left-hand side denotes the probability that there are exactly k components from $\mathcal{A}$ and the right-hand side gives the probability of obtaining exactly k components from $\mathcal{A}$ in a particular way, i.e. all from among $H_1, \ldots, H_M$. Moreover

$p_0 > (1-\epsilon/3)^2$ since if there is a giant component with $n-g$ elements and it is $\equiv_m$-equivalent to $\mathfrak{G}_{U(m)}$ then there cannot be a component with more than g vertices from $\mathcal{A}$. We then have

$$\begin{aligned} p_n &> p'_n p_0 \\ &> p'_n (1-\epsilon/3)^2 \\ &> p(1-\epsilon/3)^3 \\ &> p-\epsilon, \end{aligned}$$

where the second line follows by substituting the lower bound for p_0 and the third line follows from the fact that $|p'_n - p| < \epsilon/3$.

For the other direction, note that

$$p_n < p''_n p_0 + (1-p'_n)(1-p_0),$$

where the first term on the right is as above and the second term is an upper bound on the probability that k components from $\mathcal{A}$ are obtained in some way other than by having k components with g vertices or fewer and 0 components with more than g vertices. We then have

$$\begin{aligned} p_n &< p'_n p_0 + (1-p'_n)(1-p_0) \\ &< p'_n + (1-p_0) \\ &< p'_n + (1-(1-\epsilon/3)^2) \\ &< p + \epsilon/3 + (1-(1-\epsilon/3)^2) \\ &< p+\epsilon, \end{aligned}$$

where the second line is obtained by substituting the upper bound of 1 for p_0 and $1-p'_n$, the third line by substituting the lower bound of $(1-\epsilon/3)^2$ for p_0 and the fourth by the fact that $|p'_n - p| < \epsilon/3$.

For the case that $\mathfrak{G}_{U(m)} \in \mathcal{A}$, an entirely analogous argument can be used to show that $|p_n - p| < \epsilon$ where p is the limit of the sequence p''_n of probabilities that $R_n(\mathcal{C})$ contains exactly $k-1$ components that are isomorphic to one of $H_1, \ldots, H_M$. □

Fix a small addable class of graphs $\mathcal{C}$ and $m \in \mathbb{N}$. Let $r = r_m(\emptyset)$ and $t = t_m(\emptyset)$. Suppose $\mathcal{A}_1, \ldots, \mathcal{A}_t$ enumerates all the $\equiv_m$ classes of graphs in $\mathcal{C}$. We call an *m-profile* a function $f : [t] \to \{0, \ldots, r\}$. We say that a graph G *matches* the m-profile f if the following conditions hold:

(1) for each $i \in [t]$, if $f(i) < r$ then G has exactly $f(i)$ connected components which are in $\mathcal{A}_i$; and
(2) for each $i \in [t]$, if $f(i) = r$ then G has at least r distinct connected components which are in $\mathcal{A}_i$.

The following lemmas are now immediate from our previous constructions.

Lemma 6.4. *If $\mathcal{C}$ is a smooth addable class of graphs then for any m and any m-profile f, the probability that $R_n(\mathcal{C})$ matches f converges to a value p_f as n goes to infinity.*

Proof. Define p_f to be

$$p_f = \left(\prod_{i : f(i) < r} p_{f(i)}(\mathcal{A}_i) \right) \left(\prod_{i : f(i) = r} \sum_{k \geq r} p_k(\mathcal{A}_i) \right).$$

The result than follows by the asymptotic independence asserted in Theorem 6.2 above. □

Note that every graph matches some profile f and these events are mutually exclusive for distinct f. Thus, the sum of p_f over all profiles f must be 1. The reason for considering profiles is, of course, that they completely determine the $\equiv_m$ class of a graph.

Lemma 6.5. *If G_1 and G_2 are graphs that both match the same m-profile f, then $G_1 \equiv_m G_2$.*

Proof. This is immediate from Lemma 4.1. □

We can now establish the main convergence law for MSO.

Theorem 6.6. *If $\mathcal{C}$ is a smooth, addable class of graphs and ϕ is a sentence of MSO, then the probability that $R_n(\mathcal{C})$ satisfies ϕ tends to a limit as n goes to infinity.*

Proof. Let m be the quantifier rank of ϕ. By Lemma 6.5, if f is an m-profile then either all graphs matching f satisfy ϕ or none do. Let us say that f implies ϕ if the former case holds. Then taking p to be the sum of p_f (as in Lemma 6.4) over all f that imply ϕ, we see that the probability that $R_n(\mathcal{C})$ satisfies ϕ tends to p. □

We can say somewhat more about the possible values p to which the probability that ϕ holds in $R_n(\mathcal{C})$ may converge. Note that the property of being connected is definable by a sentence of MSO and thus the probability that $R_n(\mathcal{C})$ is connected converges to a limit. By Theorem 5.1, this value is at least $1/\sqrt{e}$ and therefore greater than $1/2$. This, together with the result below tells us that the limiting probabilities of MSO sentences on smooth, addable classes cluster near 0 and 1.

Theorem 6.7. *Let $\mathcal{C}$ be a smooth, addable class and let c be the limiting probability that $R_n(\mathcal{C})$ is not connected. Then, for any MSO sentence ϕ, the probability that $R_n(\mathcal{C})$ satisfies ϕ converges to a value p such that either $p \leq c$ or $p \geq 1 - c$.*

Proof. Let m be the quantifier rank of ϕ and let $\mathfrak{G}_{U(m)}$ be the universal connected structure defined in Sec. 4.1. If $\mathfrak{G}_{U(m)} \models \phi$ then ϕ is true in $R_n(\mathcal{C} \cap \mathrm{Conn})$ with probability tending to 1. Thus the probability that $R_n(\mathcal{C})$ satisfies ϕ tends to at least $1 - c$. On the other hand, if $\mathfrak{G}_{U(m)} \not\models \phi$, then ϕ is false in $R_n(\mathcal{C} \cap \mathrm{Conn})$ with probability tending to 1, and so the probability that $R_n(\mathcal{C})$ satisfies ϕ tends to at most c. □

For many interesting classes of graphs, the value of c is quite small. As noted above, for planar graphs the value of c is about 0.036746, giving us the result mentioned in the introduction.

7. Future work

We have shown that the zero-one law holds for random connected graphs from smooth addable classes of graphs, for formulas of first order and monadic second-order logic. Moreover, a limit law holds for random graphs of such classes which do not have to be connected. This includes many of the tame classes of graphs that have been studied in finite model theory in recent years. Still, there are other classes one could explore. Most interesting would be proper minor-closed classes which are not addable, such as the graphs embeddable in a torus or, more generally, the class of graphs of genus at most k for a fixed value of k.

Another general direction would be to explore logics beyond first-order logic (other than MSO) such as fixed-point logics. These are known to admit a zero-one law over the class of all graphs. However, their study is based on equivalence relations $\equiv^k$ stratified by the

number of variables rather than the quantifier rank. These equivalence relations do not have finite index and that makes many of the methods we have used here infeasible to use.

References

[BGK07] Manuel Bodirsky, Clemens Gröpl, and Mihyun Kang. Generating labeled planar graphs uniformly at random. *Theor. Comput. Sci.*, 379:377–386, July 2007.

[Com89] K.J. Compton. 0-1 laws in logic and combinatorics. In I. Rival, editor, *NATO Advanced Study Institute on Algorithms and Order*, pages 353–383. Kluwer, 1989.

[CP16] Guillaume Chapuy and Guillem Perarnau. Connectivity in bridge-addable graph classes: The McDiarmid-Steger-Welsh conjecture. In *Proceedings of the Twenty-seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '16, pages 1580–1588, 2016.

[Daw07] A. Dawar. Finite model theory on tame classes of structures. In *MFCS*, volume 4708 of *Lecture Notes in Computer Science*, pages 2–12. Springer, 2007.

[DG10] A. Dawar and E. Grädel. Properties of almost all graphs and generalized quantifiers. *Fundam. Inform.*, 98(4):351–372, 2010.

[DVW96] Alain Denise, Marcio Vasconcellos, and Dominic J. A. Welsh. The random planar graph. *Congressus Numerantium*, 113:61–79, 1996.

[Fag76] R. Fagin. Probabilities on finite models. *Journal of Symbolic Logic*, 41(1):50–58, March 1976.

[GKLT69] Y. V. Glebskiĭ, D. I. Kogan, M.I. Ligon'kiĭ, and V. A. Talanov. Range and degree of realizability of formulas in the restricted predicate calculus. *Kibernetika*, 2:17–28, 1969.

[GN05] Omer Giménez and Marc Noy. Asymptotic enumeration and limit laws of planar graphs. *Mathematics of Computation*, 2005.

[HMNT18] P. Heinig, T. Müller, M. Noy, and A. Taraz. Logical limit laws for minor-closed classes of graphs. *J. Comb. Theory, Ser. B*, 130:158–206, 2018.

[KV90] Ph. G. Kolaitis and M. Y. Vardi. 0-1 laws and decision problems for second-order logic. *Information and Computation*, 87:302–338, 1990.

[KV92] Ph. G. Kolaitis and M. Y. Vardi. Infinitary logics and 0-1 laws. *Information and Computation*, 98(2):258–294, 1992.

[Lib04] L. Libkin. *Elements of Finite Model Theory*. Springer, 2004.

[Mak04] J. A. Makowsky. Algorithmic uses of the feferman-vaught theorem. *Ann. Pure Appl. Logic*, 126:159–213, 2004.

[McC02] Gregory L. McColm. MSO zero-one laws on random labelled acyclic graphs. *Discrete Mathematics*, 254:331–347, 2002.

[McD09] Colin McDiarmid. Random graphs from a minor-closed class. *Comb. Probab. Comput.*, 18:583–599, July 2009.

[MSW05] Colin McDiarmid, Angelika Steger, and Dominic J.A. Welsh. Random planar graphs. *Journal of Combinatorial Theory, Series B*, 93(2):187 – 205, 2005.

[MSW06] Colin McDiarmid, Angelika Steger, and Dominic Welsh. Random graphs from planar and other addable classes. In R. L. Graham, B. Korte, L. Lovász, A. Wigderson, G. M. Ziegler, Martin Klazar, Jan Kratochvíl, Martin Loebl, Jiří Matoušek, Pavel Valtr, and Robin Thomas, editors, *Topics in Discrete Mathematics*, volume 26 of *Algorithms and Combinatorics*, pages 231–246. Springer Berlin Heidelberg, 2006. 10.1007/3-540-33700-8_15.