# First order logic with equicardinality in random graphs

## ABSTRACT

We answer a question of Blass and Harari about the validity of the zero-one law in random graphs for extensions of first order logic. For several graph properties $P$ (e.g. Hamiltonicity), it is known that every (regular) logic able to express $P$ is also able to interpret 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 first order logic 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.

## CCS CONCEPTS

• **Mathematics of computing → Random graphs**; • **Theory of computation → Finite Model Theory**.

## KEYWORDS

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

**ACM Reference Format:**

. 2024. First order logic with equicardinality in random graphs. In *Proceedings of Logic in Computer Science (LICS'24)*. ACM, New York, NY, USA, 6 pages. https://doi.org/XXXXXXX.XXXXXXX

## 1 INTRODUCTION

In this paper we study zero-one laws in the random binomial graph $G(n,p)$. Recall that $G(n,p)$ is defined as a probability distribution over the set of all labeled graphs with vertex set $\{1,2,\ldots,n\}$, by requiring that each of the $\binom{n}{2}$ potential edges appears with probability $p$ and independently of all other edges. Zero-one laws — the phenomena where all properties in some (logical) set of properties are either almost surely valid or almost surely invalid for a given random graph — teaches us about the set of properties itself and the underlying random graph. It is thus considered an important part of Finite Model Theory. In this work we prove that a certain quantifier allows interpretation of a segment of arithmetic, which is the extreme opposite of a zero-one law, in the sense of having no control over asymptotic behavior of properties. We show that this quantifier explains all known cases of properties $P$ for which, if augment first order logic with $P$, we can interpret Arithmetic. Besides of its intrinsic value and the short proof, we see this result as an explanation, or a step towards one, of the phenomena.



*LICS'24, July 8–12, 2024, Tallinn, Estonia*



ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/XXXXXXX.XXXXXXX

The study of random graphs was pioneered by Erdös and Rényi in the 1960s, originating from two seminal papers [8, 9]. 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, consider $G(n,1/2)$, which is the uniform distribution over all labeled graphs on $n$ vertices. For a graph property $P$, we say that it holds *asymptotically almost surely* (a.a.s. for short) in $G(n,1/2)$ if

$$\lim_{n\to\infty} \mathbb{P}\left(G(n,1/2) \text{ satisfies A}\right) = 1$$

and that it holds *asymptotically almost never* (a.a.n. for short) in $G(n,1/2)$ if

$$\lim_{n\to\infty} \mathbb{P}\left(G(n,1/2) \text{ satisfies A}\right) = 0.$$

The observation that natural graph properties hold either a.a.s. or a.a.n. in $G(n,1/2)$ remains valid when $1/2$ is replaced by any constant probability strictly between 0 and 1, and so from now we consider $G(n,p)$ for a fixed $p \in (0,1)$.

The above observation is not a formal statement, because of the lack of a formal definition of "natural graph properties". From a logician's point of view, it is natural to consider 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 dictionary consists of a single binary relation $\sim$ representing adjacency.[1] Indeed, the classic zero-one law, proven independently by Glebskii et al. [11] and Fagin [10], states that every first order property holds either a.a.s. or a.a.n. in $G(n,p)$ for a constant $p \in (0,1)$. 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 a zero-one law. For example, as noted by Fagin [10], the property of having an even number of vertices is second order, but has no limiting probability.

It is therefore natural to ask for extensions of first order logic which are strong enough to express a given graph property $P$ on the one hand, but still obey a zero-one law on the other hand. This question was posed by Blass and Harari [2], in particular for Hamiltonicity and rigidity (asymmetry). They suggested monadic second order logic, denoted $\mathcal{MSO}$, as an extension which expresses rigidity and might obey the zero-one law, and also asked about extensions of $\mathcal{FO}$ with equicardinal quantifiers. These questions has been studied in many papers. The following review is far from being comprehensive; for a survey of the results in this field see, e.g., [4].

Note that a trivial extension of $\mathcal{FO}$ which includes $P$ is the union $\mathcal{FO} \cup \{P\}$. However, this class of properties clearly lacks a basic notion of closure. To avoid such trivialities, the extensions of first order logic which are required to be *regular*. A regular logic can be described as a logic that is closed under negation, conjunction,

[1] From now on we shall identify logical sentences with the properties they describe.

existential quantification, relativization and substitution (see [6] for more details). For a given property $P$, the minimal regular extension of first order logic which expresses $P$ is obtained by adding the *Lindström quantifier* of $P$, which we denote $Q_P$ [17, 18]. Syntactically, we define $\mathcal{FO}[Q_P]$ as the set of sentences obtained from the set of atomic formulas by closing it with respect to first order operations and quantification with $Q_P$, which is done as follows: given a formula $\varphi_V(x)$ with one free variable $x$ and a formula $\varphi_E(x, y)$ with two free variables $x, y$, return the formula $Q_P x, y\,(\varphi_V(x), \varphi_E(x, y))$ in which $x, y$ are quantified. Semantically, the truth value of this formula is defined as follows. For a given graph $G = (V, E)$, let $V_0 = \{v \in V : G \models \varphi_V(v)\}$ and let $E_0$ be the set of edges $\{u, v\} \in E$ with $u, v \in V_0$ such that $G \models \varphi(u, v)$ or $G \models \varphi(v, u)$. Then

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

For a more detailed treatment of these notions, see [7].

The answer to the first question of Blass and Harari, regarding Hamiltonicity and rigidity, was given by Dawar and Grädel [5], and in a stronger form by Haber and Shelah [12]. Let $Q_{\text{rig}}$ be the Lindström quantifier of rigidity and let $Q_{\text{Ham}}$ be the Lindström quantifier of Hamiltonicity. Then $\mathcal{FO}[Q_{\text{rig}}]$ obeys the zero-one law, while for $\mathcal{FO}[Q_{\text{Ham}}]$ the zero-one law fails in a very strong sense: this language can interpret a segment of arithmetic in $G(n, p)$, which allows (among other things) to express a property with no limiting probability. The answer to the second question, about monadic second order logic, was given much earlier by Kaufmann and Shelah [16]: $\mathcal{MSO}$ can interpret arithmetic, and so the zero-one law fails colossally.

As for other graph properties, Haber and Shelah also proved that the zero-one law holds for the extensions of $\mathcal{FO}$ with the Lindström quantifiers of connectivity and $k$-colorability for every fixed $k$. These results also follow a more general theorem by Dawar and Grádel [5], from which Planarity follows as well. On the other hand, there are additional graph properties for which it is known that the corresponding Lindström-extension of $\mathcal{FO}$ can interpret arithmetic: these include regularity, having a perfect matching and having a $C_4$-factor.

Common to all the Lindström-extensions of $\mathcal{FO}$ which are known to be able to interpret arithmetic is the ability to express the *equicardinality quantifier*, also known as the *Härtig quantifier* [13], which we denote $Q_=$. This quantifier allows for testing if two definable sets are of the same size. Syntactically, given formulas $\varphi(x)$ and $\psi(x)$ with one free variable $x$, the Härtig quantifier returns a formula $Q_= x\,(\varphi(x), \psi(x))$ in which $x$ is quantified. Semantically, for a given graph $G = (V, E)$ we have $G \models Q_= x\,(\varphi(x), \psi(x))$ if and only if

$$|\{v \in V : G \models \varphi(v)\}| = |\{v \in V : G \models \psi(v)\}|.$$

The Härtig quantifier is a natural extension of first order languages and was studied quite extensively in the context of general model theory and abstract logic [14].

In this paper we prove that the extension $\mathcal{FO}[Q_=]$ is able to express arithmetic in $G(n, p)$. This answers the third question of Blass and Harari, and provides a general result which immediately implies all known cases. We demonstrate the ability to interpret arithmetic with the existence of a property with no limiting probability.

Theorem 1.1 (Main Theorem). *Let $p \in (0, 1)$ be a constant. Then there exists a sentence $\varphi \in \mathcal{FO}[Q_=]$ such that the limit*

$$\lim_{n\to\infty} \mathbb{P}\,(G(n, p) \models \varphi)$$

*does not exist.*

The proof consists of two parts: probabilistic and logical. In the probabilistic part (Section 2) we show how $\mathcal{FO}[Q_=]$ can express sets of vertices of logarithmic size (which we call *logarithmic sets* for short). We also show how to express arbitrary subsets of a given logarithmic set. This allows the interpretation of $\mathcal{MSO}$ logic on logarithmic sets. For the logical part (Section 3) we apply Kauffman and Shelah [15], which, as mentioned above, show that $\mathcal{MSO}$ logic can interpret arithmetic in $G(n, p)$.

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

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 $\mathbf{G} \sim G(n, p)$ when $\mathbf{G}$ has distribution $G(n, p)$. For two vertices $u, v \in V$, let $u \sim v$ denote that they are adjacent in $\mathbf{G}$. For a subset $S \subseteq V$, let $\mathbf{G}[S]$ denote the subgraph of $\mathbf{G}$ induced by $S$.

Given two sequences of positive random variables $(\mathbf{X}_n)_{n=1}^{\infty}$, $(\mathbf{Y}_n)_{n=1}^{\infty}$, we define the following notions of asymptotic equivalence:

(1) We say that $\mathbf{X}_n = (1 + o(1))\,\mathbf{Y}_n$ with high probability if there exists a sequence $\varepsilon_n = o(1)$ such that

$$\mathbb{P}\left(\left|\frac{\mathbf{X}_n}{\mathbf{Y}_n} - 1\right| \le \varepsilon_n\right) = 1 - o(1). \tag{1}$$

(2) 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

$$\mathbb{P}\left(\left|\frac{\mathbf{X}_n}{\mathbf{Y}_n} - 1\right| \le \varepsilon_n\right) = 1 - \exp\left(-n^{\Omega(1)}\right).$$

For convenience, we often omit dependency on $n$ from our notation.

Finally, recall the following tail bounds on binomial and Poisson variables, following from Chernoff's inequality (see [1], Appendix A). Let $X \sim \text{Bin}(n, p)$ and $\mu = \mathbb{E}X$. Then for every $0 < \delta < 1$,

$$\mathbb{P}\,(|X - \mu| \ge \delta\mu) \le 2\exp\left(-\frac{\delta^2}{3}\mu\right). \tag{2}$$

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

$$\mathbb{P}\,(|X - \mu| \ge \delta\mu) \le 2\exp\left(-\frac{\delta^2}{4}\mu\right). \tag{3}$$

## 2 THE PROBABILISTIC PART

From now fix a constant $0 < p < 1$ and consider the binomial Erdős-Rényi graph $\mathbf{G} \sim G(n, p)$.

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

$$\begin{aligned} \mathbf{A} &= \left\{v \in V' : v \sim u_1 \wedge v \sim u_2\right\}, \\ \mathbf{B} &= \left\{v \in V' : v \sim u_1 \wedge v \nsim u_2\right\}, \\ \mathbf{C} &= \left\{v \in V' : v \nsim u_1 \wedge v \sim u_2\right\}, \\ \mathbf{D} &= \left\{v \in V' : x \nsim u_1 \wedge v \nsim u_2\right\}. \end{aligned}$$

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

From (2), with exponentially high probability we have

$$\begin{aligned} |\mathbf{A}| &= (1+o(1))p^2 n, && (4)\\ |\mathbf{B}|, |\mathbf{C}| &= (1+o(1))p(1-p)n, && (5)\\ |\mathbf{D}| &= (1+o(1))(1-p)^2 n. && (6) \end{aligned}$$

It will be convenient to condition by the values of the variables $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$; that is, to condition by an event of the form

$$Q_{A,B,C,D} = \{\mathbf{A} = A, \mathbf{B} = B, \mathbf{C} = C, \mathbf{D} = D\}$$

where $A, B, C, D$ are possible values of $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$. Note that conditioning by $Q_{A,B,C,D}$ does not affect the distribution of $\mathbf{G}[V']$.

The rest of the section is as follows. In subsection 2.1 we introduce a construction that can be used to express logarithmic sets. In subsection 2.2 we show how to express arbitrary subsets of any logarithmic set. In both subsections, all the probabilities and expected values are assumed to be conditioned by $Q_{A,B,C,D}$, where $A, B, C, D$ satisfy the asymptotic equivalences given in (4), (5), (6). Finally, in Subsection 2.3 we apply the law of total probability to obtain non-conditioned results.

## 2.1 Expressing sets of logarithmic size

We construct subsets of $A$ in terms of the edges between $A$ and $B$.

*Definition 2.1.*
(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 $\mathbf{S}_k = \{v \in A : \mathbf{d}_B(v) = k\}$.
(3) For every $x \in A$ let $\mathbf{S}[x] = \mathbf{S}_{\mathbf{d}_B(x)} = \{v \in A : \mathbf{d}_B(v) = \mathbf{d}_B(x)\}$.

*Remark 2.2.* Given a vertex $x \in A$, the statement $v \in \mathbf{S}[x]$ is expressible as a formula in $\mathcal{FO}_=(u_1, u_2, x, v)$:

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

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 $(\mathbf{d}_B(x))_{x \in A}$ are i.i.d. with distribution $\mathrm{Bin}(|B|, p)$.

Theorem 2.3. *Let $c > 0$ be a constant. Then, with exponentially high probability, there exists $0 \le k \le |B|$ such that $|\mathbf{S}_k| = (1 + o(1))c \ln n$.*

To prove Theorem 2.3, we introduce some notations and a lemma.

*Definition 2.4.* Let $n_A = |A|$ and $n_B = |B|$. For every $0 \le k \le n_B$ let $p_k = \mathbb{P}(\mathrm{Bin}(n_B, p) = k)$. Also let $\mu = pn_B$ and $\sigma = \sqrt{p(1-p)n_B}$.

In the following lemma we apply normal approximations to estimate the binomial probabilities $p_k$.

Lemma 2.5. *Let $c > 0$ be a constant. 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 2.5. We apply Theorem 1.2 and Theorem 1.5 from Bollobás [3].

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 Theorem 1.2 from [3] (with $h = t\sigma$ and $n = n_B$),

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

Similarly, applying Theorem 1.5 from [3] (with $h = t\sigma$ and $n = n_B$),

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

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

Proof of Theorem 2.3. Note that $\mathbf{s}_k := |\mathbf{S}_k| \sim \mathrm{Bin}(n_A, p_k)$ for every $0 \le k \le n_B$. 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 $\mathrm{Bin}(n_B, p)$ and let $\mathbf{N} \sim \mathrm{Pois}(n_A)$ 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 $0 \le k \le n_B$ let $\tilde{\mathbf{s}}_k$ count the number of times the value $k$ appears in the process; that is, $\tilde{\mathbf{s}}_k = |\{0 \le i \le \mathbf{N} : \mathbf{d}_i = k\}|$. 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 $\tilde{\mathbf{s}}_k \sim \mathrm{Pois}(n_A p_k)$ for every $k$.

We now apply Lemma 2.5 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

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

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

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

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

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

## 2.2 Expressing arbitrary subsets

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

*Definition 2.7.* For a set $S \subseteq A$, a subset $T \subseteq S$ and a vertex $z \in C$, we say that $z$ *defines* $T$ in $S$ if $T = \{s \in S : s \sim z\}$.

Proposition 2.8. *There exists a positive constant $c_0$ such that the following holds with exponentially high probability. For every set $S \subseteq A$ of size $|S| \leq c_0 \ln n$ and for every subset $T \subseteq S$, there exists $z \in C$ which defines $T$ in $S$.*

Proof. Let $p_0 = \min\{p, 1-p\}$ and choose $c_0$ be a positive constant such that $\gamma := -c_0 \ln p_0 < 1$.

Fix $S \subseteq A$ of size $|S| \leq c_0 \ln n$ and $T \subseteq S$. For a given $z \in C$, it defines $T$ in $S$ with probability

$$p^{|T|}(1-p)^{|S|-|T|} \geq p_0^{|S|} \geq p_0^{c_0 \ln n} = n^{-\gamma}.$$

Crucially, the subsets of $S$ defined by different vertices $z \in C$ are independently distributed. Thus the probability that there exists no $z \in C$ which defines $T$ in $S$ is

$$\left(1 - p^{|T|}(1-p)^{|S|-|T|}\right)^{|C|} \leq \left(1 - n^{-\gamma}\right)^{|C|} = \exp\left(-\Theta(n^{1-\gamma})\right).$$

Taking a union bound over $n^{\Theta(\ln n)}$ possible choices of $S$ and $T$, the probability that exist $T$ and $S$ such that no $z \in C$ defines $T$ in $S$ is

$$n^{\Theta(\ln n)} \cdot \exp\left(-n^{1-\gamma}\right) = \exp\left(\Theta(\ln^2 n) - n^{1-\gamma}\right) = \exp\left(-n^{\Omega(1)}\right).$$

That finishes the proof. □

## 2.3 Non-conditioned results

Finally, we can apply the law of total probability over the events $Q_{A,B,C,D}$ and lose the conditioning by the values of $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$. The following theorem summarizes the probabilistic ingredients required for the proof of Theorem 1.1.

Theorem 2.9. *There exists a positive constant $c_0$ such that, with exponentially high probability, for every set $S \subseteq \mathbf{A}$ of size $|S| \leq c_0 \ln n$ and for every subset $T \subseteq S$ there exists $z \in \mathbf{C}$ which defines $T$ in $S$. Moreover, with exponentially high probability there exists $x \in \mathbf{A}$ such that $|\mathrm{S}[x]| = (1+o(1))\frac{c_0}{2}\ln n$.*

From now on, $c_0$ always refers to the constant promised by Theorem 2.9.

The event described in Theorem 2.9 depends on our initial choice of two vertices $u_1, u_2$. Let us denote it $Q(u_1, u_2)$. The theorem states that $\mathbb{P}(Q(u_1, u_2)) = 1 - \exp\left(-n^{\Omega(1)}\right)$. From symmetry considerations, this probability does not depend on $u_1, u_2$.

# 3 THE LOGICAL PART

We start by recalling the results of Kauffman and Shelah [15] about the expressive power of monadic second order logic in $G(n,p)$ with a constant $p \in (0,1)$. These results are summarized by two main theorems. Theorem 1 states that there exist $\mathcal{MSO}$-formulas $\phi_+, \phi_\times$ which (with high probability) express addition and multiplication operations on the vertices. Theorem 2 demonstrates the expressive power of these formulas by constructing $\mathcal{MSO}$-sentences with complicated sets of subsequential limits.

Another useful fact is mentioned in [15] as a closing remark. It states that unary relations on the vertex set $V = \{1, 2, \ldots, n\}$ can be used to encode binary relations on the subset $\left\{1, 2, \ldots, \lfloor\sqrt{n}\rfloor\right\}$, using the mapping $(i,j) \mapsto i + j\lfloor\sqrt{n}\rfloor$. To state this result more formally, let $\mathcal{BSO}$ denote the class of second order sentences (in the language of graphs), with quantification only over unary and binary relations (here BSO stands for *binary second order*).

Theorem 3.1 (Kauffman-Shelah). *Let $\mathbf{G} \sim G(n,p)$ for a constant $0 < p < 1$ and let $S = \left\{1, 2, \ldots, \lfloor\sqrt{n}\rfloor\right\}$. Then for every $\psi' \in \mathcal{BSO}$ there exists $\psi \in \mathcal{MSO}$ such that, with high probability, $\mathbf{G} \models \psi \iff \mathbf{G}[S] \models \psi'$.*

For the proof of Theorem 1.1, we need a sentence which alternates very slowly between near-0 probabilities and near-1 probabilities. More specifically, we aim for a sentence which refers to size $n$ of the graph, and alternates depending on the modular residue of $\log^* n$. Here $\log^*$ is the iterated logarithm function; its significance comes from its insensitivity to a replacement of $n$ with $\Theta(\ln n)$, which will be necessary later when we restrict to a logarithmic subset. This approach is described in a subsequent paper by Shelah and Spencer [20], and also in Spencer's book [19] (see Subsection 8.3.3), from which we borrow some of the notation.

In $\mathcal{BSO}$, by using the $\mathcal{MSO}$-formulas $\phi_+, \phi_\times$ which (with high probability) express addition and multiplication in $\mathbf{G}$, it is not hard to construct a formula $\mathrm{LogStar}(x,y)$ expressing the equality $y = \log^* x$, and a formula $\mathrm{Mod}(x)$ expressing the property $x \equiv 0, 1, \ldots, 49 \mod 100$ (see [19] for more details).

We now define two $\mathcal{BSO}$-sentences:

(1) Arith′, which states that there exist a binary relation LogStar and a unary relation Mod as above.
(2) BigGap′, which also states the existence of LogStar and Mod, and additionally states that

$$\exists x, y\,[\mathrm{Max}(x) \wedge \mathrm{LogStar}(x, y) \wedge \mathrm{Mod}(y)]$$

where $\mathrm{Max}(x)$ is a formula stating that $x$ is the maximal vertex.

Given that $\phi_+, \phi_\times$ indeed express addition and multiplication in $\mathbf{G}$ (which happens with high probability), we have:

(1) $\mathbf{G} \models \mathrm{Arith}'$ (every finite graph satisfies Arith′).
(2) If $\log^*(n) \equiv 0, 1, \ldots 49 \mod 100$ then $\mathbf{G} \models \mathrm{BigGap}'$.
(3) If $\log^*(n) \equiv 50, 51, \ldots 99 \mod 100$ then $\mathbf{G} \models \neg\mathrm{BigGap}'$.

Now we use Theorem 3.1 to convert the $\mathcal{BSO}$-sentences Arith′ and BigGap′ into $\mathcal{MSO}$-sentences Arith and BigGap. Letting $S = \{1, 2, \ldots, \lfloor\sqrt{n}\rfloor\}$ in $\mathbf{G}$, we have $\mathbf{G}[S] \sim G(\lfloor\sqrt{n}\rfloor, p)$. Therefore, with high probability,

(1) $\mathbf{G} \models \mathrm{Arith}$.
(2) If $\log^* \lfloor\sqrt{n}\rfloor \equiv 0, 1, \ldots 49 \mod 100$ then $\mathbf{G} \models \mathrm{BigGap}$.
(3) If $\log^* \lfloor\sqrt{n}\rfloor \equiv 50, 51, \ldots 99 \mod 100$ then $\mathbf{G} \models \neg\mathrm{BigGap}$.

Also note that $\log^* n - 1 \leq \log^* \lfloor\sqrt{n}\rfloor \leq \log^* n$.

The next step involves another conversion, this time from $\mathcal{MSO}$-sentences on a subset of logarithmic size to $\mathcal{FO}_=$-sentences on the entire graph. As in Section 2, let us arbitrarily fix two vertices $u_1, u_2$, and use them to define (random) vertex sets $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$.

Proposition 3.2. *Let $\psi \in \mathcal{MSO}$. Then there exists a formula $\psi^*(u_1, u_2, x) \in \mathcal{FO}_=(u_1, u_2, x)$ such that, given the event $Q(u_1, u_2)$, for every $x \in \mathbf{A}$ with $|\mathbf{S}[x]| \leq c_0 \ln n$ we have*

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

*Remark 3.3.* When writing $\mathbf{G} \models \psi^*(u_1, u_2, x)$, we implicitly assume that the variable symbols $u_1, u_2, x$ are interpreted as the corresponding vertices $u_1, u_2, x \in V$.

Proof. Given $\psi$, define $\psi^*(u_1, u_2, x)$ 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 $\mathcal{FO}_=(u_1, u_2, x, v)$.
- Convert unary relations: for every unary relation $R$ introduced by $\psi$, 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 as a formula in $\mathcal{FO}(u_1, u_2, z)$.

Given the event $Q(u_1, u_2)$, for every $x \in \mathbf{A}$ with $|\mathbf{S}[x]| \leq c_0 \ln n$, we know that every subset of $\mathbf{S}[x]$ is defined by some $z \in \mathbf{C}$. Therefore

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

as we wanted. □

We shall use Proposition 3.2 with $\psi = \mathrm{Arith}$ and $\psi = \mathrm{BigGap}$; this will give us access to arithmetization over logarithmic sets $\mathbf{S}[x]$.

Another important ingredient is the ability to compare logarithmic sets.

*Definition 3.4.* Let $S, S' \subseteq \mathbf{A}$. We say that $S$ is *pseudo-smaller* than $S'$ if there exists $z \in \mathbf{C}$ which defines a subset $T' \subseteq S'$ such that $|T'| = |S|$.

*Remark 3.5.* Note that is $S$ is pseudo-smaller than $S'$ then $|S| \leq |S'|$. In the other direction, given the event $Q(u_1, u_2)$, if $|S| \leq |S'| \leq c_0 \ln n$ then $S$ is pseudo-smaller than $S'$. Also note that given two vertices $x, x' \in \mathbf{A}$, the statement "$\mathbf{S}[x]$ is pseudo-smaller than $\mathbf{S}[x']$" is expressible as a formula in $\mathcal{FO}_=(u_1, u_2, x, x')$.

We are now ready to complete the proof of Theorem 1.1. In the proof we employ the following notation. For a vertex set $U \subseteq V$ let $\mathbf{E}(U)$ be the set of edges of $\mathbf{G}$ with both endpoint from $U$. For two disjoint vertex sets $U_1, U_2 \subseteq E$ let $\mathbf{E}(U_1, U_2)$ be the set of edges of $\mathbf{G}$ with one endpoint from $U_1$ and the other from $U_2$. Given two vertices $u_1, u_2$, note that the event $Q(u_1, u_2)$ is determined by the values $A, B, C$ of $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and the values of the edge sets $\mathbf{E}(A, B), \mathbf{E}(A, C)$.

Proof of Theorem 1.1. We define Alt as a sentence in $\mathcal{FO}_=$ stating the existence of two vertices $u_1, u_2$ and a vertex $x \in \mathbf{A}$ such that:

(1) $\mathbf{G} \models \mathrm{Arith}^*(u_1, u_2, x)$.
(2) If $x' \in \mathbf{A}$ is another vertex such that $\mathbf{G} \models \mathrm{Arith}^*(u_1, u_2, x')$ then $\mathbf{S}[x]$ is not pseudo-smaller than $\mathbf{S}[x']$.
(3) $\mathbf{G} \models \mathrm{BigGap}^*(u_1, u_2, x)$.

As we have seen, all these statements are indeed expressible in $\mathcal{FO}_=$. We now show that $\lim_{n\to\infty} \mathbb{P}(\mathbf{G} \models \mathrm{Alt})$ does not exist.

First, restrict to a subsequence of $n$ with $\log^* n \equiv 25 \mod 100$. We show that $\mathbb{P}(\mathbf{G} \models \mathrm{Alt}) \to 1$ on this subsequence. Fix $u_1, u_2$ arbitrarily (e.g. $u_1 = 1$ and $u_2 = 2$). We know that $\mathbb{P}(Q(u_1, u_2)) = 1 - o(1)$, so it is sufficient to prove $\mathbb{P}(\mathbf{G} \models \mathrm{Alt} \mid Q(u_1, u_2)) \to 1$ on the subsequence.

Condition by the values $A, B, C$ of $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and by the values of the edge sets $\mathbf{E}(A, B), \mathbf{E}(A, C)$ such that $Q(u_1, u_2)$ holds. The event $Q(u_1, u_2)$ guarantees a vertex $x' \in A$ such that $|\mathbf{S}[x']| = (1 + o(1))\frac{c_0}{2}\ln n$. The conditioning determines the set $S = \mathbf{S}[x']$. The important observation is that the induced subgraph $\mathbf{G}[S]$ depends only on $\mathbf{E}(A)$, and therefore, given the conditioning, its distribution is still binomial, with vertex set of size $(1+o(1))\frac{c_0}{2}\ln n$ and parameter $p$. We deduce that, given the conditioning, $\mathbf{G}[\mathbf{S}[x']] \models \mathrm{Arith}$ with high probability. Applying Proposition 3.2 and the law of total probability over all possible values of $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{E}(A, B), \mathbf{E}(A, C)$, we conclude that given $Q(u_1, u_2)$, with high probability there exists $x' \in \mathbf{A}$ such that $\mathbf{G} \models \mathrm{Arith}^*(u_1, u_2, x')$ and $|\mathbf{S}[x']| = (1 + o(1))\frac{c_0}{2}\ln n$.

Now pick $x \in \mathbf{A}$ such that $\mathbf{G} \models \mathrm{Arith}^*(u_1, u_2, x)$ and $|\mathbf{S}[x]|$ is maximal. Let $m = |\mathbf{S}[x]|$. By definition, $x$ satisfies parts 1 and 2 of Alt. From maximality, $m \geq |\mathbf{S}[x']| = \Omega(\ln n)$ and so $23 \leq \log^* m \leq 25$. Therefore $\mathbf{G}[\mathbf{S}[x]] \models \mathrm{BigGap}$. From Proposition 3.2 we get $\mathbf{G} \models \mathrm{BigGap}^*(u_1, u_2, x)$ and so $x$ also satisfies part 3 of Alt.

Second, restrict to $n$ with $\log^* n \equiv 75 \mod 100$. We show that $\mathbb{P}(\mathbf{G} \models \mathrm{Alt}) \to 0$ on this subsequence. Let $Q = \bigcap_{u_1,u_2} Q(u_1, u_2)$. Theorem 2.9 implies $\mathbb{P}(Q(u_1, u_2)) = 1 - \exp\left(-n^{\Omega(1)}\right)$ for every $u_1, u_2$. A union bound over $\Theta(n^2)$ possible pairs $u_1, u_2$ then shows that $\mathbb{P}(Q) = 1 - o(1)$. We prove that $\mathbf{G} \models \neg\mathrm{Alt}$ given the event $Q$, and thus complete the proof.

Assume that there exist $u_1, u_2, x$ such that parts 1 and 2 of Alt hold. From part 1 and Proposition 3.2 we have $\mathbf{G}[\mathbf{S}[x]] \models \mathrm{Arith}$.

From part 2 and the assunption that $Q(u_1, u_2)$ holds we again deduce that $m := |\mathrm{S}[x]| = \Omega(\ln n)$. Therefore $73 \leq \log^* m \leq 75$. Combining both facts we deduce $\mathbf{G}\left[\mathrm{S}[x]\right] \models \neg\mathrm{BigGap}$, which implies $\mathbf{G} \models \neg\mathrm{BigGap}^*(u_1, u_2, x)$ from Proposition 3.2. Therefore $x$ does not satisfy part 3 of Alt. $\square$

## 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. Properties of almost all graphs and generalized quantifiers. *Fundam. Inform.*, 98(4):351–372, 2010.

[6] H.-D Ebbinghaus. *Extended Logics: The General Framework*, pages 25–76. Cambridge University Press, March 2017.

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

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

[9] 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.

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

[11] 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).

[12] Simi Haber and Saharon Shelah. Random graphs and Lindström quantifiers for natural graph properties. *Submitted*. HbSh:986 in Shelah's archive.

[13] 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.

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

[15] Matt Kauffman and Saharon Shelah. On random models of finite power and monadic logic. *Discrete Mathematics*, 54:285–293, 1985.

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

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

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

[19] Joel Spencer. *The Strange Logic of Random Graphs*. Springer, 2001.

[20] Joel Spencer and Saharon Shelah. Zero-one laws for sparse random graphs. *Journal of the American Mathematical Society*, 1:97–115, 1988.