

Efficiently Constructible Huge Graphs that Preserve First Order Properties of Random Graphs

Moni Naor* Asaf Nussboim†

Eran Tromer

Department of Computer Science and Applied Mathematics,

Weizmann Institute of Science, Rehovot 76100, Israel

{moni.naor, asaf.nussbaum, eran.tromer}@weizmann.ac.il

April 9, 2008

Abstract

Consider the problem of producing random graphs (as inputs for some graph algorithm) when the order of the graphs N is huge, say $N = 2^n$. While the canonical random graphs $\mathcal{G}(N, p)$ are too expensive to sample and to store, we construct efficiently computable random-looking graphs that faithfully emulate $\mathcal{G}(N, p)$ graphs w.r.t. arbitrary first-order (FO) properties. FO properties are graph properties that can be expressed by a formula ϕ where variables stand for vertices and the only relations are equality and adjacency (e.g. having an isolated vertex is a FO property: $\exists x \forall y (\neg \text{EDGE}(x, y))$). Random graphs are known to exhibit remarkable structure w.r.t. FO properties, as indicated by the famed 0/1-law: for a variety of choices of $p = p(N)$, any *fixed* FO property ϕ holds for $\mathcal{G}(N, p)$ with probability tending either to 0 or to 1 as N grows to infinity.

We show that similar 0/1-laws are satisfied by $\mathcal{G}(N, p)$ even w.r.t. sequences of formulas $\{\phi_N\}$ with bounded quantifier depth, $\text{depth}(\phi_N) \leq D^* = \frac{\log(N)}{\log(1/p)}$ but no longer w.r.t. $\text{depth-}2D^*$ properties. We then efficiently emulate $\mathcal{G}(N, p)$ by constructing graphs that satisfy similar $\Theta(D^*)$ -0/1-laws. We provide both probabilistic constructions (that exhibit other random-looking qualities such as k -wise independence and computational indistinguishability from $\mathcal{G}(N, p)$), and deterministic constructions (producing a single random-looking graph per each graph size). Finally, we show that no efficiently computable graphs can simultaneously capture all $\text{poly}(\log(N))$ -depth properties of random graphs, not even under much weaker notions of resemblance than satisfying appropriate 0/1-laws.

*Partly supported by a grant from the Israel Science Foundation.

†Partly supported by the Minerva Foundation 2-8495.

1 Introduction

We deal with (cheap) random-looking graph-distributions that resemble some desired (yet prohibitively expensive) truly random distribution. Throughout, graphs are assumed to be huge, so efficient algorithms run in $\text{poly}(n)$ bounded resources, whereas the number of vertices is, say, $N = 2^n$. Huge graphs like that are not represented explicitly, but rather by an efficient procedure that evaluates edge-queries using a succinct representation (a seed) of the graph. Such random looking distributions are sampled by randomly picking the seed.

We attempt to capture a large class of properties of the canonical truly random graphs $\mathcal{G}(N, p)$ where N vertices are fixed and the edges are independently picked each with probability $p = p(N)$. A prominent class of properties is that of first-order (FO) properties, namely those that can be expressed by a formula ϕ in the canonical language where variables stand for vertices and the only relations are equality and adjacency (e.g containing a triangle is a FO property of quantifier depth 3 written as $\exists x \exists y \exists z (\text{EDGE}(x, y)) \wedge (\text{EDGE}(x, z)) \wedge (\text{EDGE}(y, z))$). Random graphs are known to exhibit remarkable structure w.r.t. FO properties, namely the famed 0/1-law: any fixed FO property ϕ holds for $\mathcal{G}(N, p)$ with probability tending either to 0 or to 1 as N grows to infinity. Thus one can view this work as dealing with graphs that look random to distinguishers that are expressible as FO properties.

We demonstrate that for sufficiently large k , any *exact* k -wise independent graphs (defined below) preserve the 0/1-law of random graphs (this is not true for *almost* k -wise independent graphs). We also provide a construction of computationally pseudorandom graphs that satisfy this 0/1-law (in general, computational pseudorandomness does not imply preserving such properties). Finally, for each graph size a single efficiently computable ‘random-looking’ graph is provided that preserves the FO properties of random graphs. Those results can be extended to FO properties of quantifier depth up to $\Theta(\frac{\log(N)}{\log(1/p)})$. On the other hand, it is shown that no efficiently computable graphs can simultaneously capture all $\text{poly}(\log(N))$ -depth properties of random graphs, not even under much weaker notions of resemblance than satisfying appropriate 0/1-laws.

1.1 Background on Random-Looking Graphs

We survey several characterizations for the notion of a ‘random-looking’ graph.

Combinatorial pseudorandom graphs. This term refers here to a collection of definitions that consider a single graph G_N per each size N and intend to capture the edge distribution of $\mathcal{G}(N, p)$ by requiring that any induced subgraph of G_N has density $\approx p$. One variant is Thomason’s *jumbledness* [27], where for each vertex set U , $|e(U) - p\binom{|U|}{2}| \leq \alpha|U|$ should hold (here $e(U)$ counts the internal edges of U). It

is well known that $\alpha = \Theta(\sqrt{pN})$ is the best-possible accuracy-parameter and it is achieved by $\mathcal{G}(N, p)$. A weaker (yet highly influential) definition is *quasirandomness*, which requires only that $\forall U |e(U) - p\binom{|U|}{2}| \leq o(N^2)$. Several efficient constructions of quasirandom (and jumbled) graphs were provided in last decades, and they are known to exhibit some prominent properties of $\mathcal{G}(N, p)$ graphs, leading to a variety combinatorial applications (see a survey by Krivelevich and Sudakov [20]).

$k(N)$ -wise independent graphs. These graphs (first considered in [16, 21]) are a natural relaxation of $\mathcal{G}(N, p)$ where again each edge appears w.p. $p = p(N)$, but it is only required that the distribution of any fixed $k(N)$ potential edges should be mutually independent. Efficient constructions of n^c -wise independent graphs are known for any fixed c and a wide variety of densities $p(N)$. These constructions immediately follow from efficient constructions for $k(N)$ -wise independent bits due to Joffe [19], and Alon, Babai and Itai [1] - the appearance of each potential edge is simply decided by a single random bit (that has probability of success $p(N)$).

Computationally pseudorandom graphs. These are defined [16, 21] as graphs which are computationally indistinguishable from random graphs in the sense of Goldreich, Goldwasser and Micali [15]. Namely, no $poly(n)$ -time distinguishing algorithm that performs edge-queries of its choice can tell apart a pseudorandom graph from a random $\mathcal{G}(N, p)$ graph. Again, computationally pseudorandom graphs are immediately given from efficient constructions of Boolean pseudorandom functions, PRF (where the value of the function on each specific input decides a specific edge). Such PRF are known to exist iff one-way functions (OWF) exist [15][18]. Thus, whenever computational pseudorandomness is mentioned, the existence of OWF is assumed.

Graphs that preserve specific combinatorial properties of random graphs. Random graphs are known to exhibit a remarkable combinatorial structure (see Bollobás' book [6]). For instance, consider $\mathcal{G}(N, \frac{1}{2})$ which is the uniform distribution on all N -vertices graphs. Then for some $s(N) = (2 - o(1)) \log(N)$, it holds that with overwhelming probability $\mathcal{G}(N, \frac{1}{2})$ is:

1. Hamiltonian and thus connected, and contains a perfect matching.
2. Has clique number and independence number precisely $s(N) \pm 1$.
3. Has chromatic number precisely $\frac{N}{s(N)}(1 \pm o(1))$.
4. Has connectivity number precisely $\frac{1}{2}N(1 \pm o(1))$.
5. Has maximal and minimal degree precisely $\frac{1}{2}N(1 \pm o(1))$.

At least some of these properties are met by $\text{poly}(\log(N))$ -wise independent graphs, and by combinatorial pseudorandom graphs. However, it can be shown that computational pseudorandomness does not imply such combinatorial properties. Indeed, [16][21] show specific computationally pseudorandom graphs that boldly defy properties 1–5 above. Conversely, [16][21] provide efficient constructions of graphs which are simultaneously: computationally pseudorandom (w.r.t. $\mathcal{G}(N, \frac{1}{2})$), almost n^c -wise independent, preserve properties 1–3 above, and approximate properties 4 and 5.

Our work. While the constructions of [16][21] are tailor-made to preserve a small number of *prescribed* prominent properties, the current work deals with capturing *arbitrary* FO properties of random graphs. The strength of these FO properties is incomparable w.r.t. the properties studied in [16][21]. To present our results we always have some order-function $N : \mathbb{N} \rightarrow \mathbb{N}$ at the background, which decided the number of vertices $N = N(n)$ per each n . Invariably, $n^{\omega(1)} \leq N \leq 2^{\text{poly}(n)}$, where the first \leq makes the production of $\mathcal{G}(N, p)$ graphs non-trivial, and the second \leq enables efficient constructions. We let $p = p(N) = p(N(n))$, $D = D(N) = D(N(n))$, $k = k(N) = k(N(n))$, and denote sequences of distributions on N -vertex graphs by $\mathcal{G} = \{\mathcal{G}_N\} = \{\mathcal{G}_{N(n)}\}_{n \in \mathbb{N}}$. Sequences of FO formulas are denoted by $\Phi = \{\phi_N\} = \{\phi_{N(n)}\}_{n \in \mathbb{N}}$.

1.2 Preserving First-Order Properties of Random Graphs

First-order properties are graph properties that can be expressed in FO canonical language, where the variables stand for vertices and the only relations are equality and adjacency. For instance, having an isolated vertex can be written as $\exists x \forall y \neg \text{EDGE}(x, y)$; see Section 2 for definitions.

From the FO lens, random graphs exhibit a remarkable structure (see Spencer’s [23] for an excellent survey). The following 0/1-law is known to hold for $\mathcal{G}(N, p)$: every FO property ψ holds with probability tending either to 0 or to 1 as the size of the graph grows to infinity. The case where p is constant is due to Fagin [12] and independently to Glebskii, Kogan, Liagonkii and Talanov [14]. The other known case where $p(N) = N^{-\alpha}$ for an irrational α is due to Shelah and Spenser [26].

Can one efficiently construct random-looking graphs that resemble $\mathcal{G}(N, p)$ and satisfy this 0/1-law? The answer is positive, but we shall actually consider graphs that meet a much stronger requirement.

Generalized 0/1-Laws. Rather than fixing a single FO formula, we shall consider sequences of formulas $\Phi = \{\phi_N\}$. Such a sequence can specify much richer properties than a single formula. For instance, containing a clique of size $f(N)$ can be expressed by the sequence where $\phi_N = \exists x_1 \dots \exists x_{f(N)} \bigwedge_{i \neq j} ((x_i \neq x_j) \wedge \text{EDGE}(x_i, x_j))$, and the

quantifier depth is $\text{depth}(\phi_N) = f(N)$ (quantifier depths are formally defined in Section 2). We will mainly consider the quantifier depth of ϕ_N , rather than the entire length of ϕ_N ; this choice of complexity measure will be well-motivated by the discussed results.

We will define the $D(N)$ -0/1-law as a natural generalization of the basic 0/1-law. Roughly speaking, it means that for any sequence $\Phi = \{\phi_N\}$ with quantifier depth $\text{depth}(\phi_N) \leq D(N)$ it holds that $\lim_{N \rightarrow \infty} \Pr[\mathcal{G}_N \models \phi_N] \in \{0, 1\}$. The precise definition, which addresses some technicalities, is given in Section 2.

Equivalence. Let $\mathcal{G}^1 = \{\mathcal{G}_N^1\}$ and $\mathcal{G}^2 = \{\mathcal{G}_N^2\}$ be two sequences of distributions over graphs. We wish to formalize the notion of \mathcal{G}^1 preserving the FO properties of \mathcal{G}^2 (of special interest is the case where $\mathcal{G}_N^2 = \mathcal{G}(N, p)$). Having a 0/1-law hold for both \mathcal{G}^1 and \mathcal{G}^2 may not suffice as it might be the case that $\Pr[\mathcal{G}_N^1 \models \phi_N] \xrightarrow{N \rightarrow \infty} 1$, whereas $\Pr[\mathcal{G}_N^2 \models \phi_N] \xrightarrow{N \rightarrow \infty} 0$. Therefore the following definition is introduced: \mathcal{G}^1 and \mathcal{G}^2 are said to be weakly $D(N)$ -equivalent, if for any sequence Φ with quantifier depth $\text{depth}(\phi_N) \leq D(N)$, it holds that $\lim(\Pr[\mathcal{G}_N^1 \models \phi_N] - \Pr[\mathcal{G}_N^2 \models \phi_N]) \xrightarrow{N \rightarrow \infty} 0$. Note that weak equivalence enables to discuss the resemblance of \mathcal{G}^1 and \mathcal{G}^2 , even when the 0/1-law no longer holds (say when $\Pr[\mathcal{G}_N^j \models \phi_N] \xrightarrow{N \rightarrow \infty} 1/3$ for both $j = 1, 2$).

Weak equivalence is used only to pronounce our impossibility results, while our positive results relate to the significantly more powerful notion of strong $D(N)$ -equivalence (equivalence for short). Strong equivalence means that as we (independently) sample one graph from \mathcal{G}_N^1 and another from \mathcal{G}_N^2 , then the chance that these 2 graphs disagree on *even a single* depth $D(N)$ formula is vanishing. Note that (strong) equivalence between any \mathcal{G}^1 and \mathcal{G}^2 implies the $D(N)$ -0/1-law for both \mathcal{G}^1 and \mathcal{G}^2 . Yet, we usually redundantly stress meeting the 0/1-law, even when strong equivalence is achieved.

1.3 Our Results

Maximal 0/1-laws for random graphs. We start by establishing the maximal depth for which 0/1-laws hold for random graphs. For some $D^* = \frac{(1-o(1)) \log(N)}{\log(1/p(N))}$, the following holds: For any density $p = p(N)$,¹ $\mathcal{G}(N, p)$ satisfies the D^* -0/1-law, whereas for a large variety of densities, the $2D^*$ -0/1-law is strongly violated.

Probabilistic constructions. For D^* as above, we show that arbitrary k -wise independent graphs satisfy the D^* -0/1-law and are D^* -equivalent to $\mathcal{G}(N, p)$ whenever $k \gg (D^*)^2$. Thus, k -wise independent graphs provide the best possible emulation of random graphs w.r.t. FO properties: indeed, even $\mathcal{G}(N, p)$ graphs cannot achieve

¹Throughout this subsection we assume that $p = p(N) \leq \frac{1}{2}$. Otherwise each term $p(N)$ concerning quantifier depths should be replaced by $\min\{p(N), 1 - p(N)\}$.

$2D^*$ -equivalence to themselves (because 0/1-laws no longer hold for such depths). Assuming that one-way functions exist, it is easy to efficiently achieve computational indistinguishability from $\mathcal{G}(N, p)$ simultaneously with k -wise independence (and optimal equivalence). This is done by combining known constructions of pseudorandom graphs and k -wise independent graphs.

Deterministic construction (Paley graphs). We show that for every n and p there exists a single efficiently computable Paley graph of size $N = 2^{\Theta(n)}$ and edge density $p' = p \pm \epsilon$, which is $D(N)$ -equivalent to $\mathcal{G}(N, p')$. Here $D(N)$ depends on ϵ ; for example, for any $\epsilon(N) \geq \Theta(1/\log(N))$ we can obtain $D(N) \geq \Theta(\frac{\log(N)}{\log \log(N)})$. For the special case $p = \frac{1}{2}$ the edge density is exactly p and $D(N) \sim \frac{\log(N)}{2}$ which is optimal up to a factor of $4 + o(1)$. Paley graphs were already known to maintain quite a few qualities of random graphs, and in particular, our techniques are a (non-trivial) generalization of Graham and Spencer's analysis for the directed version of Paley graphs with $p = \frac{1}{2}$ [17]. Our positive results regarding both probabilistic and deterministic constructions rely on powerful sufficient conditions for 0/1-laws due to Fagin [12] and Spencer [23].

Impossibility results. While the above positive results are nearly optimal (w.r.t. strong-equivalence), one may still desire to achieve weak equivalence to random graphs, when $D(N)$ is too large for $D(N)$ -0/1-laws to hold. In this context we obtain the following impossibility result: any efficiently constructed graphs $\mathcal{G} = \{\mathcal{G}_N\}$ with seed length $m = m(N)$ are never even weakly $D(N)$ -equivalent to $\mathcal{G}(N, \frac{1}{2})$ for some $D(N) = \Theta(\sqrt{m} + \log(N))$. Separating formulas that have not only small depth but also small $\text{poly}(m, \log(N))$ total length are also provided. Analogous results can be obtained for a wide range of densities p . These results rule out the possibility of efficiently producing graphs that simultaneously capture all $\text{poly}(\log(N))$ -depth properties of random graphs, under any reasonable notion of resemblance.

1.4 Comparing Alternative Notions of Pseudorandomness

We now compare all aforementioned notions of pseudorandomness: $\text{poly}(\log(N))$ -wise independence, computational pseudorandomness, optimal $\Theta(\sqrt{pN})$ -jumbledness and optimal FO-equivalence (strong $\Theta(\frac{\log(N)}{\log(1/p)})$ -equivalence to $\mathcal{G}(N, p)$). We will also mention (the relaxed notion of) almost k -wise independence, a.k.a. (k, ϵ) -wise independence. Whereas k -wise independence means that the joint distribution of any k potential edges is identical to corresponding distribution \mathcal{D} induced by $\mathcal{G}(N, p)$; then almost k -wise independence only requires to achieve statistical distance $\epsilon = o(1)$ from \mathcal{D} . Interestingly, while no notion implies all the others, a single construction can simultaneously capture all four. All of the following arguments apply to arbitrary $n^{\omega(1)} \leq N \leq 2^{\text{poly}(n)}$ and $1/\text{poly}(\log(N)) \leq p \leq 1$.

Computational pseudorandomness and k -wise independence are incomparable by the following elementary facts. Any k -wise independent graphs provided using the standard construction via polynomials of degree k , are easily distinguished from random graphs using only $k + 1$ edge queries. On the other hand, any computable graphs with seed length m (and in particular any computationally pseudorandom graphs) are statistically far from all (k, ϵ) -wise independent graphs whenever $k \gg \frac{m}{p^3}$.

Next, since jumbledness, quasirandomness and FO-equivalence may hold even for a single graph per size, they cannot guarantee neither (k, ϵ) -wise independence nor computational pseudorandomness. In the other direction, for random looking distributions $\{\mathcal{G}_N\}$ to achieve jumbledness (or quasirandomness) it means that as we sample a single graph G_N independently from each distribution \mathcal{G}_N , then with probability 1 the sequence $\{G_N\}$ should be jumbled (or quasirandom).

One of the main results in [2] shows that $\log(N)$ -wise independence guarantees optimal jumbledness whenever $p \geq \Theta(\frac{\log(N)}{N})$. In contrast, computational pseudorandomness (and almost k -wise independence) fail to imply jumbledness because by [16, 21] whenever $p \gg \frac{n^{\omega(1)}}{N}$ then huge cliques can be forced into the resulting graphs while retaining computational pseudorandomness (and almost k -wise independence). Yet, (the much weaker) quasirandomness condition is implied by computational pseudorandomness. The latter implication holds because having the ‘correct’ $(pN)^4(1 \pm o(1))$ number of labeled 4-cycles is (surprisingly) equivalent to quasirandomness (by Chung, Graham and Wilson [9]), and this 4-cycles condition readily follows from computational pseudorandomness as long as $p \geq 1/\text{poly}(n)$.

Next, computational pseudorandomness, optimal jumbledness and almost k -wise independence all fail to imply even depth-2 equivalence to $\mathcal{G}(N, p)$. Indeed, isolated vertices can be forced into the graphs while preserving any of these random looking criteria. In the other direction, optimal FO-equivalence does not guarantee quasirandomness either. For instance, one can partition the vertex set into 2 equal size sets V_1, V_2 s.t. the edges connecting these sets form a random bipartite graph with density p while the subgraphs induced by V_1 and by V_2 are random $G(\frac{N}{2}, p(1 + \Delta))$ and $G(\frac{N}{2}, p(1 - \Delta))$ graphs, respectively. These graphs are clearly not quasirandom whenever Δ is constant, whereas their optimal FO-equivalence follows analogously to Theorems 2 and 3.

Finally, we show that $\omega\left(\frac{\log^2(N)}{\log^2(1/p(N))}\right)$ -wise independence guarantees optimal FO-equivalence. It is well known that k -wise independence can be easily strengthened to maintain computational pseudorandomness as well - resulting in a single construction that simultaneously meets all four criteria of random-lookingness.

2 Preliminaries

2.1 First Order Logic on Graphs

Formally, the alphabet of FO (first-order) logic on graphs is made of:

1. Infinitely many variable symbols such as ‘ x ’, ‘ y ’, ‘ z ’ .
2. The binary relation symbols ‘ $=$ ’ and ‘EDGE’.
3. The quantifier symbols ‘ \forall ’ and ‘ \exists ’, the connective symbols ‘ \neg ’, ‘ \vee ’, ‘ \wedge ’, and the signs ‘(’ and ‘)’.

A FO formula is a formula written in graphs’ FO logic. A FO property is a graph property that can be expressed by a FO formula where the variables x, y, z stand for vertices, ‘ $=$ ’ stands for equality and ‘EDGE’ stands for adjacency.

The quantifier depth $depth(\phi)$ of a formula ϕ is defined inductively:

1. For atomic expressions, $depth(x = y) = depth(EDGE(x, y)) = 0$.
2. $depth(\neg\phi) = depth((\phi)) = depth(\phi)$.
3. $depth(\phi \vee \psi) = depth(\phi \wedge \psi) = \max\{depth(\phi), depth(\psi)\}$
4. $depth(\exists x\phi) = depth(\forall x\phi) = depth(\phi) + 1$.

For instance, the property of being either an empty graph or containing a triangle is a FO property that can be expressed by the following formula of quantifier depth 3: $(\forall u\forall v\neg EDGE(u, v)) \vee (\exists x\exists y\exists z (EDGE(x, y) \wedge (EDGE(x, z)) \wedge (EDGE(y, z)))$.

2.2 Asymptotics

Throughout, let some $N : \mathbb{N} \rightarrow \mathbb{N}$ decide the number of vertices $N = N(n)$ per each n . Our results are most interesting when $n^{\omega(1)} \leq N \leq 2^{poly(n)}$, but actually, given the dependency of the parameters p, D, k on N , then most claims hold for any $N \xrightarrow{n \rightarrow \infty} \infty$ (say, when $N = n$ and the graphs are not huge). To emphasize this fact we typically use the notation $p(N), D(N), k(N), \{\mathcal{G}_N\}, \{\phi_N\}$.

2.3 Distributions on Huge Graphs

Definition 1 (distributions on huge graphs) *Let $\ell : \mathbb{N} \rightarrow \mathbb{N}$ be a $poly(n)$ -bounded length function. Distributions on huge graphs with vertex sets $\{V_n\}_{n \in \mathbb{N}}, V_n \subseteq \{0, 1\}^{\ell(n)}$ are a sequence of distributions $\mathcal{G} = \{\mathcal{G}_N\} = \{\mathcal{G}_{N(n)}\}_{n \in \mathbb{N}}$, where each \mathcal{G}_N is taken over the set of simple, labeled undirected graphs over V_n .*

For our probabilistic constructions the vertex sets are often $V_n = \{0, 1\}^n$. For our deterministic constructions the distributions \mathcal{G}_N are degenerate (i.e., have support of size 1), and possibly $V_n \subsetneq \{0, 1\}^{\ell(n)}$.

Definition 2 (efficiently constructible distributions on huge graphs) *Let $\ell_1, \ell_2 : \mathbb{N} \rightarrow \mathbb{N}$ be $\text{poly}(n)$ -bounded length functions. Distributions on huge graphs $\mathcal{G} = \{\mathcal{G}_N\}$ with vertex sets $\{V_n\}_{n \in \mathbb{N}}$, $V_n \subseteq \{0, 1\}^{\ell_1(n)}$ are efficiently constructible if the following holds.*

- *Efficient Indexing: the support of each \mathcal{G}_N can be indexed as $\{G_s\}_{s \in \{0, 1\}^{\ell_2(n)}}$ (the string s is the index of the graph G_s).*
- *Efficient Evaluation: there exists a deterministic polynomial-time evaluation algorithm E s.t. for any index $s \in \{0, 1\}^{\ell_2(n)}$, and for any vertex pair $\{u, v\} \in V_n$, it holds that $E(s, u, v) = 1$ when the edge $\{u, v\}$ appears in G_s , and $E(s, u, v) = 0$ otherwise.*
- *Identity of Distributions: the graphs-distribution induced by $E(s, \cdot, \cdot)$ when s is uniformly taken from $\{0, 1\}^{\ell_2(n)}$ is identical to \mathcal{G}_N .*

Definition 3 (K-wise independent graphs) *Let $p : \mathbb{N} \rightarrow (0, 1)$, and $k : \mathbb{N} \rightarrow \mathbb{R}^+$. Distributions on huge graphs $\mathcal{G} = \{\mathcal{G}_N\}$ are $(p(N), k(N))$ -wise independent if in \mathcal{G}_N every potential edge appears w.p. $p(N)$, and the distribution of any $k(N)$ potential edges is mutually independent.*

Computational indistinguishability. Computational indistinguishability between distributions on huge graphs is defined exactly like (standard) computational indistinguishability between distributions over functions, with function evaluation replaced by graph edge queries (For more details consult [16, 21]).

2.4 New Definitions: Generalized 0/1-Laws and Equivalence

We suggest the $D(N)$ -0/1-law as a natural generalization of the basic 0/1-law. For the current discussion, let's consider the specific case where $\mathcal{G} = \{\mathcal{G}_N\}_{N \in \mathbb{N}}$ and $\Phi = \{\phi_N\}_{N \in \mathbb{N}}$ (we return to the more general case $\mathcal{G} = \{\mathcal{G}_{N(n)}\}_{n \in \mathbb{N}}$ and $\Phi = \{\phi_{N(n)}\}_{n \in \mathbb{N}}$ in the formal definition). As a naive first candidate, consider the following definition: distributions on huge graphs \mathcal{G} satisfy the $D(N)$ -0/1-law if for any sequence Φ with quantifier depth $\text{depth}(\phi_N) \leq D(N)$ it holds that

$$\lim_{N \rightarrow \infty} \Pr[\mathcal{G}_N \models \phi_N] \in \{0, 1\}. \quad (1)$$

However, this definition is never satisfiable: given any sequence Φ satisfying condition (1), consider the sequence Ψ obtained from Φ by negating all formulas for even N (for

odd N formulas are kept intact). Clearly, Ψ has the same quantifier depth as Φ , but condition (1) no longer holds for Ψ . Because of this, the definition is weakened by requiring (instead of the above) that for each sequence satisfying $\text{depth}(\phi_N) \leq D(N)$ there should exist a similar sequence $\Phi' = \{\phi'_N\}$ s.t. $\phi'_N \in \{\phi_N, \neg\phi_N\}$, and

$$\lim_{N \rightarrow \infty} \Pr[\mathcal{G}_N \models \phi'_N] = 1. \quad (2)$$

Unfortunately, with this weakened definition $D(N)$ -0/1-laws no longer imply the basic 0/1-law. For instance, if \mathcal{G}_N is (almost surely) the empty graph for even N , and the complete graph for odd N , then the basic 0/1-law no longer holds although condition (2) is satisfied. Our following final definition contains a specific condition (item 2) solely designed to guarantee that the basic 0/1-law is implied by $D(N)$ -0/1-laws.

Definition 4 (D(N)-0/1-law) *Let $\mathcal{G} = \{\mathcal{G}_N\}$ be distributions on huge graphs, and let $D : \mathbb{N} \rightarrow \mathbb{N}$. The $D(N)$ -0/1-law holds for \mathcal{G} if for any sequence of formulas $\Phi = \{\phi_N\}$ with quantifier depth $\text{depth}(\phi_N) \leq D(N)$ the following holds:*

- *There exist a sequence $\Phi' = \{\phi'_N\}$, such that $\phi'_N \in \{\phi_N, \neg\phi_N\}$, and $\Pr[\mathcal{G}_N \models \phi'_N] \xrightarrow{N \rightarrow \infty} 1$.*
- *For any fixed formula ϕ the limit $\lim_{N \rightarrow \infty} \Pr[\mathcal{G}_N \models \phi]$ exists.*

We next define equivalence between distributions on huge graphs; see Section 1.2 for motivation.

Definition 5 (Weak equivalence) *Let $D : \mathbb{N} \rightarrow \mathbb{N}$, and let $\mathcal{G}^1 = \{\mathcal{G}_N^1\}$, $\mathcal{G}^2 = \{\mathcal{G}_N^2\}$ be distributions on huge graphs. Then \mathcal{G}^1 and \mathcal{G}^2 are weakly $D(N)$ -equivalent if for any sequence of formulas $\Phi = \{\phi_N\}$ with quantifier depth $\text{depth}(\phi_N) \leq D(N)$ it holds that $\lim(\Pr[\mathcal{G}_N^1 \models \phi_N] - \Pr[\mathcal{G}_N^2 \models \phi_N]) \xrightarrow{N \rightarrow \infty} 0$.*

Definition 6 (Strong equivalence) *Let $D : \mathbb{N} \rightarrow \mathbb{N}$, and let \mathbb{D}_N denote the set of formulas with depth $\leq D(N)$. Let $\mathcal{G}^1 = \{\mathcal{G}_N^1\}$, $\mathcal{G}^2 = \{\mathcal{G}_N^2\}$ be distributions on huge graphs. Then \mathcal{G}^1 and \mathcal{G}^2 are (strongly) $D(N)$ -equivalent if for G_N^1, G_N^2 independently sampled from $\mathcal{G}_N^1, \mathcal{G}_N^2$, it holds that $\Pr[\exists \phi_N \in \mathbb{D}_N (G_N^1 \models \phi \wedge G_N^2 \not\models \phi)] \xrightarrow{N \rightarrow \infty} 0$.*

3 Extension properties and 0/1-laws

We now describe extension properties. Relying on Spencer's variant of Fagin's proof for the basic 0/1-law (cf. [12, 23]), extension properties will be used for establishing $D(N)$ -0/1-laws and equivalence for both random and k -wise independent graphs.

Definition 7 (Extension properties)

- A single graph G achieves the t -extension property, \mathbf{P}_t , if for all distinct vertices v_1, \dots, v_t and any bits b_1, \dots, b_t there exists an extending vertex $u \notin \{v_1, \dots, v_t\}$ s.t. the edge $\{u, v_i\}$ appears in G iff $b_i = 1$.
- Let $T : \mathbb{N} \rightarrow \mathbb{N}$. Distributions on huge graphs $\mathcal{G} = \{\mathcal{G}_N\}$ achieve the $T(N)$ -extension property if $\Pr[\mathcal{G}_N \models \mathbf{P}_{T(N)}] \xrightarrow{N \rightarrow \infty} 1$.

We now state the sufficiency of $D(N)$ -extension to $D(N)$ -0/1-laws and to $D(N)$ -equivalence. Although Spencer considers only the case of a single formula (rather than a sequence of formulas), Theorems 1 and 2 are actually implicit in [23, Sec. 2.5].

Theorem 1 *Let \mathcal{G} be distributions on huge graphs, and let $D : \mathbb{N} \rightarrow \mathbb{N}$ be an arbitrary increasing function. If \mathcal{G} achieves $D(N)$ -extension, then \mathcal{G} satisfies the $D(N)$ -0/1-law.*

Theorem 2 *Let \mathcal{G}^1 and \mathcal{G}^2 be distributions on huge graphs, and let $D : \mathbb{N} \rightarrow \mathbb{N}$ be an arbitrary increasing function. If both \mathcal{G}^1 and \mathcal{G}^2 achieve $D(N)$ -extension, then \mathcal{G}^1 and \mathcal{G}^2 are $D(N)$ -equivalent.*

This motivates studying the maximal extension T^* achieved by random graphs.

Theorem 3 *For arbitrary $p : \mathbb{N} \rightarrow (0, 1)$, let $p'(N) = \min\{p(N), 1 - p(N)\}$, and let $T^* = \frac{\log(N)}{\log(1/p'(N))}$. Then there exist $D_1, D_2 = T^*(1 \pm o(1))$ s.t. $\mathcal{G}(N, p)$ achieves D_1 -extension but fails to achieve D_2 -extension.*

Proof. It turns out that random graphs and k -wise independent ones achieve the same (asymptotic) maximal extension. As random graphs are in particular k -wise independent, it suffices to prove the claim for arbitrary $(k(N), p(N))$ -wise independent graphs, with sufficiently large $k = k(N)$. We assume w.l.o.g. that $p(N) \leq \frac{1}{2}$ so $p'(N) = p(N)$.

Proving the upper-bound. Here we require $k(N) \geq D_2(N)$. Let $k = k(N)$, $d = D_2(N)$. Fix arbitrary distinct vertices v_1, \dots, v_d , and consider the extension-requirement that the vertex u should be connected to all v_i s. For any u , we have $\Pr[u \text{ is an extending vertex}] = p^d$. Applying a union bound gives $\Pr[\exists \text{ an extending vertex}] \leq Np^d$ which is $\leq 1/\log(N)$ when $d = \log[N \log(N)]/\log(1/p) = T^*(1 + o(1))$.

Proving the lower-bound. Here we require $k(N) \gg (T^*)^2$. Let $p' = p'(N)$, $k = k(N)$, $r = r(N)$, $d = D_1(N)$ and $\epsilon = \epsilon(N)$, where

$$\epsilon \stackrel{\text{def}}{=} \frac{\log^2(N)}{k \log^2(1/p')}, \quad (3)$$

$$r \stackrel{\text{def}}{=} N^{6\epsilon},$$

$$d \stackrel{\text{def}}{=} \frac{\log(N/r)}{\log(1/p')}. \quad (4)$$

Assume w.l.o.g. that $p' > 1/\sqrt{N}$ (otherwise $d \leq 2$ and there is nothing to prove). Note that $r = N^{o(1)}$, so as required $d = (1 - o(1)) \frac{\log(N)}{\log(1/p')} = (1 - o(1))T^*$.

We first upper-bound the probability that for some fixed distinct vertices v_1, \dots, v_d and bits b_1, \dots, b_d , the graph fails to exhibit an extending vertex u s.t. $\forall i, \text{EDGE}(u, v_i) \Leftrightarrow (b_i = 1)$. We later conclude with an upper bound over all possible choices of v_i s and b_j s. Let \mathbb{U} denote the set of all vertices excluding v_1, \dots, v_d . Consider the $N - d$ random variables $\{X_u\}_{u \in \mathbb{U}}$, where $X_u = 1$ iff u is an extending vertex and $X_u = 0$ otherwise. Thus, no extending vertex exists iff $X \stackrel{\text{def}}{=} \sum_{u \in \mathbb{U}} X_u = 0$. As the edges of the graph are k -wise independent, then the variables X_u are \tilde{k} -wise independent for $\tilde{k} \stackrel{\text{def}}{=} \lfloor k/d \rfloor$.

Let $\mu \stackrel{\text{def}}{=} \Pr[X_u = 1]$. As $k \geq d$, the k -wise independence gives $\mu = p^{c_1}(1 - p)^{c_0}$, where c_1, c_0 resp. denote the number of 1s and the number of 0s in the vector (b_1, \dots, b_d) . Thus, $\mu \geq (p')^d = r/N$. Next, since $k \gg d$, then $\tilde{k} = (1 - o(1)) \frac{k}{d}$ and $\lfloor \frac{\tilde{k}}{2} \rfloor = (1 - o(1)) \frac{k}{2d}$. Now assume w.l.o.g. that $\epsilon \geq \frac{\log \log(N)}{\log(N)}$ (for smaller ϵ we have larger k , so we prove for the \bar{k} which gives $\epsilon = \frac{\log \log(N)}{\log(N)}$ and then use the fact that k -wise independence guarantees \bar{k} -wise independence whenever $k > \bar{k}$). By this assumption we get $r \geq \log^6(N)$ and $k/d \leq \log^2(N)$, so $\frac{3\tilde{k}}{r} \leq r^{-(1 - o(1))\frac{2}{3}} = N^{-(1 - o(1))\frac{2}{3}(6\epsilon)} = 2^{-(1 - o(1))4\epsilon \log(N)}$. To minimize k in the current Theorem we apply Lemma 2 in [2] which strengthens standard tail-bounds for k -wise independent variables. This Lemma implies the first following \leq (the consequent \leq were justified above):

$$\Pr[X = 0] \leq \left[\frac{2(1 - \mu)\tilde{k}}{\mu(N - d)} \right]^{\lfloor \frac{\tilde{k}}{2} \rfloor} \leq \left[\frac{3\tilde{k}}{\mu N} \right]^{\lfloor \frac{\tilde{k}}{2} \rfloor} \leq \left[\frac{3\tilde{k}}{r} \right]^{\lfloor \frac{\tilde{k}}{2} \rfloor} \leq 2^{-(1 - o(1))4\epsilon \log(N) \frac{k}{2d}}.$$

On the other hand, the number of possible choices of v_i s and b_j s is only

$$\binom{N}{d} 2^d \leq 2N^d = 2^{(1 + o(1)) \log(N)d},$$

so $\Pr[X = 0]$ vanishes when $4\epsilon \frac{k}{2d} \geq (1 + \Omega(1))d$. The latter holds by equations 3 and 4 and the Theorem follows. \blacksquare

Remark 1 *Interestingly, by Theorem 3, very sparse graphs and very dense graphs look the same from the FO perspective. This is formally expressed by the fact that $\mathcal{G}(N, p)$ and $\mathcal{G}(N, 1 - p)$ have the same extension. For instance, depth $\frac{\log(N)}{10}$ properties can not distinguish between $\mathcal{G}(N, 0.001)$ and $\mathcal{G}(N, 0.999)$, and depth- $\frac{\log(N)}{2 \log \log(N)}$ properties can not tell apart $\mathcal{G}(N, \frac{1}{\log(N)})$ from $\mathcal{G}(N, 1 - \frac{1}{\log(N)})$.*

Is the $D(N)$ -extension property not only a sufficient but also a *necessary* condition for $D(N)$ -0/1-laws? While for arbitrary graph-distributions the answer is negative²,

²Consider any distribution where each graph is a union of $\Theta(\sqrt{N})$ disjoint cliques all of sizes $\Theta(\sqrt{N})$. Clearly, these graphs fail to exhibit even 2-extensions, yet, they can be shown to retain $\Theta(\sqrt{N})$ -0/1-laws.

we now show that for $\mathcal{G}(N, p)$ graphs the maximal extension and the maximal depth of 0/1-laws are roughly the same.

Theorem 4 *Let $p : \mathbb{N} \rightarrow (0, 1)$ satisfy $2^{-o(\sqrt{\log(N)})} \leq p(N) \leq 1 - 2^{-o(\sqrt{\log(N)})}$. Then there exists $q(N) = (1 \pm o(1))p(N)$, s.t. $\mathcal{G}(N, q(N))$ defies the $2D^*$ -0/1-law for some $D^* = \frac{(1-o(1))\log(N)}{\log(1/q'(N))}$ where $q'(N) = \min\{q(N), 1 - q(N)\}$.*

Remark 2 *The 0/1-law is defied here by formulas of the simplest possible form: while the strength of FO language is known to stem from alternating between existential and universal quantifiers, we use separating formulas with only existential quantification.*

Proof. Assume w.l.o.g. $p(N) \leq 1/2$. The claim will follow by presenting $q(N)$ as above and a sequence of FO formulas $\Phi = \{\phi_N\}$ with $\text{depth}(\phi_N) = \frac{(2 \pm o(1))\log(N)}{\log(1/q'(N))}$ s.t. $1/4 \leq \Pr[\mathcal{G}(N, q(N)) \models \phi_N] \leq 3/4$. To make our example more convincing we even guarantee that there is no limit to $\{\Pr[\mathcal{G}(N, q(N)) \models \phi_N]\}$.

Our formulas ϕ_N state the existence of cliques of size $\sim \frac{2\log(N)}{\log(1/p(N))}$ in the graph (for $p > \frac{1}{2}$, ϕ_N would state the existence of independent sets of that size). By Bollobás and Erdős[7], there exists $S^* = S^*(N, p(N)) = \frac{(2-o(1))\log(N)}{\log(1/p(N))}$ s.t. the maximal clique size of $\mathcal{G}(N, p)$ is almost surely either S^* or $S^* + 1$. In particular, for $\phi_N = \exists v_1 \dots v_{S^*} \bigwedge_{i \neq j} ((v_i \neq v_j) \wedge \text{EDGE}(v_i, v_j))$, we have $\Pr[\mathcal{G}(N, p) \models \phi_N] = 1 - o(1)$.

Fix sufficiently large N s.t. $\Pr[\mathcal{G}(N, p) \models \phi_N] \geq 3/4$, and let

$$\begin{aligned} \Gamma(r) &\stackrel{\text{def}}{=} \Pr[\mathcal{G}(N, r) \models \phi] = \sum_{G \models \phi} \Pr[\mathcal{G}(N, r) \models \phi] \\ &= \sum_{G \models \phi} r^{e(G)} (1-r)^{\binom{N}{2} - e(G)}. \end{aligned}$$

Here $e(G)$ denotes the number of edges in G . Let $p = p(N)$, $q = q(N)$, $q' = q'(N)$, $\mu = \mu(N)$ and $\phi = \phi_N$. Clearly, $\Gamma(\cdot)$ is continuous in r , and ϕ_N is a monotone property³. Thus, for any choice of $1/4 \leq \mu \leq 3/4$ there exists (a unique) $q \leq p$ s.t. $\Pr[\mathcal{G}(N, q) \models \phi_N] = \mu$. Thus, $\mathcal{G}(N, q(N))$ defies the $\frac{2\log(N)}{\log(1/p(N))}$ -0/1-law. In particular, we can choose μ s.t. $\{\mu(N)\}$ has no limit.

Define $\delta = \delta(N)$ by $q = p(1 - \delta)$. It suffices to prove that $\delta = o(1)$ and that $1/\log(1/p) = (1 \pm o(1))/\log(1/q')$. Intuitively, the argument for bounding δ relies on the remarkable concentration of the clique-number on either S^* or $S^* + 1$: if we assume towards contradiction that δ is constant, then $S^*(N, q)$ is smaller by a constant multiplicative factor from $S^*(N, p)$. On the other hand, the maximal clique size of a $\mathcal{G}(N, q)$ graph is almost surely $S^*(N, q) \pm 1 \lesssim S^*(N, p)$, which contradict the fact that $\Pr[\mathcal{G}(N, q) \models \phi] \geq 1/4$.

³Namely, if $G \models \phi$ and G' is obtained by adding edges to G , then $G' \models \phi$ as well.

Formally, let $\mathbb{E}_{S,N,r}$ denote the expected number of S -cliques in $\mathcal{G}(N, r)$. Let $S^* = S^*(N, p)$. By [7] $\mathbb{E}_{S^*,N,p} = N^{\Theta(1)}$ so Markov's inequality gives the second \leq in the following

$$\begin{aligned} 1/4 &\leq \mu \stackrel{\text{def}}{=} \Pr[\mathcal{G}(N, q) \models \phi] \leq \mathbb{E}_{S^*,N,q} \\ &= \binom{N}{S^*} \cdot q^{\binom{S^*}{2}} = \binom{N}{S^*} \cdot p^{\binom{S^*}{2}} (1-\delta)^{\binom{S^*}{2}} \\ &= \mathbb{E}_{S^*,N,p} (1-\delta)^{\binom{S^*}{2}} \leq N^{\Theta(1)} e^{-\Theta(\delta(S^*)^2)} \\ &= N^{\Theta(1)} e^{-\Theta\left(\delta \left(\frac{\log(N)}{\log(1/p)}\right)^2\right)}. \end{aligned}$$

This implies that $\Theta(\log(N)) \geq \delta \left(\frac{\log(N)}{\log(1/p)}\right)^2$ so $\delta \leq \Theta\left(\frac{\log(1/p)^2}{\log(N)}\right)$ and by our assumption that $\frac{1}{p} = 2^{o(\sqrt{\log(N)})}$ then $\delta \ll 1$. Finally as $0 < \delta \leq \frac{1}{2}$ we get $\frac{1}{1-\delta} = 1 + \frac{\delta}{1-\delta} \leq 1 + 2\delta \leq e^{2\delta}$. Consequently, the entire claim follows since

$$\frac{\log \frac{1}{q}}{\log(1/p)} = \frac{\log \frac{1}{p} + \log \frac{1}{1-\delta}}{\log(1/p)} \leq 1 + \frac{\log e^{2\delta}}{\log(1/p)} = 1 + \Theta\left(\frac{\delta}{\log(1/p)}\right) = 1 + o(1) \quad \blacksquare$$

Combining Theorems 1, 3, and 4 establishes the maximal depth of random graphs' 0/1-law.

Theorem 5 *Given $p : \mathbb{N} \rightarrow (0, 1)$, let $p' = \min\{p, 1-p\}$, and let $D^* = \frac{\log(N)}{\log(1/p')}$. Then there exist $D_1, D_2 = (1 \pm o(1))D^*$ s.t. :*

1. *For any density p , the D_1 -0/1-law holds for $\mathcal{G}(N, p)$.*
2. *For infinitely many densities p , the $2D_2$ -0/1-law fails to hold for $\mathcal{G}(N, p)$.*

4 Random Looking Distributions with Optimal Equivalence

In light of Theorem 5, our aim is to efficiently construct graphs that satisfy $\Theta\left(\frac{\log(N)}{\log(1/p)}\right)$ -0/1-laws and are $\Theta\left(\frac{\log(N)}{\log(1/p)}\right)$ -equivalent to $\mathcal{G}(N, p)$. This aim is met by arbitrary $\omega\left(\frac{\log^2(N)}{\log^2(1/p)}\right)$ -wise independent graphs, due to the fact that these graphs achieve optimal extensions.

Theorem 6 *Given $p : \mathbb{N} \rightarrow (0, 1)$, let $p'(N) = \min\{p(N), 1-p(N)\}$. Then for some $D(N) = (1-o(1))\frac{\log(N)}{\log(1/p')}$, and for any $k(N) = \omega\left(\frac{\log^2(N)}{\log^2(1/p')}, it holds that all $(p(N), k(N))$ -wise independent graphs satisfy the $D(N)$ -0/1-law and are $D(N)$ -equivalent to $\mathcal{G}(N, p)$ graphs.$*

Proof. Theorem 6 is precisely the lower-bound part of the proof of Theorem 3. ■

Recall that for arbitrary $p(N)$, one can construct (as in [16, 21]), $\text{poly}(\log(N))$ -wise independent graphs that are also computationally pseudorandom w.r.t. $\mathcal{G}(N, p)$. Combining this with Theorem 6 one can show the following.

Theorem 7 *Let $c > 0, p : \mathbb{N} \rightarrow (0, 1), p' = \min\{p, 1 - p\}$. Then there exists an explicit efficient construction of distributions on huge graphs \mathcal{G} that for some $D(N) = (1 - o(1)) \frac{\log(N)}{\log(1/p'(N))}$ and some $\bar{p}(N)$ s.t. $|\bar{p}(N) - p(N)| \leq N^{-3}$ are:*

1. $(\bar{p}(N), \log^c(N))$ -wise independent.
2. Satisfy the $D(N)$ -0/1-law and are $D(N)$ -equivalent to $\mathcal{G}(N, p)$.
3. Computationally indistinguishable from $\mathcal{G}(N, p)$ if one-way functions exist.

5 A Single Graph Equivalent to Random Graphs

In this section we present a single huge graph (per each order N) that is efficiently constructible and resembles $\mathcal{G}(N, p)$: the sequence achieves high equivalence to $\mathcal{G}(N, p)$ and has edge density $p \pm \epsilon$. The construction is based on Paley graphs, which are known to preserve a variety of properties of random graphs (cf. [4]). Our argument non-trivially generalizes Graham and Spencers' analysis for (directed) Paley graphs with density $p = 1/2$, into (undirected) Paley graphs with various possible densities $p = p(N)$. To stress the efficiency of our generalized construction we use indexing by n rather than by N (we write $p(n), D(n)$ instead of $p(N), D(N)$). We employ the following generalized definition:

Definition 8 (Paley graph) *Let \mathcal{F} be a finite field of size N , let $M \in \mathbb{N}$ such that $2M \mid (N - 1)$, and let $p \in \{\frac{1}{M}, \frac{2}{M}, \dots, \frac{M-1}{M}\}$. Let $Z \subset \{a \in \mathcal{F} : a^M = 1\}$ with $|Z| = pM$. Then the Paley graph $G_{\mathcal{F}, M, p, Z} = (\mathcal{F}, E_{\mathcal{F}, M, p, Z})$ is given by*

$$E_{\mathcal{F}, M, p, Z} = \{\{u, v\} : u, v \in \mathcal{F}, (u - v)^{(N-1)/M} \in Z\} \quad (5)$$

It is readily verified that every node has exactly $p(N - 1)$ neighbors, and that the graph is undirected since the exponent in (5) is even.

The rest of this section is structured as follows. First, as a technical aid we define sets of linear equalities that contain certificates to “ $x \not\equiv 0 \pmod{M}$ ”, and observe that for certain M these sets can be small. Then, we show that the $D(n)$ -FO properties of a Paley graph $G_{\mathcal{F}, M, p, Z}$ are related to the size of the smallest such certifying set for M . Next, we show that for appropriate parameters we can efficiently compute edge queries in $G_{\mathcal{F}, M, p, Z}$. Finally, we describe two concrete sequences of Paley graphs, and invoke the aforementioned lemmas to derive their efficient computability and $D(n)$ -FO properties.

Definition 9 (nonzero-certifying set) A set $C \subset \mathbb{N} \times \mathbb{Z}$ is nonzero-certifying modulo M if $\sum_{(y,z) \in C} y < M$ and for all $x \in \mathbb{Z}$:

$$x \not\equiv 0 \pmod{M} \quad \text{iff} \quad \exists (y, z) \in C : y_j x \equiv z_j \pmod{M} \quad (6)$$

For example, for any $M \in \mathbb{N}$ the set $\{(1, r)\}_{r \in \{1, \dots, M-1\}}$ is nonzero-certifying modulo M . Smaller sets can be obtained by the following:

Lemma 1 Let $M = q_1^{e_1} q_2^{e_2} \cdots q_s^{e_s}$ for distinct primes q_i and $e_i \in \mathbb{N}$. Then there exists a set C which is nonzero-certifying modulo M and $|C| = \sum_{t=1}^s e_t (q_t - 1)$.

Proof sketch. Denote $\pi_t = \prod_{t'=t+1}^s q_{t'}^{e_{t'}}$, and set $C = \{(\pi_t q_t^i, \pi_t (M/q_t) r)\}_{t,i,r}$ where $t \in \{1, \dots, s\}$, $i \in \{0, \dots, e_t - 1\}$, $r \in \{1, \dots, q_t - 1\}$. Then $|C| = \sum_{t=1}^s e_t (q_t - 1) < \sum_{t=1}^s (\lg q_t^{e_t})(B-1) = (B-1) \lg M$, and it is readily verified that $\sum_{(y,z) \in C} y = M-1$. To show that (6) indeed holds, show that it holds modulo each $q_t^{e_t}$ by considering the q_t -ary representation of $z \pmod{q_t^{e_t}}$; then apply the Chinese Remainder Theorem.⁴ ■

The next lemma shows that Paley graphs satisfy $D(n)$ -0/1-laws with $D(n)$ that is related to the size of nonzero-certifying sets. The analysis follows Graham and Spencer's proof of the connection between similar Paley graphs (restricted to $M = 2$) and tournament problems [17][5]. Recall that for a finite field \mathcal{F} , a character $\chi : \mathcal{F} \rightarrow \mathbb{C}$ of order M is a multiplicative homomorphism from \mathcal{F}^* onto the M -th roots of unity, extended with $\chi(0) = 0$; such χ exist whenever $M \mid (N - 1)$. We will invoke Weil's theorem:

Theorem 8 (Weil) Let \mathcal{F} be a finite field, let $N = |\mathcal{F}|$, and let χ be a character of order M . Let $f(x) \in \mathcal{F}[x]$ be a monic polynomial that is not an M -th root of any polynomial in $\mathcal{F}[x]$. Then:

$$\left| \sum_{u \in \mathcal{F}} \chi(f(u)) \right| < (\deg F - 1) \sqrt{N}$$

Lemma 2 Let $\mathcal{G} = \{G_{\mathcal{F}, M, p, Z}\}_n$ be a sequence of Paley graphs with $\mathcal{F} = \mathcal{F}(n)$, $M = M(n)$, $p = p(n)$, $Z = Z(n)$, $N = |\mathcal{F}(n)|$ such that $N > M^{\omega(1)}$. Let $\ell = \ell(n)$, and suppose that for every n there exist a set of size ℓ which is nonzero-certifying modulo M . Then \mathcal{G} satisfies the $D(n)$ -0/1-law for $D(n) = \frac{\lg N}{2\ell} (1 - o(1))$.

Proof. By Theorem 11, it suffices to show that $G_{\mathcal{F}, M, p, Z}$ satisfies the $D(n)$ -extension law. Denote $d = D(n)$, $\ell = \ell(n)$. Let $C = \{(y_j, z_j)\}_{j=1}^{\ell}$ be nonzero-certifying modulo M , and let $\chi : \mathcal{F} \rightarrow \mathbb{C}$ be a character of order M .

⁴Essentially, we are forming a system of linear equations which expresses a special case of the additive analogue of the Pohlig-Hellman-Silver algorithm [22].

Let $v_1, \dots, v_d \in \mathcal{F}$ be arbitrary vertices, and let $b_1, \dots, b_d \in \{0, 1\}$. We wish to show that there exists an extending vertex $u \in \mathcal{F} \setminus \{v_1, \dots, v_d\}$ such that $\{u, v_i\} \in E_{\mathcal{F}, M, p, Z}$ iff $b_i = 1$ for all $i = 1, \dots, d$. Let $w_1, \dots, w_d \in \mathcal{F}$ be chosen arbitrarily subject to $w_i^{(N-1)/M} \in Z$ iff $b_i = 1$, for $i = 1, \dots, d$. Then by definition of $E_{\mathcal{F}, M, p, Z}$, it suffices to show that there exists a vertex $u \notin \{v_1, \dots, v_d\}$ such that $(u - v_i)^{(N-1)/M} = w_i^{(N-1)/M}$ for all i . This further reduces to $\chi(u - v_i) = \chi(w_i)$, since in this case $\mu_i = (u - v_i)/w_i$ is in $\text{Ker}_\chi = \chi^{-1}(1)$ so the order of μ_i divides $|\text{Ker}_\chi| = (N-1)/M$, whence $(u - v_i)^{(N-1)/M} / w_i^{(N-1)/M} = \mu_i^{(N-1)/M} = 1$.

It thus suffices to show that there exists $u \in \mathcal{F}$ such that $\chi(u - v_i) = \chi(w_i)$ for all i . Let α be a generator of \mathcal{F}^* , and denote:

$$h(u) = \prod_{i=1}^d h_i(u) \quad \text{where} \quad h_i(u) = \prod_{j=1}^{\ell} \left(1 - \frac{\chi(u - v_i)^{y_j}}{\chi(w_i^{y_j} \alpha^{z_j})} \right) \quad (i = 1, \dots, d)$$

Note that $h_i(u) = 0$ iff there exists $j \in \{1, \dots, \ell\}$ such that $\chi(u - v_i)^{y_j} / \chi(w_i^{y_j} \alpha^{z_j}) = 1$. Since $\chi(\alpha)$ is a generator of the multiplicative group of M -th roots of unity in \mathbb{C} , which has order M , for $u \neq v_i$ we can take discrete logs to base $\chi(\alpha)$. Then:

$$h_i(u) = 0 \quad \text{iff} \quad \exists j \in \{1, \dots, \ell\} : y_j \log_{\chi(\alpha)}((u - v_i)/w_i) \equiv z_j \pmod{M}$$

Since C_n is nonzero-certifying modulo M , by considering $x = \log_{\chi(\alpha)}((u - v_i)/w_i)$ we get that $h_i(u) = 0$ iff $x \equiv 0 \pmod{M}$, i.e., iff $\chi(u - v_i) \neq \chi(w_i)$. Our task is thus reduced to showing the existence of an ‘‘extending vertex’’ $u \in \mathcal{F} \setminus \{v_1, \dots, v_d\}$ such that $h(u) \neq 0$.

Denote $S = \sum_{u \in \mathcal{F}} h(u)$. By the triangle inequality:

$$|S| \leq \sum_j h(u) \neq 0 \prod_{i=1}^d \prod_{j=1}^{\ell} \left(1 + \left| \frac{\chi(u - v_i)^{y_j}}{\chi(w_i^{y_j} \alpha^{z_j})} \right| \right) \leq \sum_{\substack{u \in \mathcal{F} \\ h(u) \neq 0}} 2^{d\ell} = d2^{d\ell} + \sum_{\substack{u \in \mathcal{F} \setminus \{v_1, \dots, v_d\} \\ h(u) \neq 0}} 2^{d\ell} \quad (7)$$

Thus, if $|S| > d2^{d\ell}$ then there exists an extending vertex. To lower bound $|S|$, we first expand the product over i and j . Denote $\mathcal{I} = \{1, \dots, d\} \times \{1, \dots, \ell\}$. Then:

$$\begin{aligned} S &= \sum_{u \in \mathcal{F}} \prod_{i=1}^d \prod_{j=1}^{\ell} \left(1 + \frac{\chi(u - v_i)^{y_j}}{-\chi(w_i^{y_j} \alpha^{z_j})} \right) = \sum_{u \in \mathcal{F}} \sum_{I \subseteq \mathcal{I}} \prod_{(i,j) \in I} \frac{\chi(u - v_i)^{y_j}}{-\chi(w_i^{y_j} \alpha^{z_j})} \\ &= \sum_{u \in \mathcal{F}} \sum_{I \subseteq \mathcal{I}} P_I \left(\prod_{(i,j) \in I} \chi(u - v_i)^{y_j} \right) \quad \text{where} \quad P_I = \prod_{(i,j) \in I} \frac{1}{-\chi(w_i^{y_j} \alpha^{z_j})} \end{aligned}$$

By separating the case $I = \emptyset$ and, changing order of summation and using the multiplicativity of χ , we then obtain:

$$S = N + \sum_{\substack{I \subseteq \mathcal{I} \\ I \neq \emptyset}} P_I \sum_{u \in \mathcal{F}} \chi(f_I(u)) \quad \text{where} \quad f_I(u) = \prod_{(i,j) \in I} (u - v_i)^{y_j}$$

For all $I \subseteq \mathcal{I}$ with $I \neq \emptyset$, $f_I(u)$ has at least one root v_i and the multiplicity of any root v_i is at most $\sum_{j=1}^{\ell} y_j < M$ by Definition 9, so $f_I(u)$ is not an M -th power of any polynomial in $\mathcal{F}[u]$. Also, $\deg f_I \leq d(M-1)$. Invoking Weil's theorem, we obtain for all such I :

$$\left| \sum_{u \in \mathcal{F}} \chi(f_I(u)) \right| \leq (d(M-1) - 1) \sqrt{N}$$

Then by the triangle inequality,

$$|S| \geq N - \sum_{\substack{I \subseteq \mathcal{I} \\ I \neq \emptyset}} P_I \left| \sum_{u \in \mathcal{F}} \chi(f_I(u)) \right| > N - 2^{d\ell} d(M-1) \sqrt{N}$$

By (7), there remains to show that $2^{d\ell} d \geq N - 2^{d\ell} d(M-1) \sqrt{N}$. Indeed:

$$\left(N - 2^{d\ell} d(M-1) \sqrt{N} \right) - 2^{d\ell} d \geq \sqrt{N} \left(\sqrt{N} - 2^{d\ell} dM \right)$$

and the latter is greater than 0 when $\lg N > 2(d\ell + \lg d + \lg M)$, i.e., when $d > \frac{\lg N - 2\lg M}{(2+o(1))\ell} > \frac{\lg N - 2\lg N/\omega(1)}{(2+o(1))\ell} = \frac{\lg N}{2\ell}(1-o(1))$. ■

Remark 3 Since the choice $w_1, \dots, w_d \in \mathcal{F}$ in the above proof was arbitrary, we have actually shown a stronger result: for the same parameters as in Lemma 2, there exists an edge labeling $L : \mathcal{F} \times \mathcal{F} \rightarrow \{1, \dots, M\}$ of the full graph of size N , such that for any d vertices v_1, \dots, v_d and labels a_1, \dots, a_d there exists a vertex $u \in \mathcal{F} \setminus \{v_1, \dots, v_d\}$ such that $L(u, v_i) = a_i$ for all $i = 1, \dots, d$. ■

Recall that $M \in \mathbb{N}$ is called B -smooth if no prime divisor of M is larger than B .

Corollary 1 Let $\mathcal{G} = \{G_{\mathcal{F}, M, p, Z}\}_n$ be a sequence of Paley graphs with $\mathcal{F} = \mathcal{F}(n)$, $M = M(n)$, $p = p(n)$, $Z = Z(n)$, $N = |\mathcal{F}(n)|$ such that $N > M^{\omega(1)}$ and M is B -smooth for $B = B(n)$. Then \mathcal{G} satisfies the $D(n)$ -0/1-law for

$$D(n) = \frac{\lg N}{2^{(B-1)\lg M}(1-o(1))}$$

Proof. Let $M = q_1^{e_1} q_2^{e_2} \cdots q_s^{e_s}$ for distinct primes $q_i \leq B$ and $e_i \in \mathbb{N}$. Then by Lemma 1, there exists a set C which is nonzero-certifying modulo M and $\ell(n) = |C| = \sum_{t=1}^s e_t (q_t - 1) < \sum_{t=1}^s (\lg q_t^{e_t}) (B - 1) = (B - 1) \lg M$. The claim follows by Lemma 2. ■

We now address the issue of efficient computability. The following lemma shows that there are sequences of Paley graphs in which edge queries can be computed efficiently, under constraints which will be addressed by the concrete sequences described later.

Lemma 3 *There exists a deterministic algorithm A which, for any \mathcal{F} , N , M and p as in Definition 8, evaluates edge queries in a Paley graph $G_{\mathcal{F}, M, p, Z}$ in the following sense: given an oracle $\mathcal{O}_{\mathcal{F}}$ which computes the basic operations in \mathcal{F} , and given an element $g \in \mathcal{F}$ of order M in \mathcal{F}^* , there exists Z as in Definition 8 such that $A^{\mathcal{O}_{\mathcal{F}}}(N, M, p, g, u, v) = 1$ iff $(u, v) \in E_{\mathcal{F}, M, p, Z}$. Moreover, if M is B -smooth then A runs in time $\text{poly}(\log N, B)$.*

Proof. Note that $\langle g \rangle = \{a \in \mathcal{F} : a^M = 1\}$, and set $Z = \{a \in \langle g \rangle : \log_g a < pM\}$. For $u \neq v$, to test whether $a = (u - v)^{(N-1)/M}$ fulfills $a \in Z$, it suffices to compute discrete logarithms in the group $\langle g \rangle$, whose order is B -smooth. This can be done deterministically in time $\text{poly}(\log N, B, |C_{\mathcal{F}}|)$ using the Pohlig-Hellman-Silver algorithm [22]. ■

We can now proceed to describe two specific efficiently computable huge graphs based on sequences of Paley graphs. As we have seen, it suffices to find a deterministically computable sequence of pairs (N, M) such that N is a prime power, $2M \mid (N - 1)$, M is highly smooth, and we can deterministically find an efficient representation of the finite field $\mathcal{F} = \text{GF}(N)$ and an element $g \in \mathcal{F}^*$ of order M . Moreover, we wish the sequence to be dense: for every $n \in \mathbb{N}$ there should be (N, M) fulfilling $M = 2^{\Theta(n)}$.

Recall the following results about finite fields, from [24] and [25].

Theorem 9 (Shoup) **(a)** *Let q be prime and $m \in \mathbb{N}$. Then there exists a deterministic algorithm that computes an irreducible polynomial $I(X)$ of degree m in $\text{GF}(q)[X]$ in time $\text{poly}(q, m)$.* **(b)** *Let $I(X)$ be any an irreducible polynomial of degree m in $\text{GF}(q)[X]$, and let $\mathcal{F} = \text{GF}(q)[X]/(I(X))$. There exists a deterministic algorithm which, given $I(X)$, runs in time $\text{poly}(q, m)$ and outputs a set of elements in \mathcal{F} which contains at least one generator of \mathcal{F}^* .*

The following is an explicit construction which approximates any desired edge density $p(n)$ up to an additive term of $\epsilon(n) < \Theta(1/n)$, and achieves $D(n)$ which is optimal up to a constant. Here, we choose N and M using Euler's theorem.

Theorem 10 *Let $p = p(n) \in (0, 1)$ and let $\epsilon = \epsilon(n) \geq c_0/n$ for a certain constant $c_0 > 0$. Then there exists a deterministically efficiently computable huge graph $\mathcal{G} =$*

$\{g_n\}_n$ which satisfies the $D(n)$ -0/1-law for $D(n) = \frac{n}{2^{\log(1/\epsilon)}(1-o(1))}$, and g_n has size $2^{\theta(n)}$ and edge density $p'(n)$ such that $|p'(n) - p(n)| < \epsilon(n)$.

Proof. Set $c_0 = 2 \lg 3$. Let $N = 3^{n'}$ where $n' = 2^k$ and $k = \lceil \lg(n/\lg 3) \rceil$. Let $M = 2^{\lceil \lg(1/\epsilon) \rceil}$. Note that $2^n < N \leq 2^{2n}$, and that $M < 2^{\lg(1/\epsilon)+1} < 2^{\lg(n/2 \lg 3)+1} = 2^{\lg(n/\lg 3)} \leq n'$, so $M \mid n'$. Since 3 is relatively prime to $2n'$, Euler's theorem yields $3^{\varphi(2n')} \equiv 1 \pmod{2n'}$, where $\varphi(2n') = n'$. Hence $2M \mid (N - 1)$. We have $\epsilon/2 < \frac{1}{M} \leq \epsilon$, and can choose $p'(n) \in \{\frac{1}{M}, \frac{2}{M}, \dots, \frac{M-1}{M}\}$ such that $|p'(n) - p(n)| \leq \frac{1}{M} \leq \epsilon$.

By Theorem 9(a), we can deterministically compute an irreducible polynomial of degree n' in $\text{GF}(3)[X]$ in time $\text{poly}(n') = \text{poly}(n)$, and can thus efficiently calculate in the field $\mathcal{F} = \text{GF}(3^{n'})$.⁵ To deterministically find an element of order M in time $\text{poly}(n)$, run the algorithm of Theorem 9(b) and, for each output element β , directly test whether $\gamma = \beta^{(N-1)/S}$ has order M by computing the first M powers of γ . Note that when β generates \mathcal{F}^* , γ indeed has order M .

By the above and Lemma 3 there exists a set Z such that $G_{\mathcal{F}, M, p, Z}$ is a Paley graph whose edge queries can be computed deterministically in time $\text{poly}(\log N) = \text{poly}(n)$. Then $\mathcal{G} = \{G_{\mathcal{F}(n), M(n), p'(n), Z(n)}\}_n$ is a deterministically efficiently computable huge graph with density $p' = p \pm \epsilon$. Since M is 2-smooth, by Corollary 1 \mathcal{G} satisfies the $D(n)$ -0/1-law for $D(n) = \frac{\lg N}{2^{\lg M}}(1-o(1)) \geq \frac{n}{2^{\lg(1/\epsilon)}(1-o(1))}$. ■

The above allows only $\epsilon(n) > \Theta(1/n)$, which means we cannot meaningfully approximate graphs with density $p \ll 1/n$. To enable better approximation ϵ , and also to obtain N closer to 2^n (albeit at some cost in the extension $D(n)$), we will replace Euler's totient function $\varphi(\cdot)$ with Carmichael's function $\lambda(\cdot)$, which likewise satisfies that $b^{\lambda(a)} \equiv 1 \pmod{a}$ for any relatively prime $a, b \in \mathbb{N}$. The benefit is that $\lambda(a)$ occasionally assumes much smaller values than $\varphi(a)$ (cf. [11]). For square-free $a \in \mathbb{N}$, $\lambda(a) = \text{lcm}\{q-1 : q \text{ prime}, q \mid a\}$. For $b \in \mathbb{N}$, let $\eta(b) = \prod_{q \text{ prime}, q-1 \mid b} q$. Note that $\lambda(\eta(b)) = b$. Then by [3]:

Theorem 11 (Pomerance, Odlyzko) *There exists a constant $c_1 > 0$ such that for all sufficiently large A , there exists $b < (\ln A)^{c_1 \ln \ln \ln A}$ s.t. $\eta(b) > A$.*

Theorem 12 *Let $p = p(n) \in (0, 1)$ and let $\epsilon > 2^{-n^{1/c_2 \ln \ln n}}$ for a constant $c_2 > 0$. Then there exists a deterministically efficiently computable huge graph $\mathcal{G} = \{g_n\}_n$ which satisfies the $D(n)$ -0/1-law for $D(n) = n/\log(1/\epsilon)^{\Theta(\log \log \log(1/\epsilon))}$, and g_n has size $2^{n(1+o(1))}$ and edge density $p'(n)$ such that $|p'(n) - p(n)| < \epsilon(n)$.*

Proof. We first find appropriate N, M . Let $B = (\ln(6/\epsilon))^{c_1 \ln \ln \ln(6/\epsilon)}$. Then by Theorem 11, for sufficiently large n there exists $b < B$ such that $\eta(b) > 6/\epsilon$. We can deterministically find such b by exhaustive search in time $\text{poly}(B) < \text{poly}(n)$. Fix any c_2 larger than c_1 . It is readily verified that $\log(6/\epsilon)^{c_1 \ln \ln n} < n/\sqrt{\ln n}$ for sufficiently

⁵Alternatively replace 5 with 3, and by [13], $X^{2^k} - 2$ is irreducible in $\text{GF}(3)[X]$.

large n , and since $n > \ln(6/\epsilon)$ we get $B < n/\sqrt{\lg n}$ and thus $b < n/\sqrt{\ln n} = o(n)$. Let n' be the smallest multiple of b that is larger than n , and let $N = 3^{n'}$. Then $2^n \leq N \leq 2^{n(1+o(1))}$.

Let $M = \prod_{\text{prime } q|\eta(b), q>\kappa} q$ where κ is the largest such that $M \geq 1/\epsilon$. Note that $M \mid \eta(b)$ and $2 \mid \eta(b)$ but $2 \nmid M$, so $2M \mid \eta(b)$, and from the definition of λ we get $\lambda(2M) \mid \lambda(\eta(b)) = b$. Thus $\lambda(2M) \mid n'$, and since $3 \nmid M$ we get $3^{2^r b} \equiv 1 \pmod{2M}$, i.e., $2M \mid (N - 1)$. Also note that all prime factors of M are at most $b + 1$, so M is $(B + 1)$ -smooth and $M < (B + 1)/\epsilon = (1/\epsilon)^{1+o(1)}$. Since $\frac{1}{M} \leq \epsilon$, we can choose $p'(n) \in \{\frac{1}{M}, \frac{2}{M}, \dots, \frac{M-1}{M}\}$ such that $|p'(n) - p(n)| \leq \frac{1}{M} \leq \epsilon$.

Conclude as in Theorem 10, with two differences. First, to test whether $\gamma = \beta^{(N-1)/M}$ is of order M , use the fact that M is $(B + 1)$ -smooth and square-free: by the Chinese Remainder, γ has order M iff $\gamma^{S/q} \neq 1$ (and thus $\gamma^{M/q}$ has order q) for every prime $q \mid M$, and this can be checked in time $\text{poly}(B \lg M) = \text{poly}(n)$. Second, M is $(B + 1)$ -smooth so we get $D(n) = \frac{n}{2^{B \lg M} (1-o(1))} = n/B^{1+o(1)}$. ■

6 The Limits of Efficiently Computable Families

We now argue that no efficiently computable family of graphs can achieve even weak $D(N)$ -equivalence to $G(N, \frac{1}{2})$ once $D(N)$ is an arbitrary polynomial in n . In fact, any computable graph-distributions that are efficient in their randomness complexity (but may still use, say, exponential running time) are distinguished from random graphs. These results can be generalized to a wide range of densities p .

Theorem 13 *Let \mathcal{G} be a computable distribution on huge graphs with seed length $0 \leq m(N) \leq \frac{1}{2} \binom{N}{2}$. Then \mathcal{G} is not even weakly $D(N)$ -equivalent to $G(N, \frac{1}{2})$ for some $D(N) = \Theta(\sqrt{m(N)} + \log(N))$.*

Remark 4 *One can insist on providing separating formulas of small $\text{poly}(m, \log(N))$ total length (instead of requiring only small quantifier depth). Then the same proof provides the desired formulas (at the expense of polynomially increasing the depth).*

Remark 5 *Efficiently computable graphs are in fact separated by the simplest form of formulas: whereas the strength of FO language is known to rely on alternating between existential and universal quantifiers, our minimal-depth separating formulas use only universal quantification. Similarly, our $\text{poly}(m, \log(N))$ -length separating formulas use only a single alternation (from universal to existential quantification).*

Remark 6 *To separate deterministic constructions from $G(N, \frac{1}{2})$, Theorems 13 and 4 provide (entirely different) separating formulas of depth $2D^* = (2 \pm o(1)) \log(N)$. The precise $2 + o(1)$ term arises in Theorem 13, because the proof actually only requires to solve $\binom{r}{2} \geq r \log(N) + \omega(1)$. For Theorem 4, the above holds simply because 0/1-laws no longer hold at depth $2D^*$, and consequently, even weak-equivalence cannot*

be achieved by a single graph. In particular, note that Paley graphs are strongly separated from $\mathcal{G}(N, \frac{1}{2})$ by depth $2D^*$ formulas, although they meet the strongest possible criteria of resemblance w.r.t. depth $\frac{1}{2}D^*$ properties.

Proof. We provide a separating condition that holds for \mathcal{G} , but rarely holds for $\mathcal{G}(N, p)$, and then translate this condition into FO formulas. Eventually the separating condition will stem from the succinct representation of efficiently computable graphs that is uncharacteristic for random graphs.

First attempt. Fix N and let $m = m(N)$. Let E be the evaluating algorithm of \mathcal{G} . The most natural separating condition is simply being in the support of \mathcal{G} , which is stated by “there exists a seed $s \in \{0, 1\}^m$ s.t. $\forall u, v$, it holds that $E(s, u, v) = 1$ iff $\text{EDGE}(u, v)$ ”. To translate this into FO language (were there are no algorithms), we use the Cook-Levin reduction from Turing machines to Boolean formulas. Thus, the term “ $E(\cdot)$ ” is replaced by some formula “ $\chi(\cdot)$ ” that satisfies $\chi(s, u, v) = E(s, u, v)$ for all inputs s, u, v of appropriate length.

Unfortunately, the terms “ $\chi(s, u, v)$ ” and “ $E(s, u, v)$ ” refer to some specific labeling of u, v as strings in $\{0, 1\}^{\log(N)}$, while the “ $\forall u, v$ ” term stands for arbitrary unlabeled vertices. The problem is not syntactic since one can encode any 0-bit as a FO expression that evaluates to FALSE, and any 1-bit as an expression that evaluates to TRUE. Instead, the problem is *semantic*: algorithms always have $\log(N)$ -bits names for vertices, and changing these names may entirely switch the output of the computation. In contrast, in FO language vertices are nameless, so all statements are closed under graph-isomorphism. This inherent difference rules out the possibility of expressing (in FO language) statements like “the j 'th bit of vertex v is 0”. Consequently, one cannot impose any semantic connection between the vertices in the term “ $E(s, u, v)$ ” and the vertices in the term “ $\forall u, v$ ”, so the first attempt fails.

Second attempt. To bypass this, one might seek a separating condition that is closed under graph-isomorphism. The most natural condition would be being isomorphic (instead of identical) to some graph in the support of \mathcal{G} . This condition could be shown to fail as well, but it takes us a long way towards our final (though perhaps less natural) separating condition: Every “small” subgraph is isomorphic to some subgraph of some graph in the support of \mathcal{G} . The latter is formally stated as follows, where $r = r(N)$ is specified later:

Condition 1 *Every subgraph on r vertices v_1, \dots, v_r is isomorphic to some subgraph of a graph that is evaluated by χ using some seed $s \in \{0, 1\}^m$.*

We first translate Condition 1 into the expression ψ_N where $u_{i_1} \dots u_{i_{\log(N)}}$ denote the bits of a vertex $u_i \in \{0, 1\}^{\log(N)}$, and $s_1 \dots s_m$ denote the bits of the seed $s \in \{0, 1\}^m$.

$$\psi_N = \forall v_1, v_2 \dots v_r \exists u_{1_1} \dots u_{1_{\log(N)}}, \dots, u_{r_1} \dots u_{r_{\log(N)}} \exists s_1 \dots s_m$$

$$\bigwedge_{i \neq j} \text{EDGE}(v_i, v_j) \Leftrightarrow \chi(s_1 \dots s_m, u_{i_1} \dots u_{i_{\log(N)}}, u_{j_1} \dots u_{j_{\log(N)}}).$$

We next translate ψ_N into FO language. As before $\chi(\cdot)$ is not in FO language where there are vertices but no bit-strings. Syntactically, this is solved by encoding each bit b_j by an expression $\text{Enc}(b_j) = \text{EDGE}(\bar{x}_j, \bar{x}'_j)$. Thus, to encode the seed and the r vertices, exactly $q = m + r \log(N)$ vertex pairs \bar{x}_j, \bar{x}'_j are needed, so ψ_N is re-formulated as:

$$\psi'_N = \forall v_1 \dots v_r \exists \bar{x}_1 \bar{x}'_1 \dots \bar{x}_q \bar{x}'_q$$

$$\bigwedge_{i \neq j} \text{EDGE}(v_i, v_j) \Leftrightarrow \chi(\text{Enc}(s), \text{Enc}(v_i), \text{Enc}(v_j)).$$

We stress that there is no way of expressing (in FO language) the validity of the encoding, namely, the requirement that indeed $\text{EDGE}(\bar{x}_j, \bar{x}'_j) = b_j$. Yet, these encodings will (almost) preserve the meaning of Condition 1, because we never require a specific encoding to be valid. Instead, we only require the *existence* of some valid encoding, and clearly, as long as the graph is neither empty nor complete, then for each bit b_j a pair of valid encoding vertices \bar{x}_j, \bar{x}'_j exists. Thus, we define our (almost) final separating formula ϕ_N by $\phi_N = \psi'_N \vee \gamma$, where γ is a fixed formula which states that the graph is either complete or empty.

These ϕ_N indeed separate \mathcal{G}_N from $G(N, \frac{1}{2})$. First, note that $\Pr[\mathcal{G}_N \models \phi_N] = 1$ because for any (single) graph G in the support of \mathcal{G}_N , if the graph is either complete or empty we are done. Otherwise, each vertex in G has a valid encoding. Since all the encodings in ψ'_N are valid, then ψ'_N has the same meaning as Condition 1 so $G \models \psi'_N$.

On the other hand, $G(N, \frac{1}{2})$ is complete or empty with only vanishing probability. Hence it suffices to show that w.h.p. $G(N, \frac{1}{2}) \not\models \psi'_N$. Indeed, assume for a fixed graph G , that $G \models \psi'_N$. This implies that for any subgraph on r vertices G_r of G the following holds: there exist strings $\bar{s} \in \{0, 1\}^m$, and $\bar{v}_i \in \{0, 1\}^{\log(N)}$, $i = 1, \dots, r$ s.t. when the evaluator E is given all $\binom{r}{2}$ inputs in lexicographic order, then $E(\bar{s}, \bar{v}_i, \bar{v}_j)$ is exactly the adjacency string of G_r . In particular this implies that G_r has small Kolmogorov complexity $\text{KC}(G_r) \leq m + r \log(N) + \Theta(1)$. Since an r -subgraph of a random graph hardly ever has $\text{KC}(G_r) < \binom{r}{2} - \omega(1)$, then ψ'_N rarely holds when $m + r \log(N) < \binom{r}{2} - \omega(1)$, which holds for, say, $r = \max\{2\sqrt{m} + 1; 4 \log(N) + 2\}$. (here the upper-bound on m is used to guarantee that $r \leq N$ so r is indeed a valid size of a sub-graph). Thus for some $r = \Theta(\sqrt{m} + \log(N))$ our ϕ_N are separating, as desired.

At this stage, by the efficiency of the Cook-Levin reduction, not only the depth but also the entire length of ϕ_N is polynomial in the running time of the evaluator E

on graphs with $\log(N)$ -bit vertices. Finally, to minimize the quantifier depth of the separating formulas (at the expense of exponential blow-up in the length), we replace the $\exists \bar{x}_1 \bar{x}'_1 \dots \bar{x}_q \bar{x}'_q$ term by a single disjunction (a single \bigvee) over all 2^q possible strings $(u_{1_1} \dots u_{1_{\log(N)}}, \dots, u_{r_1} \dots u_{r_{\log(N)}} \text{ and } s_1 \dots s_m)$ and then we ‘hard-wire’ these strings into the $\chi(\cdot)$ term. To hard-wire a specific string we encode each 0-bit by “ $v_1 \neq v_1$ ”, and encode each 1-bit by “ $v_1 = v_1$ ”. This results in a separating formula of depth $r = \Theta(\sqrt{m} + \log(N))$. The claim follows. ■

Acknowledgments. The second author wishes to thank Nati Linial, Dan Romik and Avi Wigderson for helpful discussions and Daniel Reichman for referring him to [20]. We thank Ronen Gradwohl, Eran Ofek, Guy Rothblum, Tal Sagiv and Udi Wieder for helpful comments on an earlier draft.

References

- [1] N. Alon, L. Babai and A. Itai. *A fast and simple randomized parallel algorithm for the maximal independent set problem.* journal of Algorithms 7, 567-583, 1986.
- [2] N. Alon, A. Nussboim. *On k-Wise Independent Random Graphs.* In Preparation.
- [3] L. M. Adleman, C. Pomerance and R. S. Rumely, *On Distinguishing Prime Numbers from Composite Numbers,* Annals of Mathematics, vol. 117, no. 1, 173–206, 1983.
- [4] N. Alon and J. H. Spencer, *The Probabilistic Method,* John Wiley and Sons, 1992.
- [5] L. Babai, *Character Sums, Weil’s Estimates, and Paradoxical Tournaments,* lecture notes, <http://people.cs.uchicago.edu/~laci/reu02.dir/paley.pdf>
- [6] B. Bollobás. *Random Graphs,* Academic Press, 1985.
- [7] B. Bollobás and P. Erdős, *Cliques in Random Graphs,* Cambridge Philosophical Society Mathematical Proc., vol. 80, 419–427, 1976.
- [8] B.Chor, O.Goldreich. *On the power of two-point based sampling.* J. Complexity 5(1): 96-106 (1989).
- [9] F. R. K. Chung, R. L. Graham and R. M. Wilson, *Quasi-random graphs,* Combinatorica, vol. 9, 345–362, 1989.
- [10] A. Ehrenfeucht, *An Application of Games to the Completeness Problem for Formalized Theories,* Fundamenta Mathematicae, vol. 49, 129–141, 1961.
- [11] P. Erdős, C. Pomerance and E. Schmutz, *Carmichael’s Lambda Function,* Acta Arithmetica, vol. 58, 363-385, 1991.

- [12] R. Fagin, *Probabilities in Finite Models*, Journal of Symbolic Logic, vol. 41, 50–58, 1969.
- [13] S. Gao and D. Panario, *Tests and Constructions of Irreducible Polynomials Over Finite Fields*, Foundations of Computational Mathematics (F. Cucker, M. Shub, Eds.), 346–361, Springer, 1997.
- [14] Y. V. Glebskii, D. I. Kogan, M. I. Liagonkii, V. A. Talanov, *Range and Degree of Realizability of Formulas in the Restricted Predicate Calculus*, Cybernetics, vol. 5, 142–154, 1976.
- [15] O. Goldreich, S. Goldwasser, S. Micali, *How to Construct Random Functions*, Journal of the ACM, vol. 33, no. 4, 276–288, 1985.
- [16] O. Goldreich, S. Goldwasser, A. Nussboim, *On the Implementation of Huge Random Objects*, proc. 44th IEEE Symposium on Foundations of Computer Science, 68–79, 2003.
- [17] R. L. Graham and J. H. Spencer, *A Constructive Solution to a Tournament Problem*, Canadian Math Bulletin, vol. 14, 45–48, 1971.
- [18] J. Håstad, R. Impagliazzo, L.A. Levin and M. Luby, *A Pseudo-Random Generator from any One-Way Function*, SIAM Journal on Computing, vol. 28, num. 4, 1364–1396, 1999.
- [19] A. Joffe. *On a Set of Almost Deterministic k -wise Independent Random Variables*. In Annual of Probability 2, pages 1961-1962, 1974.
- [20] M. Krivelevich and B. Sudakov, *Pseudo-random Graphs*, preprint, <http://www.math.princeton.edu/~bsudakov/papers.html>
- [21] A. Nussboim, *Huge Pseudo-Random Graphs that Preserve Global Properties of Random Graphs*, M.Sc. Thesis, Advisor: S. Goldwasser, Weizmann Institute of Science, 2003, <http://www.wisdom.weizmann.ac.il/~asafn/psdgraphs.ps>
- [22] S. C. Pohlig and M. E. Hellman, *An Improved Algorithm for Computing Logarithms Over $GF(p)$ and Its Cryptographic Significance*, IEEE Transactions on Information Theory, Vol. IT-24, 106–110, 1978.
- [23] J. H. Spencer. *The Strange Logic of Random Graphs*. Springer Verlag, 2001.
- [24] V. Shoup, *New Algorithms for Finding Irreducible Polynomials over Finite Fields*, Mathematics of Computation, vol. 54, 435–447, 1990.
- [25] V. Shoup, *Searching for primitive roots in finite fields*, Mathematics of Computation, vol. 58, 369–380, 1992.
- [26] J. H. Spencer and S. Shelah, *Zero-One Laws for Sparse Random Graphs*, Journal of the American Mathematical Society, vol. 1, 97–115, 1988.

- [27] A. Thomason, *Pseudo-random graphs*, Proceedings of Random Graphs, Annals of Discrete Mathematics 33, 307–331, 1987.