First-Order Logic with Equicardinality in Random Graphs

Simi Haber
Bar-Ilan University, Israel

Tal Hershko
California Institute of Technology, US

Mostafa Mirabi
Taft School, US

Saharon Shelah
Hebrew University of Jerusalem, Israel

Abstract

We answer a question of Blass and Harary about the validity of the zero-one law in random graphs for extensions of first-order logic (FOL). For a given graph property P, the *Lindström extension* of FOL by P is defined as the minimal (regular) extension of FOL able to express P. For several graph properties P (e.g. Hamiltonicity), it is known that the Lindström extension by P is also able to interpret a segment of arithmetic, and thus strongly disobeys the zero-one law. Common to all these properties is the ability to express the Härtig quantifier, a natural extension of FOL testing if two definable sets are of the same size. We prove that the Härtig quantifier is sufficient for the interpretation of arithmetic, thus providing a general result which implies all known cases of Lindström extensions which are able to interpret a segment of arithmetic.

2012 ACM Subject Classification Theory of computation → Finite Model Theory; Mathematics of computing → Random graphs

Keywords and phrases finite model theory, first-order logic, monadic second-order logic, random graphs, zero-one laws, generalized quantifiers, equicardinality

Digital Object Identifier 10.4230/LIPIcs.CSL.2025.12

Funding *Saharon Shelah*: Research partially supported by the grant “Independent Theories” NSF-BSF, (BSF 3013005232) and by NSF grant no: DMS 1833363. Paper 1250 on Shelah’s publication list.

1 Introduction

In this paper we study the Erdös-Rényi binomial random graph model $G(n,p)$. Recall that $G(n,p)$ is defined as a probability distribution over the set of all labeled (simple) graphs with the vertex set $[n] := \{1, 2, \ldots, n\}$, by requiring that each of the $\binom{n}{2}$ potential edges appears with probability p and independently of all other edges. Note that $G\left(n, \frac{1}{2}\right)$ is the uniform distribution over the set of $2^{\binom{n}{2}}$ labeled graphs with vertex set $[n]$.

In what follows, we use $\mathbb{P}(\cdot)$ to denote probabilities and $\mathbb{E}(\cdot)$ to denote expected values.

1.1 Background and Previous Results

The study of random graphs was pioneered by Erdös and Rényi in the 1960s, originating from two seminal papers [9, 10]. One of the earliest phenomena recognized in their work is the fact that many natural graph properties — including connectivity, Hamiltonicity, planarity, k-colorability for a fixed k and containing H as a subgraph for a fixed graph H — hold either in almost all graphs, or in almost none of them. Formally, let P be a graph property and fix

33rd EACSL Annual Conference on Computer Science Logic (CSL 2025).
Editors: Jörg Endrullis and Sylvain Schmitz; Article No. 12; pp. 12:1–12:17
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

$p \in (0,1)$. We say that P holds *asymptotically almost surely* (a.a.s. for short) in $G(n,p)$ if

$$\lim_{n\to\infty} \mathbb{P}\left(G(n,p) \text{ satisfies } P\right) = 1$$

and that P holds *asymptotically almost never* (a.a.n. for short) in $G(n,p)$ if

$$\lim_{n\to\infty} \mathbb{P}\left(G(n,p) \text{ satisfies } P\right) = 0.$$

Then, for example, for every fixed $p \in (0,1)$,

- Connectivity holds a.a.s. in $G(n,p)$.
- Hamiltonicity holds a.a.s. in $G(n,p)$.
- Planarity holds a.a.n. in $G(n,p)$.
- For every fixed $k \in \mathbb{N}$, k-colorability holds a.a.n. in $G(n,p)$.
- For every fixed finite graph H, the property of containing H as a subgraph holds a.a.s. in $G(n,p)$.

As a reference, see any introductory text in random graphs, e.g. [19].

The observation that many natural graph properties hold either a.a.s. or a.a.n. in $G(n,p)$ motivates the following definition.

▶ **Definition 1.** *Let $\mathcal{A}$ be a set of graph properties. We say that $\mathcal{A}$* obeys the zero-one law *in $G(n,p)$ if for every property $P \in \mathcal{A}$,*

$$\lim_{n\to\infty} \mathbb{P}\left(G(n,p) \text{ satisfies } P\right) \in \{0,1\}.$$

With this definition, we may formulate the following informal observation: if $\mathcal{A}$ is a set of natural graph properties then $\mathcal{A}$ obeys the zero-one law. This is not a formal statement, due to the lack of a formal definition of a "natural graph property".

From a logician's point of view, a natural class of graph properties is the class $\mathcal{FO}$ of *first-order properties.* These are properties which can be expressed as a sentence in the first-order language of graphs, whose signature consists of a single binary relation $\sim$ representing adjacency.[1] Indeed, a classic result, proven independently by Glebskii et al. [13] and Fagin [11], states that $\mathcal{FO}$ obeys the zero-one law in $G(n,p)$.

▶ **Theorem 2** (GKLT-Fagin). *Fix $p \in (0,1)$. Then the set of first-order graph properties $\mathcal{FO}$ obeys the zero-one law in $G(n,p)$.*

The GKLT-Fagin zero-one law pioneered the study of random graphs with tools of mathematical logic. This point of view has proved to be doubly beneficial, teaching us about the properties of the underlying random graph, and also about the expressive power of logical languages. It is therefore considered an important part of finite model theory.

The GKLT-Fagin zero-one law deals with first-order properties. However, many graph properties which are considered natural — including connectivity, Hamiltonicity and k-colorability — are not first-order. On the other extreme, the class $\mathcal{SO}$ of *second-order* graph properties contains all the properties listed above, but fails to obey the zero-one law. For example, as noted by Fagin [11, p. 55], the property of having an even number of vertices is second-order, but clearly has no limiting probability.

It is therefore natural to ask for extensions of first-order logic which have a stronger expressive power on the one hand, but still obey the zero-one law on the other hand. This question was posed by Blass and Harary [2, Section 5]. In their discussion, they suggest several guiding questions:

[1] We shall often identify logical sentences with the properties they describe.

1. Is there an extension of first order logic which is strong enough to express Hamiltonicity and rigidity (asymmetry), but still obeys the zero-one law?
2. What about monadic second-order logic? It cannot express Hamiltonicity, but it is still an important extension of first-order logic — does it obey the zero-one law?
3. Can something be done with "more exotic languages", for example with equicardinality quantifiers?

These questions have been studied in many papers. We dedicate the following sections to explain the precise meaning of these questions in more detail and review previous results. Our review is not exhaustive; for a broader survey of results in this field, we refer the reader to [4] and [29]. For a more focused discussion of results directly related to our work, see [6] and [16].

Lindström Extensions

Suppose we are interested in an extension of $\mathcal{FO}$ which includes a certain graph property P. One option is to simply take the union $\mathcal{FO} \cup \{P\}$. However, this set of properties clearly lacks a basic notion of closure. To avoid such trivialities, we focus on *regular* extensions. A regular logic is a logic that is closed under negation, conjunction, existential quantification, relativization and substitution (see [7] for more details). We then have the following definition.

▶ **Definition 3.** *Let P be a graph property. The* Lindström extension *of $\mathcal{FO}$ by P, denoted $\mathcal{FO}[P]$, is the minimal regular extension of $\mathcal{FO}$ that includes P.*

The term *Lindström extension* comes from the fact that $\mathcal{FO}[P]$ can be constructed by adjoining the Lindström quantifier of P, denoted Q_P and defined as follows [26, 27, 7].

▶ **Definition 4.** *Let P be a graph property. Its Lindström quantifier Q_P is defined as follows.*

- *Syntactically, given a formula $\varphi_V(x, \vec{z})$ with $x, \vec{z}$ as free variables and a formula $\varphi_E(x, y, \vec{z})$ with $x, y, \vec{z}$ as free variables, it returns the formula $Q_P x, y\, (\varphi_V(x, \vec{z}), \varphi_E(x, y, \vec{z}))$ in which x, y are quantified and $\vec{z}$ are free.*
- *Semantically, the truth value of this formula is defined as follows. Let $G = (V, E)$ be a graph and let $\vec{a}$ be a vector of vertices, of the same length as $\vec{z}$. Let $V_0 = \{v \in V : G \models \varphi_V(v, \vec{a})\}$ and let E_0 be the set of pairs $\{u, v\}$ with $u, v \in V_0$ such that $G \models \varphi(u, v, \vec{a})$ or $G \models \varphi(v, u, \vec{a})$. Then*

$$G \models_{\vec{z}=\vec{a}} Q_P x, y\, (\varphi_V(x, \vec{z}), \varphi_E(x, y, \vec{z})) \iff G_0 = (V_0, E_0) \text{ satisfies } P.$$

It can be shown that $\mathcal{FO}[P]$ is the same as the closure of $\mathcal{FO}$ under quantification with Q_P (see [8] for more details).

We can now suggest a more precise formulation of the first question of Blass and Harary: do the Lindström extensions of $\mathcal{FO}$ by Hamiltonicity and by rigidity obey the zero-one law? The answer was given by Dawar and Grädel ([6], also in [5]).

▶ **Theorem 5** (Dawar-Grädel). *Fix $p \in (0, 1)$.*

1. *The Lindström extension $\mathcal{FO}[\text{Rigidity}]$ obeys the zero-one law in $G(n, p)$.*
2. *The Lindström extension $\mathcal{FO}[\text{Hamiltonicity}]$ does not obey the zero-one law in $G(n, p)$.*

The latter part of the theorem was demonstrated by encoding Parity—the property of having an even number of vertices. It is worth noting, however, that Parity still allows for the possibility of a modular limit law, as shown by Kolaitis and Kopparty [25]. A far more extreme violation of the zero-one law occurs through the interpretation of a segment of arithmetic, a concept we will elucidate in the subsequent section.

Arithmetization

The second question of Blass and Harary, regarding monadic second order logic $\mathcal{MSO}$, was answered much earlier than the first. The answer was given by Kaufmann and Shelah [23], who proved that $\mathcal{MSO}$ disobeys the zero-one law in a very strong sense. Their main result is that *$\mathcal{MSO}$ can interpret arithmetic* in $G(n, p)$, which roughly means that there are sentences in $\mathcal{MSO}$ that define an arithmetic structure on (a subset of) the vertex set. *If a language $\mathcal{L}$ can interpret a segment of arithmetic then it is, in a sense, the farthest possible from obeying the zero-one law.* Indeed, in such a case, it is possible to construct sentences $\varphi \in \mathcal{L}$ whose probability sequence $\{\mathbb{P}\left(G(n,p) \models \varphi\right)\}_{n=1}^{\infty}$ exhibits different kinds of complex behaviors. We shall demonstrate this fact shortly by constructing a sentence $\varphi \in \mathcal{L}$ whose probability sequence alternates between near-zero values and near-one values, and hence has no limit.

The definition of interpreting arithmetic given below is somewhat weaker than what Kaufmann and Shelah prove for $\mathcal{MSO}$, but describes a more general and more common type of arithmetization results (see review below). In particular, it includes the main result of this paper, Theorem 14).

Recall that for $n \in \mathbb{N}$ we denote $[n] = \{1, 2, \ldots, n\}$. We begin by defining the language $\mathcal{SO}[\mathrm{Arith}]$ as the second-order language of arithmetic, where:

- Addition $+$ and multiplication $\times$ are defined as ternary relations (so, for example, we write $+(a, b, c)$ instead of $a + b = c$). This is a convenient choice when working with finite models such as $[n]$, where addition and multiplication are restricted.
- Second order quantification is done only over unary and binary relations.

▶ **Example 6.** We can construct a formula in $\mathcal{SO}[\mathrm{Arith}]$ with free variables x, y expressing the property $y = 2^x$ by the following steps.

- Begin by asserting the existence of a binary relation: $\exists \mathrm{Exp}^2$.
- Require that it is single-valued:

$$\forall a \forall b \forall b' \left(\mathrm{Exp}(a, b) \wedge \mathrm{Exp}(a, b') \to b = b'\right).$$

- Define the relation inductively:

$$\mathrm{Exp}(0, 1) \wedge \forall a \forall b \forall a' \forall b' \left(\mathrm{Exp}(a, b) \wedge +(a, 1, a') \wedge \times(b, 2, b') \to \mathrm{Exp}(a', b')\right).$$

- Finally, require $\mathrm{Exp}(x, y)$.

For every $n \in \mathbb{N}$, the set $[n]$ admits an arithmetic structure with the standard addition and multiplication (restricted to $[n]$). Given an interval $[a, b] \subseteq \mathbb{R}$ and a sentence $\varphi \in \mathcal{SO}[\mathrm{Arith}]$, we say that:

1. φ *holds in* $[a, b]$ if $[n] \models \varphi$ for every $n \in [a, b] \cap \mathbb{N}$.
2. φ is *constant on* $[a, b]$ if φ holds in $[a, b]$ or $\neg\varphi$ holds in $[a, b]$.

Finally, let us say that a function $f : \mathbb{N} \to \mathbb{N}$ is *finite-to-one* if for every $m \in \mathbb{N}$, the inverse image $f^{-1}(m)$ is a non-empty finite set. Note that if f is finite-to-one then f is onto and $\lim_{n\to\infty} f(n) = \infty$. Examples include $f(n) = \lfloor \sqrt{n} \rfloor$ and $f(n) = \lfloor \log \log n \rfloor$.

▶ **Definition 7** (Arithmetization). *Let $\mathcal{L}$ be a logical language whose signature includes the binary relation symbol $\sim$, representing adjacency. Fix a parameter $p \in (0, 1)$, constants $0 < c_1 \leq c_2$ and a finite-to-one function $f : \mathbb{N} \to \mathbb{N}$. We say that $\mathcal{L}$* can interpret a segment of arithmetic *in $G(n, p)$ with constants c_1, c_2 and a scaling function $f(n)$ if the following holds. For every sentence $\varphi \in \mathcal{SO}[\mathrm{Arith}]$ there exists a sentence $\varphi^* \in \mathcal{L}$ such that, given* *a sequence* $\{n_k\}_{k=1}^{\infty}$ *with* $\lim_{k\to\infty} n_k = \infty$ *such that* φ *is constant on* $[c_1 f(n_k), c_2 f(n_k)]$ *for every* k*, we have*

$$\lim_{k\to\infty} \mathbb{P}\left(G(n_k,p) \models \varphi^* \iff \varphi \text{ holds in } [c_1 f(n_k), c_2 f(n_k)]\right) = 1.$$

To motivate the definition, we remark that it reflects a general strategy of encoding arithmetic in random graphs, which can be roughly summarized as follows:

- Restrict to a certain $\mathcal{L}$-definable subset $S \subseteq [n]$ of size $|S| \in [c_1 f(n), c_2 f(n)]$.
- Use the structure of the random graph to encode unary and binary relations and an arithmetic structure on S.
- Given a sentence $\varphi \in \mathcal{SO}[\text{Arith}]$, use the encoded structure to convert it into a sentence $\varphi^* \in \mathcal{L}$ asserting that φ is satisfied in S.

▶ **Proposition 8.** *Let* $\mathcal{L}$ *be a language that can interpret a segment of arithmetic in* $G(n,p)$*. Then there exists a sentence* $\psi \in \mathcal{L}$ *such that the limit* $\lim_{n\to\infty} \mathbb{P}\left(G(n,p) \models \psi\right)$ *does not exist. In particular,* $\mathcal{L}$ *disobeys the zero-one law.*

Proof. Let $0 < c_1 \leq c_2$ and $f(n)$ be the constants and the scaling function for $\mathcal{L}$, as in Definition 7. Let $N = \lceil \max\{|\log_2 c_1|, |\log_2 c_2|\}\rceil + 1$. It is straightforward to construct a sentence $\varphi \in \mathcal{SO}[\text{Arith}]$ such that for every $m \in \mathbb{N}$, $[m] \models \varphi$ if and only if

$$\lfloor \log_2 m \rfloor \equiv k \mod 8N \quad \text{for } k \in \{-N, -N+1, \ldots, N-1, N\}.$$

Let $\{n_k\}_{k=1}^{\infty}$ be a sequence satisfying $f(n_k) = 2^{4Nk}$ for every k. Such a sequence exists and approaches ∞, because f is finite-to-one. We claim that φ is constant on each interval $[c_1 f(n_k), c_2 f(n_k)]$. Indeed, for every $m \in [c_1 f(n_k), c_2 f(n_k)] \cap \mathbb{N}$ we have

$$\begin{aligned} \log_2 m &\in [\log_2 c_1 + 4Nk, \log_2 c_2 + 4Nk], \\ \lfloor \log_2 m \rfloor &\in [4Nk - N, 4Nk + N] \end{aligned} \tag{1}$$

where (1) follows from our choice of N. When k is even, (1) implies $[m] \models \varphi$. When k is odd, (1) implies $[m] \not\models \varphi$.

Now let $\varphi^* \in \mathcal{L}$ as in Definition 7. Then the probabilities sequence $\{\mathbb{P}\left(G(n_k,p) \models \varphi^*\right)\}_{k=1}^{\infty}$ converges to 1 on even values of k and converges to 0 on odd values of k. This implies that the limit $\lim_{n\to\infty} \mathbb{P}\left(G(n,p) \models \varphi^*\right)$ does not exist. ◀

Going back to Kaufmann and Shelah, we can now formulate the following consequence of [23] (which follows from Theorem 1 and the closing remark).

▶ **Theorem 9** (Kaufmann-Shelah). *Fix* $p \in (0,1)$*. Then* $\mathcal{MSO}$ *can interpret a segment of arithmetic in* $G(n,p)$ *(with constants* $c_1 = c_2 = 1$ *and scaling function* $f(n) = \lfloor \sqrt{n} \rfloor$*).*

As explained, this result provides a strongly negative answer to the second question of Blass and Harary.

In [16], Haber and Shelah prove an arithmetization result for the Lindström extension of Hamiltonicity, thus strengthening Part 2 of Theorem 5.

▶ **Theorem 10** (Haber-Shelah). *Fix* $p \in (0,1)$*. Then* $\mathcal{FO}[\text{Hamiltonicity}]$ *can interpret a segment of arithmetic in* $G(n,p)$ *(with scaling function* $f(n) = \Omega(\log\log\log n)$*).*

As for other graph properties, Haber and Shelah also proved in [16] that the zero-one law holds for the Lindström extensions $\mathcal{FO}[\text{Connectivity}]$ and $\mathcal{FO}[k\text{−colorability}]$ for every fixed k. These results also follow from a more general theorem by Dawar and Grädel [6], which also implies that the zero-one law holds for $\mathcal{FO}$[Planarity]. On the other hand, there are additional graph properties P for which it is known that $\mathcal{FO}[P]$ can interpret a segment of arithmetic. These include regularity, the existence of a perfect matching [15] and the existence of a C_4-factor [14]. It is noteworthy that Haber and Shelah [16] employed a strategy to encode the Rescher plurality quantifier [28], resulting in a more expressive logic. In contrast, our finding that equicardinality alone suffices to interpret a segment of arithmetic is unexpected and significantly stronger.

Equicardinality Quantifiers

Common to all the Lindström extensions of $\mathcal{FO}$ which are known to be able to interpret a segment of arithmetic is the ability to express the *equicardinality quantifier*, also known as the *Härtig quantifier* [17], which we denote by $Q_=$. This quantifier allows for testing if two definable sets are of the same size.

▶ **Definition 11.** *The Härtig quantifier $Q_=$ is defined as follows.*

- *Syntactically, given formulas $\varphi(x, \vec{z})$ and $\psi(x, \vec{z})$ with free variables $x, \vec{z}$, it returns a formula $Q_= x\,(\varphi(x, \vec{z}), \psi(x, \vec{z}))$ in which x is quantified and $\vec{z}$ are free.*
- *Semantically, the truth value of this formula is defined as follows. Let $G = (V, E)$ be a finite graph and let $\vec{a}$ be a vector of vertices, of the same length as $\vec{z}$. Then*

$$G \models_{\vec{z}=\vec{a}} Q_= x\,(\varphi(x, \vec{z}), \psi(x, \vec{z})) \iff |\{v \in V : G \models \varphi(v, \vec{a})\}| = |\{v \in V : G \models \psi(v, \vec{a})\}|\,.$$

The following proposition shows that $Q_=$ is indeed expressible in $\mathcal{FO}$[Hamiltonicity]. Similar arguments show that $Q_=$ is also expressible in the other Lindström extensions of $\mathcal{FO}$ listed above that are known to be able to interpret a segment of arithmetic.

▶ **Proposition 12.** *Let $\varphi(x, \vec{z})$ and $\psi(x, \vec{z})$ be formulas in $\mathcal{FO}$[Hamiltonicity] with free variables $x, \vec{z}$. Then the formula $Q_= x\,(\varphi(x, \vec{z}), \psi(x, \vec{z}))$ is also expressible in $\mathcal{FO}$[Hamiltonicity].*

Proof. Fix a graph $G = (V, E)$ and a vector $\vec{a}$ of vertices. Let $A = \{x \in V : G \models \varphi(x, \vec{a})\}$ and $B = \{x \in V : G \models \psi(x, \vec{a})\}$. Also let $\varphi'(x, \vec{z}) = \varphi(x, \vec{z}) \wedge \neg\psi(x, \vec{z})$ (which defines the set $A \setminus B$) and $\psi'(x, \vec{z}) = \psi(x, \vec{z}) \wedge \neg\varphi(x, \vec{z})$ (which defines the set $B \setminus A$). Define $\varphi_V(x, \vec{z}) = \varphi'(x, \vec{z}) \vee \psi'(x, \vec{z})$ and $\varphi_E(x, y, \vec{z}) = \varphi'(x, \vec{z}) \wedge \psi'(y, \vec{z})$. Consider the graph $G_0 = (V_0, E_0)$ defined by these formulas, as in Definition 4. Note that G_0 is the complete bipartite graph with sides $A \setminus B$ and $B \setminus A$. Recall that a complete bipartite graph is Hamiltonian if and only if its sides are of the same size. Therefore

$$\begin{aligned} G \models_{\vec{z}=\vec{a}} Q_{\mathrm{Ham}} x, y\,(\varphi_V(x, \vec{z}), \varphi_E(x, y, \vec{z})) &\iff |A \setminus B| = |B \setminus A| \\ &\iff |A| = |B| \iff G \models_{\vec{z}=\vec{a}} Q_= x\,(\varphi(x, \vec{z}), \psi(x, \vec{z}))\,. \end{aligned}$$

◀

Let $\mathcal{FO}[Q_=]$ denote the closure of $\mathcal{FO}$ under quantification with $Q_=$. This is a natural extension of first-order logic which has been studied quite extensively. For a survey on equicardinality quantifiers in the context of general model theory and abstract logic, see [18]. Equicardinality quantifiers have also been studied in the context of zero-one laws and convergence laws. In [12], Fayolle, Grumbach and Tollu studied zero-one laws for first-order logic enriched by generalized quantifiers, including Härtig quantifiers (equicardinality quantifiers) and Rescher quantifiers (expressing inequalities of cardinalities). Their results show that the zero-one law *does* hold for $\mathcal{FO}^*[Q_=]$, defined in the same way as $\mathcal{FO}[Q_=]$ but with the restriction that free variables are not allowed in the scope of $Q_=$-quantification.

▶ Remark 13. A simple argument, also mentioned in [12], proves that $\mathcal{FO}[Q_=]$ can express parity in the special case of $G\left(n, \frac{1}{2}\right)$. Indeed, consider the sentence $\exists z Q_= x\,(x \sim z, \neg(x \sim z))$, asserting the existence of a vertex with the same number of neighbors as non-neighbors. This sentence holds in $G\left(n, \frac{1}{2}\right)$ with probability 0 when n is even, and with probability approaching 1 when n is odd. In particular, $\mathcal{FO}[Q_=]$ does not satisfy the zero-one law in $G\left(n, \frac{1}{2}\right)$. As we shall soon see, much more can be said: $\mathcal{FO}[Q_=]$ can express not only parity, but a segment of arithmetic, and for every $p \in (0,1)$ (Theorem 14).

Finally, we mention that in addition to Lindström quantifiers and equicardinality quantifiers, other generalized quantifiers have also been studied in the context of zero-one laws and convergence laws. In a sequence of three papers [20, 21, 22], Kalia studied almost sure quantifier elimination, providing a method for proving convergence laws for logics with generalized quantifiers. In [24], Keisler and Lotfallah proved almost sure quantifier elimination logics with probability quantifiers.

1.2 Our Results

The main result of the paper is the following theorem.

▶ **Theorem 14** (Main Theorem). *Fix $p \in (0,1)$. Then $\mathcal{FO}[Q_=]$ can interpret a segment of arithmetic in $G(n,p)$ (with scaling function $f(n) = \left\lfloor \sqrt{\ln n} \right\rfloor$).*

In particular, from Proposition 8 we get the following corollary.

▶ **Corollary 15.** *Fix $p \in (0,1)$. Then there exists a sentence $\psi \in \mathcal{FO}[Q_=]$ such that the limit $\lim_{n\to\infty} \mathbb{P}\left(G(n,p) \models \psi\right)$ does not exist.*

This answers the third question of Blass and Harary [2, Section 5], and provides a general result which immediately implies all known cases of Lindström extensions of $\mathcal{FO}$ which are able to express arithmetic. As mentioned above, these include the Lindström quantifier of Hamiltonicity, regularity, the existence of a perfect matching and the existence of a C_4-factor.

The rest of the paper is dedicated to the proof of Theorem 14. The proof strategy is roughly as follows. Given a sentence $\varphi \in \mathcal{SO}[\text{Arith}]$, first apply Theorem 9 to convert it into a sentence $\varphi^* \in \mathcal{MSO}$ which expresses φ on a set of size $\lfloor \sqrt{n} \rfloor$. Then, the crux of the proof is to convert a sentence $\varphi^* \in \mathcal{MSO}$ into a sentence $\psi \in \mathcal{FO}[Q_=]$ which expresses φ^* on a set of size $\Theta(\ln n)$. To do that, we need to show that $\mathcal{FO}[Q_=]$ can define subsets of logarithmic size, and can also interpret monadic second-order logic on such sets. The proof is divided between Sections 2 and 3. In Section 2 we develop the necessary probabilistic tools, and in Section 3 we put them together in order to complete the proof.

Notation and Conventions

We denote $\mathcal{FO}_= := \mathcal{FO}[Q_=]$ for short. Given a list of variable symbols $x_1, \ldots, x_n$, let $\mathcal{FO}(x_1, \ldots, x_n)$ denote the set of first-order formulas (in the language of graphs) with $x_1, \ldots, x_n$ as free variables. Similarly define $\mathcal{FO}_=(x_1, \ldots, x_n)$ and $\mathcal{MSO}(x_1, \ldots, x_n)$.

Throughout the text we maintain the convention of denoting random variables with a boldface font.

For $n \in \mathbb{N}$ and $p \in (0,1)$, we write $\mathbf{G}_n \sim G(n,p)$ to indicate that $\mathbf{G}_n$ is a random graph with distribution $G(n,p)$. For two vertices $u, v \in [n]$, let $u \sim v$ denote that they are adjacent in $\mathbf{G}_n$. For a subset $S \subseteq [n]$, let $\mathbf{G}_n[S]$ denote the subgraph of $\mathbf{G}_n$ induced by S.

We shall use the following notions of asymptotic probabilities. Let $(E_n)_{n=1}^{\infty}$ be a sequence of events, taken from a sequence of probability spaces.

1. We say that E_n holds *with high probability* (as $n \to \infty$) if

$$\mathbb{P}(E_n) = 1 - o(1).$$

2. We say that E_n holds *with exponentially high probability* (as $n \to \infty$) if

$$\mathbb{P}(E_n) = 1 - \exp\left(-n^{\Omega(1)}\right).$$

In addition, let $(\mathbf{X}_n)_{n=1}^{\infty}$, $(\mathbf{Y}_n)_{n=1}^{\infty}$ be two sequences of positive random variables. We say that $\mathbf{X}_n = (1+o(1))\,\mathbf{Y}_n$ with (exponentially) high probability if there exists a sequence $\varepsilon_n = o(1)$ such that the event $|\mathbf{X}_n/\mathbf{Y}_n - 1| \leq \varepsilon_n$ holds with (exponentially) high probability.

For notational convenience, we sometimes omit dependency on n from our notation. The underlying assumption throughout the text is that all quantities implicitly depend on n (unless it is explicitly stated that they are constant or fixed) and $n \to \infty$. We explicitly refer to the dependency on n in cases where this convention may cause ambiguity.

Finally, recall the following tail bounds on binomial and Poisson variables, following from Chernoff's inequality (e.g. see [1, Appendix A]).

- Let $\mathbf{X} \sim \mathrm{Bin}(n,p)$ and $\mu = \mathbb{E}\mathbf{X}$. Then for every $0 < \delta < 1$,

$$\mathbb{P}(|\mathbf{X} - \mu| \geq \delta\mu) \leq 2\exp\left(-\frac{\delta^2}{3}\mu\right). \tag{2}$$

- Let $\mathbf{X} \sim \mathrm{Pois}(\lambda)$ and $\mu = \mathbb{E}\mathbf{X}$. Then for every $0 < \delta < 1$,

$$\mathbb{P}(|\mathbf{X} - \mu| \geq \delta\mu) \leq 2\exp\left(-\frac{\delta^2}{4}\mu\right). \tag{3}$$

2 Some Probabilistic Results

From now fix a constant $p \in (0,1)$ and consider a binomially distributed random graph $\mathbf{G}_n \sim G(n,p)$.

We begin by fixing, for every n, two arbitrary vertices $u_1, u_2 \in [n]$. Let $V' = [n] \setminus \{u_1, u_2\}$. Define the following (random) vertex sets:

$$\begin{aligned}
\mathbf{A} &= \{v \in V' : v \sim u_1 \wedge v \sim u_2\}, \\
\mathbf{B} &= \{v \in V' : v \sim u_1 \wedge v \not\sim u_2\}, \\
\mathbf{C} &= \{v \in V' : v \not\sim u_1 \wedge v \sim u_2\}.
\end{aligned}$$

Note that the statements $v \in \mathbf{A}, v \in \mathbf{B}, v \in \mathbf{C}$ are all expressible as formulas in $\mathcal{FO}(u_1, u_2, v)$.

From (2), with exponentially high probability we have

$$\begin{aligned}
|\mathbf{A}| &= (1+o(1))p^2 n, \\
|\mathbf{B}|, |\mathbf{C}| &= (1+o(1))p(1-p)n.
\end{aligned}$$

That is, there exists a sequence $\delta_n = o(1)$ such that the event (which we denote by $\mathcal{Q}$)

$$\frac{|\mathbf{A}|}{p^2 n}, \frac{|\mathbf{B}|}{p(1-p)n}, \frac{|\mathbf{C}|}{p(1-p)n} \in [1-\delta_n, 1+\delta_n] \tag{4}$$

holds with exponentially high probability. It will be convenient to condition on the values of the variables $\mathbf{A}, \mathbf{B}, \mathbf{C}$; that is, to condition on an event of the form $Q_{A,B,C} = \{\mathbf{A} = A, \mathbf{B} = B, \mathbf{C} = C\}$ where A, B, C are possible values of $\mathbf{A}, \mathbf{B}, \mathbf{C}$. Note that conditioning on $Q_{A,B,C}$ does not affect the distribution of $\mathbf{G}_n[V']$.

The rest of the section is as follows. In Section 2.1 we show how to define sets $S \subseteq [n]$ of logarithmic size in $\mathcal{FO}[Q_=]$. In Section 2.2 we show how to express unary relations (subsets) on such sets S in $\mathcal{FO}[Q_=]$. In both sections, all the probabilities and expected values are assumed to be conditioned on $Q_{A,B,C}$, where we assume that A, B, C satisfy Equation (4) (that is, we assume $Q_{A,B,C} \subseteq \mathcal{Q}$). Finally, in Section 2.3 we apply the law of total probability to obtain non-conditioned results.

2.1 Defining Sets of Logarithmic Size

Recall that A, B are the fixed values of the random sets $\mathbf{A}$, $\mathbf{B}$ defined above. We construct subsets of A in terms of the edges between A and B.

▶ **Definition 16.**

1. *For every vertex $x \in A$, let $\mathbf{d}_B(x)$ denote the B-degree of x, which is the number of edges between x and B.*
2. *For every $0 \le k \le |B|$, let $\mathbf{S}_k = \{v \in A : \mathbf{d}_B(v) = k\}$. That is, $\mathbf{S}_k$ is the set of vertices from A with B-degree k.*
3. *For every $x \in A$, let $\mathbf{S}[x] = \mathbf{S}_{\mathbf{d}_B(x)} = \{v \in A : \mathbf{d}_B(v) = \mathbf{d}_B(x)\}$. That is, $\mathbf{S}[x]$ is the set of vertices from A with the same B-degree as x.*

▶ Remark 17. Given a vertex $x \in A$, the statement $v \in \mathbf{S}[x]$ is expressible as a formula in $\mathcal{FO}_= (u_1, u_2, x, v)$:

$$v \in A \wedge Q_= y \left(y \in B \wedge y \sim v, y \in B \wedge y \sim x\right)$$

where $x \in A$ means $x \sim u_1 \wedge x \sim u_2$ and $y \in B$ means $y \sim u_1 \wedge \neg(y \sim u_2)$.

Importantly, note that the B-degrees $\left(\mathbf{d}_B(x)\right)_{x\in A}$ are i.i.d. with distribution $\mathrm{Bin}\left(|B|, p\right)$.

▶ **Theorem 18.** *Let $c > 0$ be a constant. Then, with exponentially high probability, there exists $k = k(n)$ such that $0 \le k \le |B|$ and $|\mathbf{S}_k| = (1 + o(1))c \ln n$.*

We can reformulate the statement of theorem more explicitly by recalling the definition of exponentially high probability (see *Notation and conventions* above). Given a positive constant c, the statement is that there exists a sequence $(\varepsilon_n)_{n=1}^{\infty}$ such that, as $n \to \infty$, we have $\varepsilon_n = o(1)$ and

$$\mathbb{P}\left(\exists 0 \le k \le |B| : \left|\frac{|\mathbf{S}_k|}{c \ln n} - 1\right| \ge \varepsilon_n\right) = \exp\left(-n^{\Omega(1)}\right).$$

For the proof of Theorem 18 we use the following normal approximations of binomial probabilities (see [3, Theorems 1.2 and 1.5]).

▷ Claim 19. Let $p \in (0, 1)$ and $n \in \mathbb{N}$. Let $\mu = np$ and $\sigma = \sqrt{p(1-p)n}$ be the mean and standard deviation of the binomial distribution $\mathrm{Bin}(n, p)$. Let $0 \le k \le n$ be an integer and let $b(k; n, p) = \mathbb{P}\left(\mathrm{Bin}\left(n, p\right) = k\right)$. Write $k = \mu + h$.

1. Assume $\mu \ge 1$ and $\frac{h(1-p)n}{3} \ge 1$. Then

$$b(k; n, p) \le \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{h^2}{2\sigma^2} + \frac{h}{(1-p)n} + \frac{h^3}{p^2 n^2}\right).$$

2. Assume $\mu \geq 1$, $h > 0$ and $k < n$. Then

$$b(k;n,p) \geq \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{h^2}{2\sigma^2} - \frac{h^3}{2(1-p)^2n^2} - \frac{h^4}{3p^3n^3} - \frac{h}{2pn} - \frac{1}{12k} - \frac{1}{12(n-k)}\right).$$

In the proof of Theorem 18 we shall use the following notation:

1. $n_A = |A|$ and $n_B = |B|$.
2. $\mu = pn_B$ and $\sigma = \sqrt{p(1-p)n_B}$.
3. $p_k = \mathbb{P}\left(\text{Bin}\left(n_B, p\right) = k\right)$ for every $0 \leq k \leq n_B$.

▶ **Lemma 20.** *Let $c > 0$ be a constant. For every $n \in \mathbb{N}$ let $t_0 \in \mathbb{R}$ be the unique positive solution of*

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{t_0^2}{2}\right) = \frac{c \ln n}{n}$$

and let $k_0 = \mu + t_0\sigma$. Then, for every integer $k \in \left[k_0 - n^{1/4}, k_0 + n^{1/4}\right]$ we have $p_k = (1+o(1))\frac{c\ln n}{n}$ (where the asymptotic term $o(1)$ is uniform with respect to k).

Proof of Lemma 20. First note that $n_B = \Theta(n)$, $\mu = \Theta(n)$, $\sigma = \Theta\left(n^{1/2}\right)$ and $t_0 = (1+o(1))\sqrt{\ln n}$. For a given integer $k \in \left[k_0 - n^{1/4}, k_0 + n^{1/4}\right]$, we can write $k = \mu + t\sigma$ for $t = t_0 + O(n^{-1/4})$. Applying Part 1 of Claim 19 (with $h = t\sigma$ and $n = n_B$),

$$\begin{aligned} p_k &\leq \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{t^2}{2}\right) \cdot \exp\left(\frac{t\sigma}{(1-p)n_B} + \frac{t^3\sigma^3}{p^2n_B^2}\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{t_0^2}{2}\right) \cdot \exp\left(O(t_0 n^{-1/4})\right) \cdot \exp\left(O(t_0 n^{-1/2})\right) \\ &= (1+o(1))\frac{c\ln n}{n}. \end{aligned}$$

Applying Part 2 of Claim 19 (with $h = t\sigma$ and $n = n_B$),

$$\begin{aligned} p_k &\geq \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{t^2}{2}\right) \\ &\quad \cdot \exp\left(-\frac{t^3\sigma^3}{2(1-p)^2b^2} - \frac{t^4\sigma^4}{3p^3b^3} - \frac{t\sigma}{2pb} - \frac{1}{12k} - \frac{1}{12(n-k)}\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{t_0^2}{2}\right) \cdot \exp\left(O(t_0 n^{-1/4})\right) \cdot \exp\left(O(t_0 n^{-1/2})\right) \\ &= (1+o(1))c\frac{\ln n}{n}. \end{aligned}$$

Overall we have $p_k = (1+o(1))c\frac{\ln n}{n}$, where the $o(1)$ term can be taken to be $O\left((\ln n)^{1/2} n^{-1/4}\right)$ and uniform with respect to k. ◀

Proof of Theorem 18. Note that $\mathbf{s}_k := |\mathbf{S}_k| \sim \text{Bin}\left(n_A, p_k\right)$ for every $0 \leq k \leq n_B$. The variables $\{\mathbf{s}_k\}_{k=0}^{n_B}$ are not independent, since $\sum_{k=0}^{n_B} \mathbf{s}_k = n_A$. However, we can replace them with independent variables by introducing a Poisson process.

Let $\{\mathbf{d}_i\}_{i=1}^{\infty}$ be i.i.d. variables with distribution $\text{Bin}\left(n_B, p\right)$ and let $\mathbf{N} \sim \text{Pois}(n_A)$ be independent of $\{\mathbf{d}_i\}_{i=1}^{\infty}$. These variables define the Poisson process $\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_{\mathbf{N}}$. For every $0 \leq k \leq n_B$ let $\tilde{\mathbf{s}}_k$ count the number of times the value k appears in the process; that is, $\tilde{\mathbf{s}}_k = |\{0 \leq i \leq \mathbf{N} : \mathbf{d}_i = k\}|$. Then the variables $\{\tilde{\mathbf{s}}_k\}_{k=0}^{n_B}$ satisfy the following two properties:

1. The distribution of $\{\tilde{\mathbf{s}}_k\}_{k=0}^{n_B}$ given $\mathbf{N} = n_A$ is identical to the distribution of $\{\mathbf{s}_k\}_{k=0}^{n_B}$.
2. $\{\tilde{\mathbf{s}}_k\}_{k=0}^{n_B}$ are independent and $\tilde{\mathbf{s}}_k \sim \text{Pois}(n_A p_k)$ for every k.

We now apply Lemma 20 with $\frac{c}{p^2}$ as the constant. For every integer $k \in \left[k_0 - n^{1/4}, k_0 + n^{1/4}\right]$ we then have

$$\mathbb{E}\left(\tilde{\mathbf{s}}_k\right) = n_A p_k = (1+o(1))p^2 n \cdot \frac{c}{p^2}\ln n = (1+o(1))c\ln n.$$

From Equation (3) we deduce that there exists a sequence $\varepsilon_n = o(1)$ such that for every integer $k \in \left[k_0 - n^{1/4}, k_0 + n^{1/4}\right]$,

$$\mathbb{P}\left(\tilde{\mathbf{s}}_k \notin [(1-\varepsilon_n)c\ln n, (1+\varepsilon_n)c\ln n]\right) \leq \frac{1}{2}.$$

Write

$$\begin{aligned} I &= [(1-\varepsilon_n)c\ln n, (1+\varepsilon_n)c\ln n]\,, \\ K &= \left[k_0 - n^{1/4}, k_0 + n^{1/4}\right] \cap \mathbb{Z} \end{aligned}$$

for short. Then, from independence,

$$\mathbb{P}\left(\tilde{\mathbf{s}}_k \notin I\ \forall k \in K\right) \leq \left(\frac{1}{2}\right)^{|K|} = \exp\left(-\Theta(n^{1/4})\right).$$

Therefore there exists k such that $\tilde{\mathbf{s}}_k \in I$ with exponentially high probability.

Finally, we condition on the event $\mathbf{N} = n_A$. By Stirling's approximation, $\mathbb{P}\left(\mathbf{N} = n_A\right) = \Theta\left(n_A^{-1/2}\right) = \Theta\left(n^{-1/2}\right)$. Overall

$$\mathbb{P}\left(\mathbf{s}_k \notin I\ \forall k \in K\right) \leq \frac{\mathbb{P}\left(\tilde{\mathbf{s}}_k \notin I\ \forall k \in K\right)}{\mathbb{P}\left(\mathbf{N} = n_A\right)} = \frac{\exp\left(-\Theta(n^{1/4})\right)}{\Theta\left(n^{-1/2}\right)} = \exp\left(-\Theta(n^{1/4})\right).$$

We conclude that, with exponentially high probability, there exists k such that $\mathbf{s}_k \in I$, and so $\mathbf{s}_k = (1+o(1))c\ln n$ as we wanted. ◀

▶ **Corollary 21.** *Let $c > 0$ be a constant. Then, with exponentially high probability, there exists $x \in A$ such that $|\mathbf{S}[x]| = (1+o(1))c\ln n$.*

Proof. Given k such that $|\mathbf{S}_k| = (1+o(1))c\ln n$, pick any $x \in \mathbf{S}_k$ and then $\mathbf{S}_k = \mathbf{S}[x]$. ◀

2.2 Expressing Unary Relations

To express subsets of a given set $S \subseteq A$, we use the edges between S and C.

▶ **Definition 22.** *For a set $S \subseteq A$ and a vertex $z \in C$ let $S_z = \{s \in S : s \sim z\}$. We say that S_z is the subset of S* defined by *z.*

▶ **Proposition 23.** *There exists a positive constant $c_1 < \frac{1}{2}$ such that the following holds with exponentially high probability. For every $x \in A$, if $|\mathbf{S}[x]| \leq 2c_1 \ln n$ then for every subset $T \subseteq \mathbf{S}[x]$ there exists $z \in C$ such that $T = \mathbf{S}[x]_z$.*

The purpose of the condition $c_1 < \frac{1}{2}$ will become apparent in Section 3 (see Lemma 30).

Proof. Let $p_1 = \min\{p, 1-p\}$ and choose c_1 to be a sufficiently small constant such that $c_1 < \frac{1}{2}$ and $\gamma_1 := -2c_1 \ln p_1 < 1$. For this proof only, let us say that a subset $S \subseteq A$ is *good* if for every subset $T \subseteq S$ there exists $z \in C$ such that $T = S_z$.

First, fix $S \subseteq A$ of size $|S| \leq 2c_1 \ln n$ and a subset $T \subseteq S$. For every $z \in C$,

$$\mathbb{P}\left(T = S_z\right) = p^{|T|}(1-p)^{|S|-|T|} \geq p_1^{|S|} \geq p_1^{2c_1 \ln n} = n^{-\gamma_1}.$$

Crucially, the subsets of S defined by different vertices $z \in C$ are independently distributed. Thus

$$\mathbb{P}\left(\forall z \in C.\ T \neq S_z\right) = \left(1 - p^{|T|}(1-p)^{|S|-|T|}\right)^{|C|} \leq \left(1 - n^{-\gamma}\right)^{|C|} = \exp\left(-\Theta(n^{1-\gamma_1})\right).$$

Taking a union bound over $2^{|S|} = 2^{\Theta(\ln n)}$ possible choices of the subset T, we deduce $\mathbb{P}\left(S \text{ is not good}\right) = \exp\left(-n^{\Omega(1)}\right)$.

Finally, for every $x \in A$ we can apply the law of total probability with respect to the possible values of $\mathbf{S}[x]$, and deduce that, with exponentially high probability, $|\mathbf{S}[x]| \leq 2c_1 \ln n$ implies that $\mathbf{S}[x]$ is good. Taking a union bound over $\Theta(n)$ possible choices of x, we get the desired result. ◀

We will also need the following analogous proposition, which will be used to control the upper bound on the size of the definable sets.

▶ **Proposition 24.** *There exists a positive constant c_2 such that $c_2 \geq 2c_1$ and the following holds with probability $1 - o\left(n^{-2}\right)$. For every $x \in A$, if $|\mathbf{S}[x]| \geq c_2 \log_2 n$ then for every $z_1, z_2 \in C$, if $z_1 \neq z_2$ then $\mathbf{S}[x]_{z_1} \neq \mathbf{S}[x]_{z_2}$.*

Again, the purpose of the condition $c_2 \geq 2c_1$ will become apparent in Section 3.

Proof. Let $p_2 = \max\{p, 1-p\}$ and choose c_2 to be a sufficiently large constant such that $c_2 \geq 2c_1$ and $\gamma_2 := -c_2 \ln p_1 > 5$. For this proof only, let us say that a subset $S \subseteq A$ is *good* if for every $z_1, z_2 \in C$, if $z_1 \neq z_2$ then $S_{z_1} \neq S_{z_2}$.

First, fix $S \subseteq A$ of size $|S| \geq c_2 \ln n$. For every pair of distinct vertices $z_1, z_2 \in C$,

$$\mathbb{P}\left(S_{z_1} = S_{z_2}\right) \leq p_2^{|S|} \leq p_2^{c_2 \ln n} = n^{c_2 \ln p_2} = n^{-\gamma_2}.$$

Taking a union bound over $\Theta(n^2)$ choices of $z_1 \neq z_2$, we get

$$\mathbb{P}\left(S \text{ is not good}\right) = O(n^{2-\gamma_2}).$$

Finally, for every $x \in A$ we can apply the law of total probability with respect to the possible values of $\mathbf{S}[x]$ and deduce that, with probability $1 - O(n^{2-\gamma_2})$, $|\mathbf{S}[x]| \geq c_2 \ln n$ implies that $\mathbf{S}[x]$ is good. Taking a union bound over $\Theta(n)$ possible choices of x, and recalling that $n^{3-\gamma_2} = o(n^{-2})$ by definition, we get the desired result.

◀

2.3 Non-Conditioned Results

Finally, we lose the conditioning on the events $Q_{A,B,C}$ which fix the values of $\mathbf{A}, \mathbf{B}, \mathbf{C}$. Note that we still have dependency on the choice of u_1, u_2; we will quantify over u_1, u_2 in the next section. The following theorem summarizes all the probabilistic results proved in this section.

▶ **Theorem 25.** *There exist positive constants c_1, c_2 with $c_1 < \frac{1}{2}$ and $c_2 \geq 2c_1$ and sequences $\delta_n = o(1)$ and $\varepsilon_n = o(1)$ such that $\mathbb{P}(\Gamma(u_1, u_2)) = 1 - o(n^{-2})$, where $\Gamma(u_1, u_2)$ is the event that:*

1.

$$\frac{|\mathbf{A}|}{p^2 n}, \frac{|\mathbf{B}|}{p(1-p)n}, \frac{|\mathbf{C}|}{p(1-p)n} \in [1 - \delta_n, 1 + \delta_n]$$

(we denote this event by $\mathcal{Q}$);

2. *There exists $x \in \mathbf{A}$ such that*
$$|\mathbf{S}[x]| \in \left[(1-\varepsilon_n)\frac{3}{2}c_1 \ln n, (1+\varepsilon_n)\frac{3}{2}c_1 \ln n\right];$$
3. *For every $x \in \mathbf{A}$, if $|\mathbf{S}[x]| \le 2c_1 \ln n$ then for every $T \subseteq \mathbf{S}[x]$ there exists $z \in \mathbf{C}$ such that $T = \mathbf{S}[x]_z$; and*
4. *For every $x \in \mathbf{A}$, if $|\mathbf{S}[x]| \ge c_2 \ln n$ then for every $z_1, z_2 \in \mathbf{C}$, if $z_1 \neq z_2$ then $\mathbf{S}[x]_{z_1} \neq \mathbf{S}[x]_{z_2}$.*

Proof. Let $\delta_n = o(1)$ be the sequence from Equation (4). In the previous sections, we conditioned all probabilities on $Q_{A,B,C}$ where $Q_{A,B,C} \subseteq \mathcal{Q}$. However, note that the proofs of Theorem 18 and Propositions 23, 24 do not depend on the specific values A, B, C, but only on the fact that they satisfy Equation (4). Therefore, they also hold when the conditioning is on the event $\mathcal{Q}$.

Now, let c_1, c_2 be the constants from Propositions 23 and 24 (respectively) and let $(\varepsilon_n)_{n=1}^{\infty}$ be the sequence from Theorem 18 for $c = \frac{3}{2}c_1$. Define $\Gamma(u_1, u_2)$ as above. Then

$$\begin{aligned}\mathbb{P}(\Gamma(u_1,u_2)) &= \mathbb{P}\left(\Gamma(u_1,u_2) \;\middle|\; \mathcal{Q}\right)\mathbb{P}(\mathcal{Q}) \\ &\ge \left(1 - o(n^{-2})\right)\left(1 - \exp\left(-n^{\Omega(1)}\right)\right) = 1 - o(n^{-2}).\end{aligned}$$

That concludes the proof. ◀

▶ Remark 26. Note that, due to symmetry considerations, $\mathbb{P}(\Gamma(u_1, u_2))$ does not depend on the choice of u_1, u_2.

3 Proof of the Main Theorem

In this section we complete the proof of Theorem 14. We begin with a sequence of short lemmas, which build upon our previous results. Once again, we fix a constant $p \in (0,1)$ and consider a binomial random graph $\mathbf{G}_n \sim G(n,p)$. Probabilities are now non-conditioned.

We begin with a refinement of Theorem 9.

▶ **Lemma 27** (Kaufmann-Shelah). *There exists a sentence* Encode$^* \in \mathcal{MSO}$ *such that:*
1. $\lim_{n\to\infty} \mathbb{P}(\mathbf{G}_n \models \text{Encode}^*) = 1$.
2. *For every sentence $\varphi \in \mathcal{SO}[\text{Arith}]$ there exists a sentence $\varphi^* \in \mathcal{MSO}$ such that, for every $n \in \mathbb{N}$ and graph $G = ([n], E)$ with $G \models \text{Encode}^*$, we have $G \models \varphi^* \iff [\sqrt{n}] \models \varphi$.*

Proof. This follows from Theorem 1 and the closing remark of [23]. ◀

Intuitively, the sentence Encode* asserts the existence of $\mathcal{MSO}$-formulas expressing a structure of addition and multiplication on the vertices, a necessary ingredient for converting any $\varphi \in \mathcal{SO}[\text{Arith}]$ into $\varphi^* \in \mathcal{MSO}$. The basic structure of the sentence Encode* is given in Theorem 1 of [23].

▶ **Lemma 28.** *For every sentence $\varphi^* \in \mathcal{MSO}$ there exists a formula $\varphi^{**}(u_1,u_2,x) \in \mathcal{FO}_=(u_1,u_2,x)$ such that the following holds. Given the event $\Gamma(u_1,u_2)$, for every $x \in \mathbf{A}$ with $|\mathbf{S}[x]| \le 2c_1 \ln n$ we have*

$$\mathbf{G}_n \models \varphi^{**}(u_1,u_2,x) \iff \mathbf{G}_n[\mathbf{S}[x]] \models \varphi^*.$$

Proof. Given φ^*, define $\varphi^{**}(u_1,u_2,x)$ as follows:

- Restrict quantification to $\mathbf{S}[x]$: replace every $\forall v\,(\theta)$ with $\forall v\,(v \in \mathbf{S}[x] \to \theta)$ and every $\exists v\,(\theta)$ with $\exists v\,(v \in \mathbf{S}[x] \wedge \theta)$. Recall that the statement $v \in \mathbf{S}[x]$ is expressible as a formula in $\mathcal{FO}_{=}\,(u_1, u_2, x, v)$ (see Remark 17).
- Convert unary relations: for every unary relation R introduced by ψ, replace $\exists R\,(\theta)$ with $\exists z_R(z_R \in \mathbf{C} \wedge \theta)$ where z_R is a new variable symbol, and also replace every $R(v)$ with $v \sim z_R$. Similarly handle $\forall R\,(\theta)$. Recall that the statement $z \in \mathbf{C}$ is expressible the formula $z \not\sim u_1 \wedge z \sim u_2$, which is in $\mathcal{FO}(u_1, u_2, z)$

Given the event $\Gamma(u_1, u_2)$, for every $x \in \mathbf{A}$ with $|\mathbf{S}[x]| \leq 2c_1 \ln n$, we know that every subset of $\mathbf{S}[x]$ is defined by some $z \in \mathbf{C}$ (see Part 3 of Theorem 25). Therefore

$$\mathbf{G}_n \models \varphi^{**}(u_1, u_2, x) \iff \mathbf{G}_n[\mathbf{S}[x]] \models \varphi^*$$

as we wanted. ◀

Next, we introduce a formula for upper-bounding the size of definable sets.

▶ **Definition 29.** *Given a choice of $u_1, u_2 \in [n]$ and $x \in \mathbf{A}$, we say that $\mathbf{S}[x]$ is* pseudo-logarithmic *if there exist $z_1, z_2 \in \mathbf{C}$ such that $z_1 \neq z_2$ but $\mathbf{S}[x]_{z_1} = \mathbf{S}[x]_{z_2}$.*

Note that this property is expressible as a formula in $\mathcal{FO}_{=}(u_1, u_2, x)$, given by

$$\exists z_1 \exists z_2\,(z_1 \in \mathbf{C} \wedge z_2 \in \mathbf{C} \wedge z_1 \neq z_2 \wedge \mathbf{S}[x]_{z_1} = \mathbf{S}[x]_{z_2})\,,$$

where $\mathbf{S}[x]_{z_1} = \mathbf{S}[x]_{z_2}$ is the formula $\forall y\,(y \in \mathbf{S}[x] \to (y \sim z_1 \leftrightarrow y \sim z_2))$.

▶ **Lemma 30.** *Given the event $\Gamma(u_1, u_2)$, for every $x \in \mathbf{A}$,*

1. *If $\mathbf{S}[x]$ is pseudo-logarithmic then $|\mathbf{S}[x]| \leq c_2 \ln n$.*
2. *If $|\mathbf{S}[x]| \leq 2c_1 \ln n$ then $\mathbf{S}[x]$ is pseudo-logarithmic.*

Proof. Part 1 follows directly from the definition of $\Gamma(u_1, u_2)$ (see part 4 of Theorem 25). Part 2 follows from the pigeonhole principle. Indeed, let $S = \mathbf{S}[x]$ and assume $|S| \leq 2c_1 \ln n$. Then the number of subsets of S is $2^{|S|} \leq 2^{2c_1 \ln n} = n^{2c_1 \ln 2}$. Recall that $c_1 < \frac{1}{2}$, so $2^{|S|} \leq n^{\ln 2} = o(n)$. However, given $\Gamma(u_1, u_2)$ we have $|\mathbf{C}| = \Theta(n)$. From the pigeonhole principle there must exist $z_1, z_2 \in \mathbf{C}$ such that $z_1 \neq z_2$ but $S_{z_1} = S_{z_2}$, hence S is pseudo-logarithmic. ◀

Finally, we introduce a formula for comparing sizes of definable sets.

▶ **Definition 31.** *Given a choice of $u_1, u_2 \in [n]$ and two vertices $x, x' \in \mathbf{A}$, we say that $\mathbf{S}[x]$ is* pseudo-smaller *than $\mathbf{S}[x']$ if there exists $z \in \mathbf{C}$ such that $|\mathbf{S}[x']_z| = |\mathbf{S}[x]|$.*

Note that this property is expressible as a formula in $FO_{=}(u_1, u_2, x, x')$, since belonging to $\mathbf{S}[x]$ and to $\mathbf{S}[x']_z$ and the equicardinality condition $|\mathbf{S}[x']_z| = |\mathbf{S}[x]|$ are all expressible in $\mathcal{FO}_{=}$.

▶ **Lemma 32.** *Given the event $\Gamma(u_1, u_2)$, for every $x, x' \in \mathbf{A}$,*

1. *If $\mathbf{S}[x]$ is pseudo-smaller than $\mathbf{S}[x']$ then $|\mathbf{S}[x]| \leq |\mathbf{S}[x']|$.*
2. *If $|\mathbf{S}[x]| \leq |\mathbf{S}[x']| \leq 2c_1 \ln n$ then $\mathbf{S}[x]$ is pseudo-smaller than $\mathbf{S}[x']$.*

Proof. Part 1 follows from the definition of pseudo-smaller (in fact, it is true deterministically). Part 2 follows from the definition of $\Gamma(u_1, u_2)$ (see Part 3 of Theorem 25). ◀

We are now ready to prove the main theorem. In the proof, we use the notation $\mathbf{E}(X, Y)$ for the set of edges in $\mathbf{G}_n$ between two disjoint sets of vertices $X, Y \subseteq [n]$.

Proof of Theorem 14. We prove that $\mathcal{FO}_=$ can interpret a segment of arithmetic with constants $\sqrt{c_1}, \sqrt{c_2}$ and scaling function $f(n) = \left\lfloor \sqrt{\ln n} \right\rfloor$, where c_1, c_2 are the constants from Theorem 25. That is, for every sentence $\varphi \in \mathcal{SO}[\text{Arith}]$ we construct a sentence $\psi \in \mathcal{FO}_=$ such that the following holds. Given a sequence $\{n_k\}_{k=1}^{\infty}$ with $\lim_{k\to\infty} n_k = \infty$ such that φ is constant on $\left[\sqrt{c_1 \ln n_k}, \sqrt{c_2 \ln n_k}\right]$ for every k, we have

$$\lim_{k\to\infty} \mathbb{P}\left(G(n_k, p) \models \psi \iff \varphi \text{ holds in } \left[\sqrt{c_1 \ln n_k}, \sqrt{c_2 \ln n_k}\right]\right) = 1. \tag{5}$$

From now on we fix a sentence $\varphi \in \mathcal{SO}[\text{Arith}]$. First, we apply Lemma 27 to obtain a sentence $\varphi^* \in \mathcal{MSO}$ such that, for every graph $G = ([n], E)$ with $G \models \text{Encode}^*$, we have $G \models \varphi^* \iff [\sqrt{n}] \models \varphi$. Second, we apply Lemma 28 to the $\mathcal{MSO}$-sentences φ^* and Encode^* (from Lemma 27) to obtain formulas $\varphi^{**}(u_1, u_2, x), \text{Encode}^{**}(u_1, u_2, x) \in \mathcal{FO}_=(u_1, u_2, x)$ such that, given the event $\Gamma(u_1, u_2)$, for every $x \in \mathbf{A}$ with $|\mathbf{S}[x]| \leq 2c_1 \ln n$ we have

$$\mathbf{G}_n \models \varphi^{**}(u_1, u_2, x) \iff \mathbf{G}_n[\mathbf{S}[x]] \models \varphi^*, \tag{6}$$
$$\mathbf{G}_n \models \text{Encode}^{**}(u_1, u_2, x) \iff \mathbf{G}_n[\mathbf{S}[x]] \models \text{Encode}^*. \tag{7}$$

Now define ψ as the sentence claiming the existence of two vertices u_1, u_2 and a vertex $x \in \mathbf{A}$ such that:

1. $\mathbf{S}[x]$ is pseudo-logarithmic and $\mathbf{G}_n[\mathbf{S}[x]] \models \text{Encode}^*$.
2. If $x' \in \mathbf{A}$ is another vertex such that $\mathbf{S}[x']$ is pseudo-logarithmic and $\mathbf{G}_n[\mathbf{S}[x]] \models \text{Encode}^*$, then $\mathbf{S}[x]$ is not pseudo-smaller than $\mathbf{S}[x']$.
3. $\mathbf{G}_n[\mathbf{S}[x]] \models \varphi^*$.

From Lemmas 30, 32 and (6), (7) above, ψ is indeed expressible as a sentence in $\mathcal{FO}_=$. It remains to verify (5).

Let $\{n_k\}_{k=1}^{\infty}$ with $\lim_{k\to\infty} n_k = \infty$ such that φ is constant on $\left[\sqrt{c_1 \ln n_k}, \sqrt{c_2 \ln n_k}\right]$ for every k. There are two cases to consider.

Case 1: φ holds in $\left[\sqrt{c_1 \ln n_k}, \sqrt{c_2 \ln n_k}\right]$.

Fix two vertices u_1, u_2 arbitrarily. We show that, with high probability, there exists a vertex $x \in \mathbf{A}$ which satisfies the three parts of ψ (along with u_1, u_2).

First, we know that $\mathbb{P}(\Gamma(u_1, u_2)) = 1 - o(1)$, so from now on we may assume that $\Gamma(u_1, u_2)$ holds. Recall that the event $\Gamma(u_1, u_2)$ guarantees a vertex $x' \in \mathbf{A}$ such that

$$|\mathbf{S}[x']| \in \left[(1-\varepsilon_n)\frac{3}{2}c_1 \ln n_k, (1+\varepsilon_n)\frac{3}{2}c_1 \ln n_k\right].$$

Fix such a vertex x'. Since $c_2 \geq 2c_1 \geq (1+\varepsilon_n)\frac{3}{2}c_1$, Lemma 32 implies that $\mathbf{S}[x']$ is pseudo-logarithmic. Let us show that $\mathbf{G}_{n_k}[\mathbf{S}[x']] \models \text{Encode}^*$ with high probability.

Condition on the values A, B, C of the sets $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and on the edge sets $\mathbf{E}(A, B)$ and $\mathbf{E}(A, C)$. These values determine the value S of the set $\mathbf{S}[x']$. Crucially, the induced subgraph $\mathbf{G}_{n_k}[S]$ depends only on the edge set $\mathbf{E}(A)$, and so, given the last conditioning, $\mathbf{G}_{n_k}[S]$ is still binomially distributed with vertex set S and edge probability p. From Lemma 27 we get that, given the last conditioning, $\mathbf{G}_{n_k}[S] \models \text{Encode}^*$ with high probability. Now apply the law of total probability over the possible values of $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{E}(A, B), \mathbf{E}(A, C)$ to conclude that $\mathbf{G}_{n_k}[\mathbf{S}[x]] \models \varphi^*$ with high probability.

Next, among all vertices $x \in \mathbf{A}$ such that $\mathbf{S}[x]$ is pseudo-logarithmic and $\mathbf{G}_{n_k}[\mathbf{S}[x]] \models \text{Encode}^*$, pick x such that $|\mathbf{S}[x]|$ is maximal. By definition, x satisfies Part 1 and Part 2 of

ψ. We show that it also satisfies Part 3. Since $\mathbf{S}[x]$ is pseudo-logarithmic, Lemma 30 implies $|\mathbf{S}[x]| \le c_2 \ln n_k$. We also know that

$$|\mathbf{S}[x]| \ge |\mathbf{S}[x']| \ge (1-\varepsilon_n)\frac{3}{2}c_1 \ln n_k \ge c_1 \ln n_k + 1.$$

Letting $\mathbf{s} = |\mathbf{S}[x]|$, we deduce $\lfloor\sqrt{\mathbf{s}}\rfloor \in \left[\sqrt{c_1 \ln n_k}, \sqrt{c_2 \ln n_k}\right]$. From the assumption of Case 1, $[\lfloor\sqrt{\mathbf{s}}\rfloor] \models \varphi$. Since $\mathbf{G}_{n_k}[\mathbf{S}[x']] \models \text{Encode}^*$, Lemma 27 implies $\mathbf{G}_{n_k}[\mathbf{S}[x']] \models \varphi^*$ as we wanted.

Case 2: $\neg\varphi$ holds in $\left[\sqrt{c_1 \ln n_k}, \sqrt{c_2 \ln n_k}\right]$.

We need to prove that with high probability, for every u_1, u_2, there is no vertex x that satisfies all three parts of ψ. Let $\Gamma = \bigcap_{u_1,u_2} \Gamma(u_1,u_2)$. Theorem 25 shows that $\mathbb{P}(\Gamma(u_1,u_2)) = 1 - o(n^{-2})$ for every u_1, u_2. Taking a union bound the over $\Theta(n^2)$ pairs u_1, u_2 we get $\mathbb{P}(\Gamma) = 1 - o(1)$. So from now on we may assume that Γ holds.

Assume that u_1, u_2, x are vertices such that Part 1 and Part 2 of ψ hold and let us show that Part 3 does not hold. Again, Γ guarantees a vertex $x' \in \mathbf{A}$ such that

$$|\mathbf{S}[x']| \in \left[(1-\varepsilon_n)\frac{3}{2}c_1 \ln n_k, (1+\varepsilon_n)\frac{3}{2}c_1 \ln n_k\right].$$

We prove $|\mathbf{S}[x]| \ge |\mathbf{S}[x']|$ by contradiction. Indeed, otherwise we have

$$|\mathbf{S}[x]| \le |\mathbf{S}[x']| \le (1+\varepsilon_n)\frac{3}{2}c_1 \ln n_k \le 2c_1 \ln n_k,$$

and from Lemma 30 we get that $\mathbf{S}[x]$ is pseudo-smaller than $\mathbf{S}[x']$. But that contradicts Part 2 of ψ. Therefore

$$|\mathbf{S}[x]| \ge |\mathbf{S}[x']| \ge (1-\varepsilon_n)\frac{3}{2}c_1 \ln n_k \ge c_1 \ln n_k + 1.$$

Moreover, $\mathbf{S}[x]$ is pseudo-logarithmic, so Lemma 30 implies $|\mathbf{S}[x]| \le c_2 \ln n$. As before, letting $\mathbf{s} = |\mathbf{S}[x]|$, we get $\lfloor\sqrt{\mathbf{s}}\rfloor \in \left[\sqrt{c_1 \ln n_k}, \sqrt{c_2 \ln n_k}\right]$. From the assumption of Case 2, $[\lfloor\sqrt{\mathbf{s}}\rfloor] \not\models \varphi$. Since $\mathbf{G}_{n_k}[\mathbf{S}[x']] \models \text{Encode}^*$, Lemma 27 implies $\mathbf{G}_{n_k}[\mathbf{S}[x']] \not\models \varphi^*$, as we wanted. ◀

References

1 Noga Alon and Joel H. Spencer. *The Probabilistic Method.* Wiley Publishing, 4th edition, 2016.

2 Andreas Blass and Frank Harary. Properties of almost all graphs and complexes. *Journal of Graph Theory*, 3(3):225–240, 1979.

3 Béla Bollobás. *Random Graphs.* Cambridge University Press, 2001.

4 Kevin J. Compton. 0-1 laws in logic and combinatorics. In Rival Ivan, editor, *Algorithms and Order*, volume 255 of *Advanced Study Institute Series C: Mathematical and Physical Sciences*, pages 353–383. Kluwer Academic Publishers, Dordrecht, 1989.

5 Anuj Dawar and Erich Grädel. Generalized quantifiers and 0-1 laws. In *Proceedings of the Tenth Annual IEEE Symposium on Logic in Computer Science (LICS 1995)*, pages 54–64. IEEE Computer Society Press, June 1995.

6 Anuj Dawar and Erich Grädel. Properties of almost all graphs and generalized quantifiers. *Fundam. Inform.*, 98(4):351–372, 2010.

7 Heinz-Dieter Ebbinghaus. Extended logics: The general framework. In Jon Barwise and Solomon Feferman, editors, *Model-Theoretic Logics*, chapter II, pages 25–76. Association for Symbolic Logic, September 1985.

8 Heinz-Dieter Ebbinghaus and Jörg Flum. *Finite Model Theory*. Springer Berlin, Heidelberg, second edition, 2006.

9 Paul Erdős and Alfréd Rényi. On random graphs, I. *Publicationes Mathematicae*, 6:290–297, 1959.

10 Paul Erdős and Alfréd Rényi. On the evolution of random graphs. In *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, number 5 in Acta Math. Acad. Sci. Hungar., pages 17–61, 1960.

11 Ronald Fagin. Probabilities on finite models. *The Journal of Symbolic Logic*, 41(1):50–58, March 1976.

12 Guy Fayolle, Stéphane Grumbach, and Christophe Tollu. Asymptotic probabilities of languages with generalized quantifiers. In *Proceedings Eighth Annual IEEE Symposium on Logic in Computer Science*, pages 199–207, 1993.

13 Yu. V. Glebskiĭ, D. I. Kogan, M. I. Liogon'kiĭ, and V. A. Talanov. Range and degree of realizability of formulas in the restricted predicate calculus. *Cybernetics and Systems Analysis*, 5(2):142–154, March 1969. (Russian original: Kibernetika 5(2):17–27, March-April 1969).

14 Simi Haber. Generalized quantifiers for simple graph properties in random graphs. , Preprint.

15 Simi Haber. Arithmetization for first order logic augmented with perfect matching. , Submitted.

16 Simi Haber and Saharon Shelah. Random graphs and Lindström quantifiers for natural graph properties. *Accepted, Annales Univ. Sci. Budapest., Sect. Math*, 2024+. HbSh:986 in Shelah's archive.

17 Klaus Härtig. Über einen quantifikator mit zwei wirk ungsbereichen. In Laszlo Kalmar, editor, *Colloquium on the Foundations of Mathematics, Mathematical Machines and their Applications*, pages 31–36. Akademiai Kiado, Budapest, September 1962.

18 Heinrich Herre, Michał Krynicki, Alexandr Pinus, and Jouko Väänänen. The Härtig quantifier: a survey. *Journal of Symbolic Logic*, 56(4):1153–1183, December 1991.

19 Svante Janson, Tomasz Łuczak, and Andrzej Ruciński. *Random Graphs*. John Wiley & Sons, 2000.

20 Risto Kaila. On probabilistic elimination of generalized quantifiers. *Random Struct. Algorithms*, 19(1):1–36, 2001.

21 Risto Kaila. Convergence laws for very sparse random structures with generalized quantifiers. *Math. Log. Q.*, 48(2):301–320, 2002.

22 Risto Kaila. On almost sure elimination of numerical quantifiers. *J. Log. Comput.*, 13(2):273–285, 2003.

23 Matt Kaufmann and Saharon Shelah. On random models of finite power and monadic logic. *Discrete Mathematics*, 54(3):285–293, 1985.

24 H. Jerome Keisler and Wafik Boulos Lotfallah. Almost everywhere elimination of probability quantifiers. *Journal of Symbolic Logic*, 74(4):1121–1142, 2009.

25 Phokion G. Kolaitis and Swastik Kopparty. Random graphs and the parity quantifier. *Journal of the Association for Computing Machinery*, 60(5):1–34, October 2013.

26 Per Lindström. First order predicate logic with generalized quantifiers. *Theoria*, 32(3):186–195, 1966.

27 Per Lindström. On extensions of elementary logic. *Theoria*, 35(1):1–11, 1969.

28 M. H. Löb. Meeting of the association for symbolic logic, leeds 1962. *The Journal of Symbolic Logic*, 27(3):373–382, 1962. First abstract: Plurality-quantification by Nicholas Rescher.

29 Peter Winkler. Random structures and zero-one laws. In Sauer N. W., Woodrow R. E., and Sands B., editors, *Finite and Infinite Combinatorics in Sets and Logic*, Advanced Science Institutes, pages 399–420. Kluwer Academic Publishers, Dordrecht, 1993.