FIRST-ORDER DISTINGUISHABILITY OF SPARSE RANDOM GRAPHS

TAL HERSHKO AND MAKSIM ZHUKOVSKII

Abstract. We study the problem of distinguishing between two independent samples $\mathbf{G}_n^1, \mathbf{G}_n^2$ of a binomial random graph $G(n,p)$ by first-order (FO) sentences. Shelah and Spencer proved that, for a constant $\alpha \in (0,1)$, $G(n, n^{-\alpha})$ obeys FO zero-one law if and only if α is irrational. Therefore, for irrational $\alpha \in (0,1)$, any fixed FO sentence does not distinguish between $\mathbf{G}_n^1, \mathbf{G}_n^2$ with asymptotical probability 1 (w.h.p.) as $n \to \infty$. We show that the minimum quantifier depth $\mathbf{k}_\alpha$ of an FO sentence $\varphi = \varphi(\mathbf{G}_n^1, \mathbf{G}_n^2)$ distinguishing between $\mathbf{G}_n^1, \mathbf{G}_n^2$ depends on how closely α can be approximated by rationals:

- for all non-Liouville $\alpha \in (0,1)$, $\mathbf{k}_\alpha = \Omega(\ln \ln \ln n)$ w.h.p.;
- there are irrational $\alpha \in (0,1)$ with $\mathbf{k}_\alpha$ that grow arbitrarily slowly w.h.p.;
- $\mathbf{k}_\alpha = O_p(\frac{\ln n}{\ln \ln n})$ for all $\alpha \in (0,1)$.

The main ingredients in our proofs are a novel randomized algorithm that generates asymmetric strictly balanced graphs as well as a new method to study symmetry groups of randomly perturbed graphs.

§1. Introduction. In this article we focus on the problem of distinguishing between two independent random (simple) graphs by first-order (FO) sentences. The vocabulary of FO language of graphs contains the adjacency and the equality relations.

Given two graphs G_1, G_2 and an FO sentence φ, we say that φ *distinguishes between* G_1 and G_2 if $G_1 \models \varphi$ and $G_2 \not\models \varphi$, or vice versa. We say that G_1, G_2 are *k-distinguishable* if there exists an FO sentence φ with quantifier depth at most k that distinguishes between them. Recall that the *quantifier depth* of a sentence φ is, roughly speaking, the maximum number of nested quantifiers in φ (for a formal definition, see [15]). We call the minimum k such that G_1, G_2 are k-distinguishable *the FO distinguishability*, and denote it by $k(G_1, G_2)$. The FO distinguishability was first studied by Spencer and St. John for random sequences [23]. Tight worst-case upper bounds on $k(G_1, G_2)$ for deterministic graphs G_1, G_2 were obtained in [17] by Pikhurko, Veith, and Verbitsky. In this article we study the FO distinguishability of random graphs. Before stating our results, we discuss some motivation of this problem as well as give an overview of its history.

Distinguishability of random graphs is closely related to zero-one laws, an important phenomenon in finite model theory. We say that a sequence of random

Received February 23, 2025 UTC.

2020 *Mathematics Subject Classification.* Primary 03C13, 05C80, 68Q19, Secondary 03B10, 05C60, 60F20.

Key words and phrases. first-order logic, random graphs, zero-one laws, quantifier depth, distinguishability of graphs, graph isomorphism, asymmetry of graphs, Liouville numbers.

© The Author(s), 2026. Published by Cambridge University Press on behalf of The Association for Symbolic Logic.
0022-4812/00/0000-0000
DOI:10.1017/jsl.2026.10271

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

graphs $\{\mathbf{G}_n\}_{n=1}^{\infty}$ satisfies the *FO zero-one law* if, for every FO sentence φ, the limiting probability $\lim_{n\to\infty}\mathbb{P}(\mathbf{G}_n \models \varphi)$ is either 0 or 1. In other words, either φ holds with high probability (w.h.p., in what follows) or $\neg\varphi$ holds w.h.p. The celebrated theorem of Glebskii, Kogan, Liogon'kii, and Talanov [8] and Fagin [7] says that the FO zero-one law holds when $\mathbf{G}_n \sim G(n,1/2)$, that is, when $\mathbf{G}_n$ is distributed uniformly on the set of all labelled graphs on $[n] := \{1,\dots,n\}$.

The Bridge Theorem (see [22, Theorems 2.5.1 and 2.3.1]) implies the following relation between zero-one laws and distinguishability: a sequence of random graphs $\{\mathbf{G}_n\}_{n=1}^{\infty}$ satisfies the FO zero-one law if and only if the respective FO distinguishability is unbounded, that is, for every $k \in \mathbb{N}$,

$$\lim_{n,m\to\infty}\mathbb{P}(k(\mathbf{G}_n,\mathbf{G}_m)\geq k) = 1.$$

In [13] Kim, Pikhurko, Spencer, and Verbitsky proved a more precise estimation: w.h.p. $k\left(\mathbf{G}_n^1,\mathbf{G}_n^2\right) = \log_2 n + O(\log\log n)$. In [4] Benjamini and the second author of the article obtained a tight bound: w.h.p. $k\left(\mathbf{G}_n^1,\mathbf{G}_n^2\right) = \log_2 n - 2\log_2\ln n + O(1)$ and is concentrated in three consecutive points. They also observed that the same result holds true for the number of variables in infinitary logic $\mathcal{L}^{\omega}_{\infty,\omega}$.

FO distniguishability is also related to the graph isomorphism problem. Note that two finite graphs are isomorphic if and only if there is no FO sentence that distinguishes between them. Since the truth value of an FO sentence of quantifier depth k on an n-vertex graph can be tested in time $O(n^k)$ [15], upper bounds on the FO distinguishability of graphs imply upper bounds on the time complexity of deciding whether or not the graphs are isomorphic. From this perspective, the problem of distinguishing between two independent copies $\mathbf{G}_n^1,\mathbf{G}_n^2 \sim G(n,1/2)$ arises naturally from the average case analysis of the graph isomorphism problem. However, this approach gives a bound that is far from being optimal: it only shows that $\mathbf{G}_n^1,\mathbf{G}_n^2$ can be distinguished in quasi-polynomial time w.h.p. A much more efficient algorithm can be obtained by considering FO logic with counting. Indeed, the results of Babai and Kučera [2] and Babai, Erdős, and Selkow [1] (combined with Immerman and Lander's [10] logical characterization of color refinement) imply that w.h.p. $\mathbf{G}_n \sim G(n,1/2)$ can be *defined* by an FO sentence with counting quantifiers of quantifier depth 4. This implies a polynomial (actually, even linear) time algorithm that distinguishes between $\mathbf{G}_n$ and *any* non-isomorphic graph w.h.p.

In sparse random graphs, however, the situation is strikingly different. The FO distinguishability can be significantly smaller, and therefore provide much more efficient algorithms for random graphs distinguishing. To state the relevant results, recall the general definition of the binomial random graph: $\mathbf{G}_n \sim G(n,p)$ is a random graph on $[n]$, in which edges between pairs of vertices appear independently with probability p. In [21], Shelah and Spencer studied the validity of the FO zero-one law for $\mathbf{G}_n \sim G(n,n^{-\alpha})$ where α is a positive constant.

THEOREM 1.1 (S. Shelah and J. Spencer [21]). *Let $\alpha \in (0,1)$. Then $\mathbf{G}_n \sim G(n,n^{-\alpha})$ obeys an FO zero-one law if and only if α is irrational.*

It follows that, for rational $\alpha \in (0,1)$ and two independent copies $\mathbf{G}_n^1,\mathbf{G}_n^2 \sim G(n,n^{-\alpha})$ we have $k(\mathbf{G}_n^1,\mathbf{G}_n^2) = O(1)$ with probability bounded away from 0.

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

Furthermore, it can be even derived that, for every $\varepsilon > 0$, $\mathbb{P}(k(\mathbf{G}_n^1, \mathbf{G}_n^2) < C) > 1 - \varepsilon$ for a certain constant $C = C(\alpha, \varepsilon)$.

For irrational α, the asymptotic behavior of $k(\mathbf{G}_n^1, \mathbf{G}_n^2)$ is more complicated. Authors of [4] suspected that similar methods to those that were applied for dense random graphs might imply $k(\mathbf{G}_n^1, \mathbf{G}_n^2) = O(\ln n)$ w.h.p. for irrational α as well and it was, however, conjectured that actually $k(\mathbf{G}_n^1, \mathbf{G}_n^2) = o(\ln n)$ w.h.p. Our first result shows that this conjecture is true.

For an irrational $\alpha \in (0, 1)$, we let $\mathbf{k}_\alpha = k(\mathbf{G}_n^1, \mathbf{G}_n^2)$ for two independent $\mathbf{G}_n^1, \mathbf{G}_n^2 \sim G(n, n^{-\alpha})$. For a sequence of random variables ξ_n and a sequence of non-zero constants a_n, we write $\xi_n = O_p(a_n)$ when ξ_n/a_n is *stochastically bounded*, that is, for every $\varepsilon > 0$, there exists $C > 0$ such that $\mathbb{P}(|\xi_n/a_n| < C) > 1 - \varepsilon$ for all n.

THEOREM 1.2. *Let $\alpha \in (0, 1)$ be irrational. Then $\mathbf{k}_\alpha = O_p\left(\frac{\ln n}{\ln \ln n}\right)$.*

On the other hand, our second result shows that it is impossible to get a uniform lower bound approaching infinity: $\mathbf{k}_\alpha$ may grow arbitrarily slow.

THEOREM 1.3. *For every function $f(n) \overset{n\to\infty}{\to} \infty$ there exists an irrational $\alpha \in (0, 1)$ and an increasing sequence of positive integers $\{n_t\}_{t=1}^{\infty}$ such that w.h.p. (as $t \to \infty$) $\mathbf{k}_\alpha(n_t) \le f(n_t)$.*

We observe that FO distinguishability $\mathbf{k}_\alpha$ depends on how well an irrational α can be approximated by rational numbers: the better α is approximable, the closer the behavior of $\mathbf{k}_\alpha$ to the case of rational α is. This allows us to prove our third result: in contrast to Theorem 1.3, for almost all $\alpha \in (0, 1)$, $\mathbf{k}_\alpha$ is at least $\frac{1-o(1)}{\ln 2} \ln \ln \ln n$ w.h.p. Let us recall that the *Liouville–Roth irrationality measure* of an irrational number α is the infimum (which can be infinite) of the set of all $d > 0$ such that at most finitely many integers p and q satisfy $|\alpha - p/q| \le q^{-d}$. Due to Roth's Theorem [18], $d = 2$ if α is algebraic (while for transcendental numbers $d \ge 2$ or $d = \infty$—in the latter case α is called *Liouville number*).

THEOREM 1.4. *Let $d_0 > 2$ and $\alpha \in (0, 1)$ be an irrational number with the Liouville–Roth irrationality measure strictly smaller than d_0. Then w.h.p. $\mathbf{k}_\alpha \ge \frac{1}{\ln d_0} \ln \ln \ln n$. In particular, for almost all irrational $\alpha \in (0, 1)$ (in the Lebesgue measure), w.h.p. $\mathbf{k}_\alpha \ge \frac{1-o(1)}{\ln 2} \ln \ln \ln n$.*

The second assertion follows from the fact that the set of α with Liouville–Roth irrationality measure strictly bigger than 2 has Lebesgue measure 0 due to Khinchin's Theorem [12]. In particular, the lower bound $\frac{1-o(1)}{\ln 2} \ln \ln \ln n$ holds true for all algebraic irrational $\alpha \in (0, 1)$ due to Roth's Theorem. In contrast, in our proof of Theorem 1.3, for a given growing $f(n)$, the corresponding irrational α is defined as a suitable Liouville number.

To prove Theorem 1.2 we distinguish $\mathbf{G}_n^1$ from $\mathbf{G}_n^2$ via an existential sentence. The strategy is to find a large family of graphs $\mathcal{F}$ with the property that w.h.p. there exists $H \in \mathcal{F}$ which appears as an induced subgraph in $\mathbf{G}_n^1$ but not in $\mathbf{G}_n^2$. Commonly, in order to verify that a representative of a given isomorphism class H appears as an induced subgraph of a random graph, the second moment of the random variable $\mathbf{X}_H$ that counts the number of such appearances is computed. When H has subgraphs that are at least as dense as H itself (i.e., H is not *strictly balanced*), the variance

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

of this random variable becomes large and does not allow to apply a concentration inequality. Moreover, symmetries of H negatively affect the expectation of $\mathbb{E}\mathbf{X}_H$. Therefore, we require $\mathcal{F}$ to be a large family of asymmetric strictly balanced graphs with a given density.

Our last contribution is the existence of a large family of asymmetric strictly balanced graphs as well as a randomized algorithm for sampling such graphs, that we have used to prove Theorem 1.2. We prove that for every density $\rho \geq 1 + 2/n$ (this restriction is tight), bounded from above by a constant, there exists a family of $n^{(\rho - o(1))n}$ strictly balanced asymmetric graphs on $[n]$ (see Theorem 2.1 in Section 2). The novel algorithm of sampling *random balanced graphs* that are w.h.p. asymmetric that we present in Section 2.1 is inspired by the proof of Ruciński and Vince [20] of the existence of a strictly balanced graph for every fixed density $\rho \geq 1$. In order to prove that the random balanced graph is asymmetric w.h.p., we develop a new approach for proving asymmetry of randomly perturbed graphs, based on the concept of *alternating cycles*, that is, cycles whose edges alternate between the edges of a fixed deterministic graph and the random edges. We believe that these results are interesting in their own right—strictly balanced graphs and their generalizations (as well as their automorphisms) naturally arise in different contexts in random graphs theory (see, e.g., [9, 19, 20] and [11, Chapter 3]).

The article is organized as follows. In Section 2 we define the random balanced graph, study its properties, and then show that it can be used to prove the existence of a large family of strictly balanced asymmetric graphs. The crucial property of the random balanced graph—asymmetry—is proved in Section 6. The existence of a large family of strictly balanced asymmetric graphs is used in Section 3 to prove Theorem 1.2. Sections 4 and 5 are devoted to proofs of Theorems 1.3 and 1.4, respectively. Note that the proof of Theorem 1.4 requires the random balanced graph and its asymmetry as well.

Notations and conventions:

Throughout the text, we often maintain a convention of denoting random variables in boldface letters.

For a graph G, we denote by $v(G)$ and $e(G)$ the number of vertices and edges in G, respectively. In addition, $\delta(G)$ is the minimum degree of G and $\Delta(G)$ is the maximum degree of G. For a set of vertices U, let $G[U]$ denote the subgraph of G induced by U.

We will sometimes need to work with multigraphs. We often use asterisks to distinguish multigraphs from simple graphs, unless the distinction is clear from the context. Recall that a *multigraph* is a pair $G^* = (V, E^*)$ where V is a set of vertices and E^* is a multiset of edges (loops are allowed). $e(G^*)$ is the number of edges counted with multiplicities. Given two multisets E_1^*, E_2^*, we define their sum $E_1^* + E_2^*$ as the multiset E^* with the property $m_{E^*}(e) = m_{E_1^*}(e) + m_{E_2^*}(e)$ for every e, where $m_{E^*}(e)$ denotes the multiplicity of e in E^*. For two multigraphs on the same vertex set $G_1^* = (V, E_1^*)$, $G_2^* = (V, E_2^*)$ we define their sum as $G_1^* + G_2^* = (V, E_1^* + E_2^*)$. Note that $e(G_1^* + G_2^*) = e(G_1^*) + e(G_2^*)$. When we talk about (multi)graphs on n vertices, we always refer to labelled graphs with vertex set $[n]$.

We use the asymptotic notations $f(n) \sim g(n)$ for $f(n) = (1 + o(1))g(n)$ and $f(n) \lesssim g(n)$ for $f(n) \leq (1 + o(1))g(n)$.

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

Note that the symbol $\sim$ is also used to denote sampling from a distribution, for example, $\mathbf{G} \sim G(n,p)$. The two meanings of the symbol are distinguished by the context.

§2. The random balanced graph. In this section we introduce a random graph model, denoted $\mathbf{H}(n,m)$, which we call the *random balanced graph* (with n vertices and m edges). We then use it to generate a large family of asymmetric, strictly balanced graphs on $[n]$ with a given density. This will be the key ingredient in our proof of Theorems 1.2 and 1.3

Recall that a non-empty graph H is *balanced* if $d(H) \geq d(H_0)$ for every non-empty subgraph $H_0 \subseteq H$, where $d(H) = \frac{e(H)}{v(H)}$ is the *density* of H. Furthermore, H is *strictly balanced* if $d(H) > d(H_0)$ for every proper subgraph $H_0 \subset H$. Also recall that a graph H is *asymmetric* if its group of automorphisms is trivial. The same definitions apply for multigraphs.

In Sections 3 and 4 we will be interested in the case where the density $\frac{m}{n}$ approaches $\frac{1}{\alpha}$ as $n \to \infty$. Here we consider a more general setting of a bounded density. The main goal of this section is to prove the following theorem.

THEOREM 2.1. *Fix a constant $c > 1$ and let $m = m(n)$ satisfy $n + 2 \leq m \leq cn$. Then, for every even positive integer* n, *there exists a family $\mathcal{H}(n,m)$ of strictly balanced and asymmetric graphs on $[n]$ with* m *edges such that $|\mathcal{H}(n,m)| = \exp(-O(n))n^m$.*

REMARK 2.2. It makes sense to consider only connected graphs since a disconnected graph is not strictly balanced, and so m should be at least $n-1$. On the other hand, any tree is strictly balanced. Up to isomorphism, there is only one connected graph with $m = n$ edges which is strictly balanced—a cycle C_n, though it is not asymmetric. Finally, every strictly balanced graph with $m = n+1$ edges is either a cycle with a path between its two vertices, or two cycles joined by a path. Clearly, both graphs are not asymmetric. Thus, the condition $m \geq n+2$ is necessary unless the graph is a tree.

REMARK 2.3. Although we believe that the assertion of Theorem 2.1 is true for odd n as well, we state it in this weaker form since an adaptation of our approach to odd n requires a more complicated case analysis to prove strict balancedness. As for our purposes it is enough to consider even n, we omit these technical complications.

In Section 2.1 we define $\mathbf{H}(n,m)$, in Section 2.2 we study its properties, and in Section 2.3 we complete the proof of Theorem 2.1.

2.1. Definition. From now on fix $m = m(n)$ and assume that n is even and $n + 2 \leq m \leq cn$, where $c > 1$ is a constant. We begin with a definition of the random balanced *multigraph* $\mathbf{H}^*(n,m)$. It is inspired by Ruciński and Vince's construction of a strictly balanced graph [20]. Write $m = q \cdot \frac{n}{2} + r$ where $q \geq 2$ and $0 \leq r < \frac{n}{2}$ are integers. $\mathbf{H}^*(n,m)$ is defined as the sum of three different (multi)graphs, which we call components: the *regular component* $\mathbf{H}_r$ with $(q-2)\frac{n}{2}$ edges, the *Hamiltonian component* $\mathbf{H}_h$ with n edges, and the *balancing component* $\mathbf{H}_b$ with r edges.

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

Let C be a cycle of order n; write it as $(v_1, v_2, \dots, v_n, v_1)$. For a given $0 \le r \le n$, define the following set of r *almost equidistributed* vertices around C:

$$R = R(C, r) = \left\{ v_i \;\middle|\; 1 \le i \le n,\ \left\lfloor (i-1) \cdot \frac{r}{n} \right\rfloor < \left\lfloor i \cdot \frac{r}{n} \right\rfloor \right\}. \tag{1}$$

Note that indeed $|R| = r$. In addition, for any sequence of k consecutive vertices in C, less than $k \cdot \frac{r}{n} + 1$ of them belong to R. Indeed, a sequence of vertices $v_{i+1}, \dots, v_{i+k}$ (with indices taken modulo n) has $\left\lfloor (i+k) \cdot \frac{r}{n} \right\rfloor - \left\lfloor i \cdot \frac{r}{n} \right\rfloor$ vertices from R, and this is less than $k \cdot \frac{r}{n} + 1$. This property shows that the vertices of R are distributed almost equally, which is useful for proving balancedness.

Definition 2.4. The random balanced multigraph with n vertices and m edges $\mathbf{H}^* = \mathbf{H}^*(n, m)$ is defined as follows:

- Let $\mathbf{H}_r = ([n], \mathbf{E}_r)$ be a uniformly distributed $(q-2)$-regular (simple) graph on $[n]$.[1]
- Let $\mathbf{H}_h = ([n], \mathbf{E}_h)$ be a uniformly random Hamilton cycle, independent of $\mathbf{H}_r$.
- Let $\mathbf{H}_b = ([n], \mathbf{E}_b)$ be defined as follows. For every $x \in \mathbf{R} := R(\mathbf{H}_h, r)$ choose a random vertex $\mathbf{y}[x] \in [n]$, such that these r vertices are chosen independently of $\mathbf{H}_r$ and $\mathbf{H}_h$, and are uniformly distributed over all possible sequences of r *distinct* vertices. Now define $\mathbf{E}_b = \{\{x, \mathbf{y}[x]\} : x \in \mathbf{R}\}$. We regard $\mathbf{E}_b$ as a multiset: edges may have multiplicity 2 and loops are allowed. Thus $\mathbf{H}_b$ may be a multigraph.

Finally, define $\mathbf{H}^* = \mathbf{H}_r + \mathbf{H}_h + \mathbf{H}_b$. Edges of $\mathbf{H}_r$ are called *regular*, edges of $\mathbf{H}_h$ are called *Hamiltonian*, and edges of $\mathbf{H}_b$ are called *balancing*.

Remark 2.5. The requirement that the vertices $\{\mathbf{y}[x] : x \in \mathbf{R}\}$ are all distinct is introduced in order to restrict the maximum degree of $\mathbf{H}^*$. Indeed, note that $\mathbf{H}_r \cup \mathbf{H}_h$ is a q-regular multigraph, and that degrees in $\mathbf{H}_b$ are 0, 1 or 2. Therefore,

$$q \le \delta(\mathbf{H}^*) \le \Delta(\mathbf{H}^*) \le q + 2. \tag{2}$$

Remark 2.6. Let $N^* = N^*(n, m)$ denote the number of possible values (H_r, H_h, H_b) of $(\mathbf{H}_r, \mathbf{H}_h, \mathbf{H}_b)$. By definition, $N^* = N_r N_h N_b$, where $N_r = \exp(-O(n))n^{\binom{q}{2}-1)n}$ is the number of $(q-2)$-regular graphs on n vertices (see [11, Corollary 9.8]); $N_h = (n-1)!/2$ is the number of cycles on $[n]$; $N_b = (n)_r$ is the number of sequences of r distinct vertices. Note that $(\mathbf{H}_r, \mathbf{H}_h, \mathbf{H}_b)$ is uniformly distributed over all N^* possible values. However, $\mathbf{H}^*$ itself is *not* uniformly distributed. This is because different multigraphs H^* may be represented as a sum $H_r + H_h + H_b$ in a different number of ways.

Definition 2.7. The *random balanced graph* $\mathbf{H} = \mathbf{H}(n, m)$ is defined as a random (simple) graph whose distribution is the distribution of $\mathbf{H}^*(n, m)$ conditioned on being simple.

Remark 2.8. Let $N = N(n, m)$ be the number of possible values (H_r, H_h, H_b) of $(\mathbf{H}_r, \mathbf{H}_h, \mathbf{H}_b)$ such that $H_r + H_h + H_b$ is a simple graph. Conditioning on the event that $\mathbf{H}^*$ is simple, $(\mathbf{H}_r, \mathbf{H}_h, \mathbf{H}_b)$ is uniformly distributed over all N possible values.

[1] We refer to [11, Chapter 9] as an introduction to random regular graphs.

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

2.2. Properties. In this section we list several important properties of $\mathbf{H}$. The main three properties are strict balancedness (Proposition 2.9), a bound on the probability that $\mathbf{H}$ contains a given set of edges (Proposition 2.13), and asymmetry (Theorem 2.15).

PROPOSITION 2.9. *Suppose* H *is a graph such that* $\mathbb{P}(\mathbf{H} = H) > 0$. *Then* H *is strictly balanced.*

Although the proof of Proposition 2.9 resembles the proof of Ruciński and Vince that the graphs that they construct in [20] are strictly balanced, below we present the full proof for completeness.

PROOF. We actually prove a stronger statement: if $\mathbb{P}(\mathbf{H}^* = H^*) > 0$ then H^* is strictly balanced.

Consider a possible value of (H_r, H_h, H_b) of $(\mathbf{H}_r, \mathbf{H}_h, \mathbf{H}_b)$ and let $H^* = H_r + H_h + H_b$. We shall prove that H^* is strictly balanced. That is, we show that every proper sub-multigraph $H_0 \subset H^*$ satisfies

$$\frac{e(H_0)}{v(H_0)} < \frac{m}{n} = \frac{q}{2} + \frac{r}{n}.$$

We now follow a sequence of reductions.

First, we claim that it suffices to prove strict balancedness of $H' = H_h + H_b$. Indeed, suppose that H' is strictly balanced. Consider a set $V_0 \subset [n]$ and denote $v_0 = |V_0|$, $H_0 = H[V_0]$, and $H_0' = H'[V_0]$. Since H' is strictly balanced,

$$\frac{e\left(H_0'\right)}{v_0} < \frac{e(H')}{n} = \frac{m - (q-2)n/2}{n} = 1 + \frac{r}{n}.$$

Now, H_0 is obtained from H_0' by adding some edges of the $(q-2)$-regular graph H_r, which increases the degree of each vertex by at most $q-2$. Therefore at most $\left(\frac{q}{2} - 1\right) v_0$ edges are added. Overall,

$$\frac{e(H_0)}{v_0} \le \frac{\left(\frac{q}{2} - 1\right) v_0 + e\left(H_0'\right)}{v_0} < \frac{q}{2} + \frac{r}{n}$$

which proves strict balancedness of H^*.

Second, we claim that it suffices to prove strict balancedness of H'', which is obtained from H' by replacing its balancing edges with loops: each edge $\{x, y[x]\}$ is replaced with a loop at x. Indeed, suppose that H'' is strictly balanced. Consider a set $V_0 \subset [n]$ and denote $v_0 = |V_0|$, $H_0' = H'[V_0]$, and $H_0'' = H''[V_0]$. It is easy to see that $e\left(H_0'\right) \le e\left(H_0''\right)$. From strict balancedness of H'' and that $e(H') = e(H'')$,

$$\frac{e\left(H_0'\right)}{v_0} \le \frac{e\left(H_0''\right)}{v_0} < 1 + \frac{r}{n},$$

which proves strict balancedness of H'.

Third, we claim that to prove strict balancedness of H'', it suffices to check only V_0 which are *segments* of the Hamiltonian component H_h; that is, sequences of consecutive vertices on the cycle. Indeed, assume that we have verified strict balancedness for segments. Consider a set $V_0 \subset [n]$ and denote $v_0 = |V_0|$ and $e_0 = e\left(H_0''\right)$. Notice that V_0 can be written as the union of mutually disjoint segments

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

$\tau_1 \sqcup \tau_2 \sqcup \cdots \sqcup \tau_t$ with no Hamiltonian edge between them. Let v_i, e_i denote the number of vertices and edges in $H''[\tau_i]$. From the assumption, $e_i < \left(1 + \frac{r}{n}\right) v_i$ for every i. Then

$$e_0 = \sum_{i=1}^{t} e_i < \left(1 + \frac{r}{n}\right) \sum_{i=1}^{t} v_i = \left(1 + \frac{r}{n}\right) v_0,$$

which proves strict balancedness.

Finally, it remains to check strict balancedness of H'' for segments. Take any segment τ with $v_0 < n$ vertices. The subgraph $H''[\tau]$ contains $v_0 - 1$ Hamiltonian edges and less than $v_0 \cdot \frac{r}{n} + 1$ loops (since the loops are almost equidistributed). Therefore,

$$e_0 < v_0 - 1 + v_0 \cdot \frac{r}{n} + 1 = \left(1 + \frac{r}{n}\right) v_0.$$

That finishes the proof. ⊣

PROPOSITION 2.10. *There exists a positive constant p_0 such that $\mathbb{P}(\mathbf{H}^*$ is simple$) \geq p_0$.*

REMARK 2.11. Following our notation from Remark 2.6 and Definition 2.7, let us note that $\mathbb{P}(\mathbf{H}^*$ is simple$) = N/N^*$, so we can rewrite the inequality as $N \geq p_0 N^*$.

PROOF OF PROPOSITION 2.10. Let A be the event that $\mathbf{H}^*$ is simple. Then $A = A_1 \cap A_2$, where

1. A_1 is the event that $\mathbf{H}_r$ and $\mathbf{H}_h$ do not share common edges.
2. A_2 is the event that $\mathbf{H}_b$ does not contain loops or edges with multiplicity 2 and does not share edges with $\mathbf{H}_r$ or with $\mathbf{H}_h$.

We write $\mathbb{P}(A) = \mathbb{P}(A_1)\mathbb{P}\left(A_2 \mid A_1\right)$ and prove that there are positive constants p_1, p_2 such that $\mathbb{P}(A_1) \geq p_1, \mathbb{P}\left(A_2 \middle| A_1\right) \geq p_2$.

$\mathbb{P}(A_1)$ is equal to the probability that the random $(q-2)$-regular graph $\mathbf{H}_r$ does not contain any edges of a given Hamilton cycle. When $q = 2$ it is clearly 1, so assume $q \geq 3$. We apply [16, Theorem 1.1], which bounds the probability that a random graph with specified degrees does not contain any edges of a given subgraph. For the special case of regular graphs it yields the following statement.

THEOREM 2.12 (B. McKay [16]). *Let $1 \leq d = d(n) \ll n$ be an integer and let $X = X(n)$ be a graph on vertex set $[n]$ and maximum degree $\Delta := \Delta(X) = o(n)$. For every even* n *let $G_{d,X}(n)$ denote the number of labelled* d*-regular graphs on $[n]$ that do not contain any edge of* X*. Then*

$$G_{d,X}(n) = \frac{(2E)!}{E!\, 2^E (d!)^n} \exp\left(-\frac{nd(d-1)}{4E} - \left(\frac{nd(d-1)}{4E}\right)^2\right) \times \exp\left(-\frac{d^2 e(X)}{2E} + O\left(\frac{d^2(d+\Delta)^2}{E}\right)\right).$$

Here $E = \frac{nd}{2}$ is the number of edges in a d*-regular graph on* n *vertices.*

In our case, let $d = q - 2$ (which is a constant) and let X be a Hamiltonian cycle, so $e(X) = n$ and $\Delta = 2$. Also let $\emptyset$ denote the empty graph (on n vertices), so $G_{d,\emptyset}(n)$

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

is the number of labelled d-regular graphs. Then

$$\mathbb{P}(A_1) = \frac{G_{d,X}(n)}{G_{d,\emptyset}(n)} = \exp\left(-\frac{nd^2}{nd} + O\left(\frac{1}{nd}\right)\right) \sim e^{-d} = \mathrm{e}^{-(q-2)}.$$

Hence $\mathbb{P}(A_1)$ is bounded from below by a positive constant.

Now let us bound $\mathbb{P}\left(A_2 \mid A_1\right)$. We fix values $\mathbf{H}_r = H_r$ and $\mathbf{H}_h = H_h$ such that A_1 holds, and prove the existence of a constant p_2 (independent of H_r, H_h) such that

$$\mathbb{P}(A_2 \mid \mathbf{H}_r = H_r,\ \mathbf{H}_h = H_h) \geq p_2.$$

Let $R = \{x_1, x_2, \ldots, x_r\}$ (the labeling may be chosen arbitrarily) be the set of r equidistributed vertices around H_h. Since A_1 holds, $H_r + H_h$ is a simple q-regular graph. The sequence of values $\mathbf{y}[x_1], \mathbf{y}[x_2] \ldots, \mathbf{y}[x_r]$ is drawn uniformly from all the $(n)_r$ sequences of r distinct vertices. We provide a lower bound on the number of choices which satisfy A_2.

Let us sequentially choose the values $y[x_1], y[x_2] \ldots, y[x_r]$, making sure that A_2 is satisfied at every step. When we get to $y[x_j]$, there are at least $n - q - j$ possible choices which assure that A_2 is still satisfied. We deduce

$$\begin{aligned}\mathbb{P}(A_2 \mid \mathbf{H}_r = H_r,\ \mathbf{H}_h = H_h) &\geq \frac{(n-q-1)_r}{(n)_r} \overset{(*)}{\geq} \frac{(n-2c-1)_r}{(n)_r} \\ &\overset{(**)}{\geq} \frac{(n-2c-1)_{\frac{n}{2}}}{(n)_{\frac{n}{2}}} \overset{(***)}{\geq} \left(\frac{\frac{n}{2}-2c-1}{\frac{n}{2}}\right)^{\frac{n}{2}} \sim \mathrm{e}^{-(2c+1)}.\end{aligned}$$

Here $(*)$ follows from $q \leq \frac{2m}{n} \leq 2c$, $(**)$ follows from $r < \frac{n}{2}$, and $(***)$ holds since $\frac{n-2c-i-1}{n-i}$ decreases as a function of i.

PROPOSITION 2.13. *There exists a positive constant c_0 such that, for every set $E_0 \subseteq \binom{[n]}{2}$, we have $\mathbb{P}(E_0 \subseteq E(\mathbf{H})) \leq (c_0/n)^{|E_0|}$.*

PROOF. The idea is to partition E_0 into three disjoint sets and separately bound the probability that they are subsets of $\mathbf{E}_r, \mathbf{E}_h, \mathbf{E}_b$. It will be simpler to work with $\mathbf{H}^*$, and for that we will rely on Proposition 2.10.

Formally, fix $E_0 \subseteq \binom{[n]}{2}$ and let $\mathcal{P}$ be the set of triplets (E_r, E_h, E_b) which form a partition of E_0. Note that $|\mathcal{P}| = 3^{|E_0|}$. For every $P = (E_r, E_h, E_b) \in \mathcal{P}$ let A_P denote the event that in $\mathbf{H}^*$ we have $E_r \subseteq \mathbf{E}_r$, $E_h \subseteq \mathbf{E}_h$, and $E_b \subseteq \mathbf{E}_b$. Then, from Proposition 2.10,

$$\mathbb{P}(E_0 \subseteq E(\mathbf{H})) \leq \frac{1}{p_0}\mathbb{P}(E_0 \subseteq E(\mathbf{H}^*)) \leq \frac{1}{p_0}\sum_{P\in\mathcal{P}} \mathbb{P}(A_P).$$

We will now prove that there exists a constant c_1 such that $\mathbb{P}(A_P) \leq (c_1/n)^{|E_0|}$ for every $P \in \mathcal{P}$. Then, taking $c_0 = \frac{1}{p_0} \cdot 3c_1$ finishes the proof.

So, let us fix a partition $P = (E_r, E_h, E_b)$ of E_0. Let $\ell_0 = |E_0|$ and similarly define ℓ_r, ℓ_h, ℓ_b.

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

Step 1. We apply the following proposition from [5].

PROPOSITION 2.14 [5, Equation (2)]. *Fix a constant $d \geq 1$ and let $\mathbf{G}_{n,d}$ be a random* d*-regular graph on* $[n]$ *vertices. Then there exists a constant $c > 0$ (depending on* d *but not on* n*) such that, for every set $E \subseteq \binom{[n]}{2}$, $\mathbb{P}\left(E \subseteq E(\mathbf{G}_{n,d})\right) \leq \left(\frac{c}{n}\right)^{|E|}$.*

In our case, it follows that there exists a positive constant c_r such that $\mathbb{P}(E_r \subseteq \mathbf{E}_r) \leq \left(\frac{c_r}{n}\right)^{\ell_r}$.

Step 2. Let us bound $\mathbb{P}(E_h \subseteq \mathbf{E}_h)$. Recall that $\mathbf{E}_h$ is the set of edges of a random Hamilton cycle. If E_h is not contained in any Hamilton cycle, this probability is trivially 0, so assume that it is not the case.

If $\ell_h = n$, then E_h is already the set of edges of a Hamilton cycle, and then $\mathbb{P}(E_h \subseteq \mathbf{E}_h) = \frac{2}{(n-1)!}$.

If $\ell_h < n$, then the graph $([n], E_h)$ has exactly $n - \ell_h$ connected components, all of them are either paths or isolated vertices (which are not considered as paths for now). Let s be the number of paths. Then the number of Hamilton cycles containing E_h equals $2^{s-1}(n - \ell_h - 1)!$. Therefore,

$$\begin{aligned}\mathbb{P}(E_h \subseteq \mathbf{E}_h) &= 2^s \cdot \frac{(n-\ell_h-1)!\cdot n}{n!} \leq \frac{2^{\ell_h}}{(n-1)_{\ell_h}} \\ &\leq \left(\frac{2\mathrm{e}}{n-1}\right)^{\ell_h} = \left(\frac{n}{n-1}\right)^{\ell_h}\left(\frac{2\mathrm{e}}{n}\right)^{\ell_h} \lesssim \mathrm{e}\left(\frac{2\mathrm{e}}{n}\right)^{\ell_h}.\end{aligned}$$

Note that the last bound also holds when $\ell_h = n$.

Step 3. Finally, let us bound $\mathbb{P}(E_b \subseteq \mathbf{E}_b)$. We shall prove that for every possible value H_h of $\mathbf{H}_h$

$$\mathbb{P}\left(E_b \subseteq \mathbf{E}_b \;\middle|\; \mathbf{H}_h = H_h\right) \leq \left(\frac{2\mathrm{e}}{n}\right)^{\ell_b}.$$

That will prove $\mathbb{P}(E_b \subseteq \mathbf{E}_b) \leq \left(\frac{2\mathrm{e}}{n}\right)^{\ell_b}$.

Fix a value $\mathbf{H}_h = H_h$. It determines the set of equidistributed vertices $R = R(H_h, r)$. Again, we arbitrarily enumerate it: $R = \{x_1, x_2, \dots, x_r\}$. Recall that $\mathbf{E}_b$ is defined as the (multi)set of edges of the form $\{x_j, \mathbf{y}[x_j]\}$, where $\mathbf{y}[x_1], \dots, \mathbf{y}[x_r]$ are drawn uniformly from all possible sequences of r *distinct* vertices. The number of ways to choose the vertices $y[x]$ such that $E_b \subseteq \mathbf{E}_b$ holds is at most $2^{\ell_b}(n - \ell_b)_{r-\ell_b}$. Indeed, there are at most 2^{ℓ_b} ways to choose a direction $x_j \to y[x_j]$ for every edge of E_b and at most $(n - \ell_b)_{r-\ell_b}$ ways to choose the remaining $y[x_j]$. Overall,

$$\mathbb{P}\left(E_b \subseteq \mathbf{E}_b \;\middle|\; \mathbf{H}_h = H_h\right) \leq 2^{\ell_b}\frac{(n-\ell_b)_{r-\ell_b}}{(n)_r} = 2^{\ell_b}\frac{1}{(n)_{\ell_b}} \leq \left(\frac{2\mathrm{e}}{n}\right)^{\ell_b}.$$

In conclusion, the three steps and the independence between the different components of $\mathbf{H}^*$ show that indeed there exists a constant c_1 such that $\mathbb{P}(A_P) \leq \left(\frac{c_1}{n}\right)^{\ell_0}$. That finishes the proof.

Finally, we address the asymmetry of $\mathbf{H}$.

THEOREM 2.15. $\mathbf{H}$ *is asymmetric w.h.p.*

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

PROOF. We consider two separate cases $m \geq \frac{3}{2}n$ and $m < \frac{3}{2}n$.

Dense case: $m \geq \frac{3}{2}n$. In this case we follow an argument due to Bollobás [5], which proves asymmetry of the random regular graph $G(n,r)$ with $r \geq 3$ fixed. This argument directly generalizes to the following result.

THEOREM 2.16 (Bollobás [5]). *Fix an integer constant* $\Delta \geq 3$. *Let* $\{\mathbf{G}_n = ([n], \mathbf{E}_n)\}_{n=1}^{\infty}$ *be a sequence of random graphs with* $\delta(\mathbf{G}_n) \geq 3$ *and* $\Delta(\mathbf{G}_n) \leq \Delta$. *Assume that there exists a constant* $c_0 > 0$ *such that* $\mathbb{P}(E_0 \subseteq E(\mathbf{G}_n)) \leq \left(\frac{c_0}{n}\right)^{|E_0|}$ *for every* $E_0 \subseteq \binom{[n]}{2}$. *Then* $\mathbf{G}_n$ *is asymmetric w.h.p.*

For the sake of completeness, we give the full proof of Theorem 2.16 in Appendix A.

Equation (2) and Proposition 2.13 show that the random balanced graph $\mathbf{H}$ satisfies the last two properties—that the maximum degree is bounded and that the probability that the given set of edges E_0 belongs to the random graph is at most $(c_0/n)^{|E_0|}$. When $m \geq \frac{3}{2}n$ it also satisfies the first property that the minimum degree is at least 3, and therefore it is asymmetric w.h.p.

Sparse case: $m < \frac{3}{2}n$. In this case $\delta(\mathbf{G}_n) = 2$ and Bollobás's argument does not apply. Instead, we develop an entirely different approach for proving asymmetry, generally applicable to randomly perturbed cycles. Indeed, in the sparse case, the regular component vanishes and $\mathbf{H} = \mathbf{H}_h + \mathbf{H}_b$. That is, $\mathbf{H}$ is a Hamilton cycle with additional $2 \leq r < \frac{n}{2}$ randomly scattered edges. Our approach is based on the observation that non-trivial automorphisms give rise to certain configurations which are very rare in the sparse case. A key concept in the proof is that of an *alternating cycle*: a cycle which alternates between Hamiltonian edges and balancing edges. Since this proof may be of its own interest and since it is long enough to interrupt the flow of the article, we present it in the separate Section 6.

2.3. The family $\mathcal{H}(n,m)$. We are now ready to prove Theorem 2.1. Given $m = m(n)$ which satisfies $n+2 \leq m \leq cn$, define $\mathcal{H}(n,m)$ as the set of all asymmetric graphs which are possible values of $\mathbf{H} = \mathbf{H}(n,m)$. That is,

$$\mathcal{H}_{n,m} = \left\{ H \text{ is asymmetric} \;\middle|\; \mathbb{P}\left(\mathbf{H}(n,m) = H\right) > 0 \right\}.$$

$\mathcal{H}(n,m)$ is indeed a family of strictly balanced and asymmetric graphs with n vertices and m edges (by the definition and from Proposition 2.9). Note that it is closed under isomorphism (since so is the family of all possible values of $\mathbf{H}$).

It remains to prove that $|\mathcal{H}(n,m)| = \exp(-O(n))n^m$. We start with an estimation of $N^* = N^*(n,m)$. Recall that

$$\begin{aligned} N^* = N_r N_h N_b &= \exp\left(-O(n)\right) n^{\left(\frac{q}{2}-1\right)n} \cdot \frac{(n-1)!}{2} \cdot (n)_r \\ &= \exp\left(-O(n)\right) n^{\left(\frac{q}{2}-1\right)n} n^n n^r = \exp\left(-O(n)\right) n^m. \end{aligned}$$

From Proposition 2.10 we have $N = \Theta(N^*)$. Combining that with Theorem 2.15, we deduce that the number of possible (H_r, H_h, H_b) such that $H_r + H_h + H_b$ is an asymmetric (simple) graph is at least $\exp(-O(n))n^m$. The only remaining issue is that we may overcount graphs: different values (H_r, H_h, H_b) may have the same sum. To handle this, the following simple lemma bounds the number of ways to

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

express a given (simple) graph as $H_r + H_h + H_b$. We allow very coarse estimations since we only care that the bound is exponential.

CLAIM 2.17. *There is a function $B(n) = \exp\big(O(n)\big)$ such that, for every graph $H \in \mathcal{H}(n,m)$, the number of possible values (H_r, H_h, H_b) of $(\mathbf{H}_r, \mathbf{H}_h, \mathbf{H}_b)$ such that $H = H_r + H_h + H_b$ is at most $B(n)$.*

PROOF. From (2) we know that $\Delta(H) \le q+2$. Therefore $(q+2)^n$ is a trivial (and very coarse) bound on the number of Hamilton cycles in H. The Hamiltonian component H_h must be one of them.

Now fix H_h, which also fixes the set R. For every $x \in R$, $y[x]$ must be its neighbor in H and can therefore be chosen in at most $q+2$ ways. So the number of choices for H_b is at most $(q+2)^r$.

The values H_h, H_b uniquely determine H_r (as the graph with all the remaining edges). Overall, the number of choices is at most

$$(q+2)^{n+r} \le (2c+2)^{\frac{3}{2}n} =: B(n),$$

where we used the inequalities $q \le 2c$ and $r < \frac{n}{2}$. That finishes the proof. ⊣

From the claim we immediately get

$$|\mathcal{H}(n,m)| \ge \frac{1}{B} \cdot \exp\big(-O(n)\big)\, n^m = \exp\big(-O(n)\big)\, n^m$$

and so Theorem 2.1 follows.

§3. The general upper bound. In this section we prove Theorem 1.2. Let us fix an irrational $\alpha \in (0,1)$ and a function $\omega = \omega(n) \to \infty$, and let $\mathbf{G}_n^1, \mathbf{G}_n^2 \sim G(n, n^{-\alpha})$ be independent. We prove that there exists a purely existential FO-sentence $\varphi = \varphi(\mathbf{G}_n^1, \mathbf{G}_n^2)$, of quantifier depth at most $\omega \frac{\ln n}{\ln\ln n}$, that distinguishes between $\mathbf{G}_n^1, \mathbf{G}_n^2$ w.h.p. To do that, it is sufficient to find a family of graphs $\mathcal{F}$ on $[v]$ where $v = \big\lfloor \omega \frac{\ln n}{\ln\ln n} \big\rfloor$ is even and satisfies the following two properties:

(A) W.h.p., $\mathbf{G}_n^1$ contains an induced subgraph which is isomorphic to some $H \in \mathcal{F}$.

(B) For any specific graph $H \in \mathcal{F}$, w.h.p. $\mathbf{G}_n^2$ does not contain an induced subgraph isomorphic to H.

Indeed, suppose we have found $\mathcal{F}$ with these properties. Let $H = H(\mathbf{G}_n^1)$ be the minimum graph from $\mathcal{F}$ (with respect to some arbitrary ordering) such that $\mathbf{G}_n^1$ contains an induced subgraph isomorphic to H. From (A), such a graph H exists w.h.p. Now let $\varphi = \varphi(\mathbf{G}_n^1)$ be an FO-sentence expressing the property of containing an induced subgraph isomorphic to H, with quantifier depth at most v. From (B), it distinguishes between $\mathbf{G}_n^1, \mathbf{G}_n^2$ w.h.p.

Our approach is to define $\mathcal{F}$ as a set of typical graphs from the family $\mathcal{H}(v,e)$ from Theorem 2.1, with e around $\frac{1}{\alpha}v$. More precisely, we set $e = \big\lceil \frac{1}{\alpha}v \big\rceil + 1$: the additional 1 makes the subgraphs dense enough to assure Property (B). We get Property (A) in the usual way using Chebyshev's inequality. To make it work, we need asymmetry and strict balancedness that are provided by Theorem 2.1. For technical reasons, we need to further refine the latter property and make sure that our family comprises

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

graphs that are *enhancely balanced*, that is, small subgraphs of $H \in \mathcal{F}$ have density slightly below $\frac{1}{\alpha}$. We will make use of the following proposition.

PROPOSITION 3.1 (Enhanced balancedness). *Consider* $\mathbf{H} = \mathbf{H}(v, e)$ *with even* $v = \left\lfloor \omega \frac{\ln n}{\ln \ln n} \right\rfloor$ *and* $e = \left\lceil \frac{1}{\alpha} v \right\rceil + 1$. *Then there exist positive constants* $\delta_0, \beta_0 > 0$ (*depending only on* α) *such that the following holds. W.h.p., for every subgraph* H_0 *of* $\mathbf{H}$ *with* $v(H_0) \leq \delta_0 v$, *we have* $e(H_0) < \frac{1}{\alpha} v(H_0) - \beta_0$.

The proof of Proposition 3.1 is postponed to Section 3.2. In Section 3.1 we use it to construct a family $\mathcal{F}$ with Properties (A) and (B) and thus complete the proof of Theorem 1.2.

3.1. A suitable family of subgraphs. Let $\mathcal{H} = \mathcal{H}(v, e)$ be the family of graphs from Theorem 2.1 with $v = \left\lfloor \omega \frac{\ln n}{\ln \ln n} \right\rfloor$, $e = \left\lceil \frac{1}{\alpha} v \right\rceil + 1$. Without loss of generality we may assume that v is even. We define $\mathcal{F} = \mathcal{F}(v, e)$ as the family of graphs $H \in \mathcal{H}$ which additionally satisfy the enhanced balancedness property from Proposition 3.1. Like $\mathcal{H}$, the family $\mathcal{F}$ is closed under isomorphism. Moreover, from Theorem 2.1, Proposition 3.1, and Claim 2.17, we have the following bound on $|\mathcal{F}|$.

CLAIM 3.2. $|\mathcal{F}| = \exp(-O(v))\,|\mathcal{H}| = \exp(-O(v))v^e$.

PROOF. The second equality follows directly from the lower bound on $|\mathcal{H}|$ from Theorem 2.1, so it remains to prove the first equality. Let N be the number of possible values (H_r, H_h, H_b) such that $H_r + H_h + H_b$ is simple, and let $N_{\mathcal{F}}$ be the number of possible values (H_r, H_h, H_b) such that $H_r + H_h + H_b \in \mathcal{F}$. Proposition 3.1 implies $\frac{N_{\mathcal{F}}}{N} = 1 - o(1)$. Claim 2.17 implies $N_{\mathcal{F}} = \exp\big(O(v)\big)\,|\mathcal{F}|$. Therefore $|\mathcal{F}| = \exp(-O(v))N_{\mathcal{F}} = \exp(-O(v))N$, implying $|\mathcal{F}| = \exp(-O(v))\,|\mathcal{H}|$. The last equality follows from the fact that $|\mathcal{H}| = \exp(O(v))N$ and $|\mathcal{F}| \leq |\mathcal{H}|$. Note that the equality $|\mathcal{F}| = \exp(-O(v))\,|\mathcal{H}|$ does not exclude the possibility that $|\mathcal{F}| = \exp(-o(v))\,|\mathcal{H}|$. ⊣

LEMMA 3.3. $\mathcal{F}$ *satisfies Property* (B).

PROOF. Fix $H \in \mathcal{F}$ and let $\mathbf{X}_H$ be the number of induced copies of H in $\mathbf{G}_2$. Then

$$\mathbb{E}\,(\mathbf{X}_H) = (n)_v\, p^e\, (1-p)^{\binom{v}{2}-e} \sim n^v\, p^e = n^{v-\alpha e},$$

since $v, e = O(\ln n)$ and $p = n^{-\alpha}$. From the definition of e we have $v - \alpha e \leq -\alpha$, so overall $\mathbb{E}\,(\mathbf{X}_H) = O\,(n^{-\alpha})$. From Markov's inequality, $\mathbb{P}\,(X_H \geq 1) = o(1)$. ⊣

LEMMA 3.4. $\mathcal{F}$ *satisfies property* (A).

PROOF. Let $\mathbf{X}_{\mathcal{F}}$ count the number of induced subgraphs $H \subseteq \mathbf{G}_n^1$ with $H \in \mathcal{F}$. We need to prove $\mathbb{P}(\mathbf{X}_{\mathcal{F}} = 0) = o(1)$. From Chebyshev's inequality, it suffices to prove $\frac{\mathrm{Var}(\mathbf{X}_{\mathcal{F}})}{[\mathbb{E}(\mathbf{X}_{\mathcal{F}})]^2} = o(1)$. Write $\mathbf{X}_{\mathcal{F}} = \sum_H \mathbb{1}_{H \subseteq \mathbf{G}_1}$, where the sum is over all subgraphs H of the complete graph K_n with $H \in \mathcal{F}$. The expected value is

$$\mathbb{E}\,(\mathbf{X}_{\mathcal{F}}) = \sum_H \mathbb{P}\left(H \subseteq \mathbf{G}_n^1\right) = \binom{n}{v} |\mathcal{F}|\, p^e\, (1-p)^{\binom{v}{2}-e} \sim n^{v-\alpha e}\, \frac{|\mathcal{F}|}{v!}.$$

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

The variance is

$$\mathrm{Var}(\mathbf{X}_{\mathcal{F}}) = \sum_{H,H'} \left[\mathbb{E}\left(\mathbb{1}_{H\subseteq \mathbf{G}_1, H'\subseteq \mathbf{G}_1}\right) - \mathbb{E}\left(\mathbb{1}_{H\subseteq \mathbf{G}_1}\right)\mathbb{E}\left(\mathbb{1}_{H'\subseteq \mathbf{G}_1}\right)\right]. \tag{3}$$

We decompose the sum in (3) by considering different possible intersection patterns of H, H'. For every non-negative integers v_0, e_0 let S_{v_0,e_0} denote the sum in (3), but only over the pairs H, H' which share exactly v_0 common vertices and e_0 common edges. Then $\mathrm{Var}(\mathbf{X}_{\mathcal{F}}) = \sum_{v_0,e_0} S_{v_0,e_0}$. When $v_0 \le 1$, the indicators $\mathbb{1}_H, \mathbb{1}_{H'}$ are independent and the corresponding summand is 0. Taking enhanced balancedness into account, we may therefore sum over the set $\mathcal{I}$ of pairs (v_0, e_0) such that $v_0 \ge 2$, $e_0 \le \frac{e}{v} v_0$ and, moreover, $e_0 \le \frac{1}{\alpha} v_0 - \beta_0$ whenever $v_0 \le \delta_0 v$. Each S_{v_0,e_0} can be written as follows:

$$\begin{aligned} S_{v_0,e_0} = &\binom{n}{v}\binom{v}{v_0}\binom{n-v}{v-v_0} \cdot T_{v_0,e_0} \\ &\times \left[p^{2e-e_0}(1-p)^{2\left(\binom{v}{2}-e\right)-\left(\binom{v_0}{2}-e_0\right)} - p^{2e}(1-p)^{2\left(\binom{v}{2}-e\right)}\right]. \end{aligned}$$

Here T_{v_0,e_0} denotes the number of possible choices of a pair of induced subgraphs H, H' on two given sets of v vertices with v_0 common vertices, such that $H, H' \in \mathcal{F}$ and they share exactly e_0 edges. Simple estimations yield

$$\frac{S_{v_0,e_0}}{\left[\mathbb{E}\left(\mathbf{X}_{\mathcal{F}}\right)\right]^2} \lesssim \frac{(v!)^2}{v_0!\left((v-v_0)!\right)^2} \cdot \frac{T_{v_0,e_0}}{|\mathcal{F}|^2} \cdot \frac{1}{n^{v_0-\alpha e_0}}. \tag{4}$$

The following claim bounds the term $T_{v_0,e_0}/|\mathcal{F}|^2$.

Claim 3.5. *For every $(v_0, e_0) \in \mathcal{I}$, $T_{v_0,e_0}/|\mathcal{F}|^2 = \exp(O(v))(\frac{c_0}{v})^{e_0}$, where c_0 is the constant from Proposition 2.13, and the term $O(v)$ does not depend on (v_0, e_0).*

Proof. Let us fix two sets of vertices V_1, V_2 with $|V_1| = |V_2| = v$ and $V_0 = V_1 \cap V_2$ having $|V_0| = v_0$. We need to bound T_{v_0,e_0}, the number of pairs H_1, H_2 such that each H_i is a graph on V_i with $H_i \in \mathcal{F}$, $H_1[V_0] = H_2[V_0]$, and the common induced subgraph contains exactly e_0 edges.

First choose $H_1 \in \mathcal{F}$ on V_1 such that $H_1[V_0]$ contains exactly e_0 edges. The number of choices is trivially bounded by $|\mathcal{F}|$. Now, given H_1, we must choose $H_2 \in \mathcal{F}$ on V_2 which contains a given set of e_0 edges. The number of choices is bounded by the number of graphs from $\mathcal{H}$ that contain the same set of edges which is, due to Proposition 2.13, at most $\left(\frac{c_0}{v}\right)^{e_0} N$, where $N = N(v, e)$ is again the number of possible values (H_r, H_h, H_b) such that $H_r + H_h + H_b$ is simple. By Claim 3.2, $|\mathcal{F}| = \exp(-O(v))N$. Therefore,

$$T_{v_0,e_0} \le |\mathcal{F}| \cdot \left(\frac{c_0}{v}\right)^{e_0} \cdot \exp\left(O(v)\right)|\mathcal{F}| = \exp\left(O(v)\right)\left(\frac{c_0}{v}\right)^{e_0}|\mathcal{F}|^2.$$

⊣

We now return to the proof of Lemma 3.4. It suffices to prove that the right-hand side in (4) is $o\left(v^{-2}\right)$ (uniformly), since the number of summands is $\Theta\left(v^2\right)$. We have

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

$$\begin{aligned}\frac{S_{v_0,e_0}}{[\mathbb{E}(\mathbf{X}_{\mathcal{F}})]^2} &= \frac{(v!)^2}{v_0!\left((v-v_0)!\right)^2}\cdot\exp\left(O(v)\right)\left(\frac{c_0}{v}\right)^{e_0}\cdot\frac{1}{n^{v_0-\alpha e_0}}\\ &\le \frac{\left((v)_{v_0}\right)^2}{\left(\frac{v_0}{\mathrm{e}}\right)^{v_0}}\cdot\exp\left(O(v)\right)\left(\frac{c_0}{v}\right)^{e_0}\cdot\frac{1}{n^{v_0-\alpha e_0}}\\ &\le \frac{v^{2v_0}}{v_0^{v_0}}\exp\left(O(v)\right)\cdot\left(\frac{c_0}{v}\right)^{e_0}\cdot\frac{1}{n^{v_0-\alpha e_0}}.\end{aligned}$$

Notice that the last expression is monotonically increasing with respect to e_0. Therefore we may bound it for every $2\le v_0\le v$ only with e_0^*, defined as the maximal e_0 such that $(v_0,e_0)\in\mathcal{I}$. Note that $e_0^*=\frac{1}{\alpha}v_0+O(1)$, so

$$\frac{S_{v_0,e_0}}{[\mathbb{E}(\mathbf{X}_{\mathcal{F}})]^2} = \underbrace{\left(\frac{v^{2-\frac{1}{\alpha}}}{v_0}\right)^{v_0}\exp\left(O(v)\right)}_{(*)}\cdot\underbrace{\frac{1}{n^{v_0-\alpha e_0^*}}}_{(**)}. \tag{5}$$

We bound (5) separately for small v_0 and for large v_0. Intuitively, for small subgraphs the enhanced balancedness property promises that $(**)$ is sufficiently small, while for large subgraphs it is $(*)$ which becomes small.

Small subgraphs. In this case we assume $v_0\le\frac{c}{\omega}v$ where $c=\frac{\alpha\beta_0}{2}$. By definition of v,

$$v_0\le\frac{c}{\omega}\left\lfloor\omega\frac{\ln n}{\ln\ln n}\right\rfloor\le c\frac{\ln n}{\ln\ln n}.$$

Of course, this implies $v_0\le\delta_0 v$, so enhanced balancedness applies: $e_0^*\le\frac{1}{\alpha}v_0-\beta_0$ and therefore $\frac{1}{n^{v_0-\alpha e_0^*}}\le n^{-\alpha\beta_0}$.

As for $(*)$, we use the simple bound

$$(*)\le v^{v_0}\left[\left(\frac{v}{v_0}\right)^{\frac{v_0}{v}}\right]^v\exp\left(O(v)\right)\le\exp\left(v_0\ln v+O(v)\right).$$

Without loss of generality, we may assume $\omega=o(\ln\ln n)$, so $v=o(\ln n)$ and

$$\exp\left(v_0\ln v+O(v)\right)\le\exp\left(c\frac{\ln n}{\ln\ln n}\ln\ln n+o(\ln n)\right)=n^{\frac{\alpha\beta_0}{2}+o(1)}.$$

Overall, $S_{v_0,e_0}/\mathbb{E}[(\mathbf{X}_{\mathcal{F}})]^2\le n^{-\frac{\alpha\beta_0}{2}+o(1)}$ uniformly. In particular it is $o\left(v^{-2}\right)$.

Large subgraphs. Now assume $v_0\ge\frac{c}{\omega}v$. In this case we only know $e_0^*\le\frac{e}{v}v_0\le\left(\frac{1}{\alpha}+\frac{2}{v}\right)v_0$, so

$$(**)=\frac{1}{n^{v_0-\alpha e_0^*}}\le n^{2\alpha\frac{v_0}{v}}.$$

As for $(*)$, we have

$$\left(\frac{v^{2-\frac{1}{\alpha}}}{v_0}\right)^{v_0}\le\left(\frac{\omega}{c}v^{1-\frac{1}{\alpha}}\right)^{v_0}=\left(\frac{\omega}{c}v^{-\frac{\varepsilon}{\zeta}}\right)^{v_0},$$

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

where $\xi = \frac{1}{\alpha} - 1$ is a positive constant. Combining both bounds,

$$\frac{S_{v_0,e_0}}{\mathbb{E}[(\mathbf{X}_{\mathcal{F}})]^2} \le \left(\omega v^{-\xi} \cdot C^{\frac{v}{v_0}} n^{\frac{2\alpha}{v}}\right)^{v_0}$$

for some constant $C > 1$. Note that

$$\omega C^{\frac{v}{v_0}} n^{\frac{2\alpha}{v}} \le \exp\left(O(\omega) + O\left(\frac{\ln n}{v}\right)\right) \le \exp\left(O(\omega) + O\left(\frac{\ln \ln n}{\omega}\right)\right).$$

Again, without loss of generality we may assume $\omega = o(\ln \ln n)$, so

$$\omega C^{\frac{v}{v_0}} n^{\frac{2\alpha}{v}} = \exp\left(o(\ln \ln n)\right) = v^{o(1)}.$$

Overall, $S_{v_0,e_0}/\mathbb{E}[(\mathbf{X}_{\mathcal{F}})]^2 \le \left(v^{-\xi+o(1)}\right)^{v_0}$ which is again $o(v^{-2})$. We have therefore completed the proof of Lemma 3.4.

3.2. Proof of Proposition 3.1. It remains to prove Proposition 3.1 about enhanced balancedness. We have $\mathbf{H} = \mathbf{H}(v, e)$ with $v \to \infty$ and $e = \left\lceil \frac{1}{\alpha} v \right\rceil + 1$. Write $\frac{1}{\alpha} = \frac{q}{2} + \xi$ where $q \ge 2$ is an integer and $0 < \xi < \frac{1}{2}$. Then $e = q \cdot \frac{v}{2} + r$ where $r = \lceil \xi v \rceil + 1 < \frac{v}{2}$.

DEFINITION 3.6. Let $\mathbf{H} = \mathbf{H}_r + \mathbf{H}_h + \mathbf{H}_b$ be a random balanced (simple) graph. Let $\mathbf{H}' = \mathbf{H}_h + \mathbf{H}_b$ be obtained by deleting the regular edges from $\mathbf{H}$. Let $\mathbf{H}''$ be obtained from $\mathbf{H}'$ by replacing every balancing edge $\{x, y[x]\}$ with a loop at x. That is, $\mathbf{H}''$ is simply the Hamilton cycle $\mathbf{H}_h$ with a loop at every vertex of $\mathbf{R}$.

DEFINITION 3.7. For a given set of vertices $V_0 \subseteq [v]$, its *segment decomposition*is $V_0 = \tau_1 \sqcup \cdots \sqcup \tau_t$, where $\tau_1, \ldots, \tau_t$ are mutually disjoint segments of the Hamiltonian component $\mathbf{H}_h$ with no Hamiltonian edge between them.

We shall use the following technical claim.

CLAIM 3.8. *For every positive integer* $\mathbf{M}$ *there exists a positive constant* β *such that the following holds. Let* $V_0 \subseteq [v]$ *be a set of* v_0 *vertices and let* $H_0'' = \mathbf{H}''[V_0]$. *If the segment decomposition of* V_0 *contains a segment of length at most* $\mathbf{M}$, *then*

$$e(H_0'') \le (1 + \xi)\, v_0 + \frac{2v_0}{v} - \beta.$$

PROOF. Fix a positive integer M. Define

$$\beta = \min_{v_0, e_0} \left((1 + \xi) v_0 - e_0\right),$$

where the minimum is over all integers v_0, e_0 with $1 \le v_0 \le M$ and $1 \le e_0 < (1 + \xi)\, v_0$. This is a minimum over a finite, fixed set, and thus β is a positive constant. Note that $\beta < 1$ by definition.

First let $V_0 = \tau \subseteq [v]$ be a segment with $v_0 \le M$ vertices. Let e_0 be the number of edges in the induced subgraph $\mathbf{H}''[\tau]$. Recall that $\mathbf{H}''$ is strictly balanced and its density is $1 + \frac{r}{v}$ (see the proof of Proposition 2.9). Also recall that $r = \lceil \xi v \rceil + 1$, therefore,

$$1 + \frac{r}{v} \in \left[(1 + \xi) + \frac{1}{v}, (1 + \xi) + \frac{2}{v}\right].$$

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

Since M is a constant, for a sufficiently large v, the inequality $e_0 < \left(1 + \frac{r}{v}\right) v_0$ implies $e_0 < (1+\xi)v_0$ (whenever $1 \le v_0 \le M$). In turn, $e_0 < (1+\xi)v_0$ implies $e_0 \le (1+\xi)v_0 - \beta$ by the definition of β.

Now let τ be a segment of any length. Again, let v_0 be its number of vertices and let e_0 be the number of edges in $\mathbf{H}''[\tau]$. In this case we can only use strict balancedness to claim that

$$e_0 < \left(1 + \frac{r}{v}\right) v_0 < (1+\xi)\, v_0 + \frac{2v_0}{v}.$$

Finally, let $V_0 \subseteq [v]$ be an arbitrary set of vertices. Let $V_0 = \tau_1 \cup \cdots \cup \tau_t$ be the segment decomposition. By definition, $\mathbf{H}''$ contains no edges between any pair τ_i, τ_j. Let v_0, e_0 denote the number of vertices and edges in $\mathbf{H}''[V_0]$ and let v_i, e_i denote the number of vertices and edges in $\mathbf{H}''[\tau_i]$. If at least one of the segments is of length $\le M$,

$$e_0 = \sum_{i=1}^{t} e_i < \sum_{i=1}^{t} \left[(1+\xi) v_i + \frac{2v_i}{v} \right] - \beta = (1+\xi)\, v_0 + \frac{2v_0}{v} - \beta.$$

That finishes the proof of the claim. ⊣

Let us now finish the proof of Proposition 3.1. Set $M = \left\lceil \frac{4}{\xi} \right\rceil$ and let β be the corresponding constant from Claim 3.8. We prove Proposition 3.1 for $\beta_0 = \frac{1}{3}\beta$ and $\delta_0 = \frac{1}{4} \min\left\{\beta, \mathrm{e}^{-\frac{2}{\xi}}\right\}$. These choices will be justified soon; for now, notice that $\beta_0 + 2\delta_0 < \beta < 1$.

Let $V_0 \subseteq [v]$ be a subset with $|V_0| = v_0 \le \delta_0 v$ and let $H_0 = \mathbf{H}[V_0]$. We begin by identifying several cases in which the desired inequality $e(H_0) < \frac{1}{\alpha} v_0 - \beta_0$ holds. The *lost edges* of V_0 are the balancing edges $\{x, \mathbf{y}[x]\}$ with $x \in V_0$ and $\mathbf{y}[x] \notin V_0$.

Fact 1. Suppose that V_0 has a lost edge. Then, from Claim 3.8,

$$e(H_0) \le \left(\frac{q}{2} - 1\right) v_0 + e\left(H_0''\right) - 1 \le \frac{1}{\alpha} v_0 + \frac{2v_0}{v} - 1 < \frac{1}{\alpha} v_0 - \beta_0.$$

Here, the first inequality follows from that fact that H_0 is obtained from $\mathbf{H}'[V_0]$ by adding some edges of the $(q-2)$-regular graph $\mathbf{H}_r$, and therefore $e(H_0) \le \left(\frac{q}{2} - 1\right) v_0 + e(\mathbf{H}'[V_0])$, while H_0'' is obtained from $\mathbf{H}'[V_0]$ by replacing its balancing edges with loops, and adding a loop for every lost edge of V_0.

Fact 2. Suppose that the segment decomposition of V_0 contains a segment of length at most M. Then, from Claim 3.8,

$$e(H_0) \le \frac{1}{\alpha} v_0 + \frac{2v_0}{v} - \beta \le \frac{1}{\alpha} v_0 + 2\delta_0 - \beta < \frac{1}{\alpha} v_0 - \beta_0.$$

Fact 3. Denote $r_0 = |V_0 \cap R|$. Then $e(H_0) \le \frac{q}{2} v_0 + r_0$. In particular, if $r_0 < \xi v_0 - \beta_0$, we again have

$$e(H_0) < \frac{1}{\alpha} v_0 - \beta_0.$$

These three facts show that in order to finish the proof, it is sufficient to prove the following statement. W.h.p., for every subset $V_0 \subseteq [v]$ with $|V_0| \le \delta_0 v$, if it is

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

composed only of segments of length at least M and satisfies $|V_0 \cap R| \geq \xi v_0 - \beta_0$, then it has a lost edge.

For $V_0 \subseteq [v]$ with $v_0 = |V_0|$ and $r_0 = |V_0 \cap R|$, the probability that it has no lost edges is precisely $\frac{(v_0)_{r_0}}{(v)_{r_0}}$ in $\mathbf{H}^*$, and $\leq \frac{1}{p_0}\frac{(v_0)_{r_0}}{(v)_{r_0}}$ in $\mathbf{H}$ (see Proposition 2.10). Assuming $r_0 \geq \xi v_0 - \beta_0$,

$$\frac{(v_0)_{r_0}}{(v)_{r_0}} \leq \left(\frac{v_0}{v}\right)^{r_0} \leq \left(\frac{v_0}{v}\right)^{\xi v_0 - \beta_0}.$$

Now consider the number of subsets $V_0 \subseteq [v]$ with v_0 vertices which are composed only of segments of length at least M. It is bounded by the number of subsets which are composed of at most $\frac{v_0}{M}$ segments, which is $O(1) \cdot \binom{v}{2v_0/M}$. By the union bound, the probability that no relevant V_0 has a lost edge is

$$\mathbb{P}\left(\text{no } V_0 \text{ loses edges}\right) = O(1) \cdot \sum_{v_0=M}^{\delta_0 v} \binom{v}{2v_0/M} \left(\frac{v_0}{v}\right)^{\xi v_0 - \beta_0}. \tag{6}$$

It remains to show that the last sum is $o(1)$. The standard bound $\binom{n}{k} \leq \left(\frac{\mathrm{e}n}{k}\right)^k$ yields

$$\begin{aligned}(6) &\leq \sum_{v_0=M}^{\delta_0 v} \left(\frac{\mathrm{e}v}{2v_0/M}\right)^{2v_0/M} \left(\frac{v_0}{v}\right)^{\xi v_0 - \beta_0} \\ &= \sum_{v_0=M}^{\delta_0 v} \left(\frac{\mathrm{e}M}{2}\right)^{\frac{2v_0}{M}} \left(\frac{v_0}{v}\right)^{\xi v_0 - \frac{2}{M}v_0 - \beta_0} \overset{(*)}{\leq} \sum_{v_0=M}^{\delta_0 v} \mathrm{e}^{v_0} \left(\frac{v_0}{v}\right)^{\frac{\xi}{2}v_0 - \beta_0}.\end{aligned}$$

The last inequality $(*)$ follows from the inequality $(\mathrm{e}x)^{\frac{1}{x}} \leq \mathrm{e}$ (applied for $x = \frac{M}{2}$) and the fact that $\frac{2}{M} < \frac{\xi}{2}$ (which follows from the definition of M).

We divide the above sum into two: the sum over $M \leq v_0 \leq \sqrt{v}$ and the sum over $\sqrt{v} \leq v_0 \leq \delta_0 v$. First,

$$\begin{aligned}\sum_{v_0=M}^{\sqrt{v}} \mathrm{e}^{v_0} \left(\frac{v_0}{v}\right)^{\frac{\xi}{2}v_0 - \beta_0} &\leq \sum_{v_0=M}^{\sqrt{v}} \mathrm{e}^{v_0} \left(\frac{1}{\sqrt{v}}\right)^{\frac{\xi}{2}v_0 - \beta_0} \\ &= \mathrm{e}^M \left(\frac{1}{\sqrt{v}}\right)^{\frac{\xi}{2}M - \beta_0} \sum_{i=0}^{\sqrt{v}-M} \left(\mathrm{e}v^{-\frac{\xi}{4}}\right)^i \lesssim \mathrm{e}^M \left(\frac{1}{\sqrt{v}}\right)^{\frac{\xi}{2}\cdot\frac{4}{\xi}-1} = o(1).\end{aligned}$$

Finally,

$$\begin{aligned}\sum_{v_0=\sqrt{v}}^{\delta_0 v} \mathrm{e}^{v_0} \left(\frac{v_0}{v}\right)^{\frac{\xi}{2}v_0 - \beta_0} &\leq \sum_{v_0=\sqrt{v}}^{\delta_0 v} \mathrm{e}^{v_0} \delta_0^{\frac{\xi}{2}v_0} \cdot \left(\frac{v}{v_0}\right)^{\beta_0} \\ &\leq v^{\beta_0/2} \sum_{v_0=\sqrt{v}}^{\delta_0 M} \left[\mathrm{e}\delta_0^{\xi/2}\right]^{v_0} = v^{\beta_0/2} \exp\left(-\Theta(\sqrt{v})\right) = o(1)\end{aligned}$$

since $\mathrm{e}\delta_0^{\xi/2} < 1$ by the definition of δ_0. That finishes the proof.

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

§4. Low upper bounds for well-approximable α. In this section we prove Theorem 1.3. Recall that for a given function $f(n) \underset{n\to\infty}{\to} \infty$ (which can grow arbitrarily slowly), our goal is to find an irrational $\alpha \in (0,1)$ and an increasing sequence of positive integers $\{n_t\}_{t=1}^{\infty}$ such that w.h.p. two independent copies $\mathbf{G}^1_{n_t}, \mathbf{G}^2_{n_t} \sim G(n_t, n_t^{-\alpha})$ can be distinguished by an FO sentence of quantifier depth at most $f(n_t)$. For the rest of this section, let us denote these copies by $\mathbf{G}_1, \mathbf{G}_2$ for notational convenience.

The construction of a suitable irrational α is explicit. The idea is to take α which is very well-approximable by rational numbers; the slower $f(n)$ grows, the better the approximations must be. Then, along a subsequence $\{n_t\}_{t=1}^{\infty}$ (which is determined by the sequence of approximations of α), distinguishing between $\mathbf{G}_1, \mathbf{G}_2$ is essentially the same as in the rational case. This idea naturally leads to the theory of Diophantine approximations, and specifically to the concept of *Liouville numbers*.

4.1. Diophantine approximations. The theory of Diophantine approximations studies approximations of real numbers by rational numbers. One of its well-known applications is the celebrated Liouville's Theorem, which was used to establish the existence of transcendental numbers[2] for the first time. A *Liouville number* is an irrational number x such that for every $d \in \mathbb{N}$, there exists a rational $\frac{p}{q}$ with $q > 1$ such that $\left|x - \frac{p}{q}\right| < \frac{1}{q^d}$. Liouville's Theorem implies that Liouville numbers are transcendental. For a comprehensive survey on this subject, see [3]. Liouville provided $x = \sum_{n=1}^{\infty} \frac{1}{2^{n!}}$ as an example of a Liouville number. For our purposes, however, we will need Liouville numbers which are much better approximated. From now on, when we write a rational number as $\frac{p}{q}$, we always assume that $q > 0$ and that $\gcd(p,q) = 1$.

Definition 4.1. Let $\varphi : \mathbb{N} \to (0,\infty)$ be a decreasing function. An irrational number x is called *φ-approximable* if, for infinitely many rational numbers $\frac{p}{q}$, $\left|x - \frac{p}{q}\right| < \varphi(q)$.

Lemma 4.2. *For every decreasing function $\varphi : \mathbb{N} \to (0,\infty)$ there exists an irrational $x \in (0,1)$ which is φ-approximable.*

Proof. First notice that if $\varphi, \psi : \mathbb{N} \to (0,\infty)$ satisfy $\varphi \le \psi$ and x is φ-approximable, then x is also ψ-approximable. Therefore, without loss of generality, we may assume $\varphi < 1$. Recursively define a sequence of natural numbers as follows: $U_1 = 1$, $U_{t+1} = \left\lceil -\log_2 \varphi\left(2^{U_t}\right)\right\rceil + 1$. Now define $x = \sum_{i=1}^{\infty} 2^{-U_i}$. Without loss of generality, we may assume that φ is decreasing sufficiently fast such that $U_{t+1} - U_t$ is a strictly increasing sequence. In that case, x is an irrational number because its binary expansion is aperiodic. For every $t \in \mathbb{N}$, consider the rational approximation $\sum_{i=1}^{t} 2^{-U_i}$ of x, which can be written as $\frac{p_t}{q_t}$ for $q_t = 2^{U_t}$ and $p_t = \sum_{i=1}^{t} 2^{U_t - U_i}$. Then

$$0 < x - \frac{p_t}{q_t} = \sum_{i=t+1}^{\infty} \frac{1}{2^{U_i}} \le \sum_{i=U_t+1}^{\infty} \frac{1}{2^i} = \frac{1}{2^{U_{t+1}-1}} \le \varphi\left(2^{U_t}\right) = \varphi(q_t).$$

⊣

[2]Recall that a real number is called *transcendental* if it is not algebraic—that is, it is not the root of a non-zero polynomial with integer coefficients.

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

As in Section 3, we distinguish between two independent $\mathbf{G}_1, \mathbf{G}_2 \sim G(n_t, n_t^{-\alpha})$ through their subgraphs; however, some details are different. First, here we count (all) subgraphs instead of induced subgraphs. We consider subgraphs with densities that approximate $\frac{1}{\alpha}$ well so that the variables counting copies of these subgraphs asymptotically behave like independent Poisson variables (as in the case of a rational α). In particular, a single subgraph suffices to distinguish $\mathbf{G}_1$ from $\mathbf{G}_2$ with positive probability. To distinguish w.h.p., we shall consider $\ell \to \infty$ different subgraphs and approximate their joint distribution.

4.2. Asymptotic Poisson behavior. We now return to the proof of Theorem 1.3. Let $g : \mathbb{N} \to \mathbb{N}$ be the inverse of f, that is, $g(k) = \min\{n \in \mathbb{N} : f(n) \geq k\}$. Without loss of generality, (1) $f : \mathbb{N} \to \mathbb{N}$ is non-decreasing and surjective, (2) $f(n) = o\left(\sqrt{\ln n}/\ln\ln n\right)$, and (3) $n/g(n)$ strictly decreases with n. Let $\alpha \in (0,1)$ be an irrational number which is φ-approximable for $\varphi(q) = 1/g(q)$; its existence is guaranteed by Lemma 4.2. Let $\{p_t\}_{t=1}^{\infty}, \{q_t\}_{t=1}^{\infty}$ be the suitable sequences from the proof of the lemma: increasing sequences of natural numbers such that, for every $t \in \mathbb{N}$, p_t, q_t are coprime and

$$0 < \alpha - \frac{p_t}{q_t} \leq \frac{1}{g(q_t)}.$$

Write $v_t = p_t$ and $e_t = q_t$ and also define $n_t = g(v_t)$. Our first assumption on f implies that $v_t = f(n_t)$. From now on we focus on the subsequence $\{n_t\}_{t=1}^{\infty}$, so the underlying parameter is now t. For convenience, we often omit the dependency on t from the notation; all quantities implicitly depend on t unless we explicitly state that they are fixed.

Let $\mathcal{H} = \mathcal{H}(v,e)$ be the set of graphs from Theorem 2.1. Recall that $\mathcal{H}$ is a family of asymmetric graphs which is closed under isomorphism. Therefore it contains $\ell = \frac{1}{v!}|\mathcal{H}|$ isomorphism classes. From the asymptotic estimation on $|\mathcal{H}|$ given in Theorem 2.1 it follows that $\ell \underset{t\to\infty}{\to} \infty$. Let $H^{(1)}, H^{(2)}, \ldots, H^{(\ell)}$ be representatives of the ℓ isomorphism classes. For every $1 \leq i \leq \ell$ let $\mathbf{X}^{(i)}$ count copies of $H^{(i)}$ in $\mathbf{G}_1$ and let $\mathbf{Y}^{(i)}$ count copies of $H^{(i)}$ in $\mathbf{G}_2$. In order to complete the proof, it is sufficient to prove the following.

PROPOSITION 4.3. *W.h.p. as $t \to \infty$, there exists $1 \leq i \leq \ell$ such that either $\mathbf{X}^{(i)} \geq 1$ and $\mathbf{Y}^{(i)} = 0$, or $\mathbf{X}^{(i)} = 0$ and $\mathbf{Y}^{(i)} \geq 1$.*

We postpone the proof of Proposition 4.3 to the very end of this section since first we have to prove several auxiliary assertions.

First of all, let us show that $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(\ell)}$ asymptotically behave like independent Poisson random variables. We do that using the method of moments (see [11, Section 6.1]). Note that the method of moments applies only when the number of variables is fixed, so we start by fixing some $k \in \mathbb{N}$ and considering only the first k variables $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(k)}$ (assuming that the underlying t is sufficiently large such that $k \leq \ell$).

Recall that for summations of Bernoulli random variables, it is easier to deal with factorial moments rather than usual moments. For a random variable $\mathbf{Z}$, its m-th factorial moment is $\mathbb{E}\left((\mathbf{Z})_m\right)$, where $(\mathbf{Z})_m = \mathbf{Z}(\mathbf{Z}-1)\ldots(\mathbf{Z}-m+1)$. For a

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

k-tuple of variables $\mathbf{Z}_1, \dots, \mathbf{Z}_k$, its $(m_1, \dots, m_k)$-th joint factorial moment equals to $\mathbb{E}\big((\mathbf{Z}_1)_{m_1} \dots (\mathbf{Z}_k)_{m_k}\big)$.

LEMMA 4.4. *Fix $k \in \mathbb{N}$ and non-negative integers $m_1, m_2, \dots, m_k$. Then, with the above definitions, we have*

$$\lim_{t\to\infty} \mathbb{E}\big((\mathbf{X}^{(1)})_{m_1} \dots (\mathbf{X}^{(k)})_{m_k}\big) = 1.$$

PROOF. Let us denote $(\vec{\mathbf{X}})_{\vec{m}} = (\mathbf{X}^{(1)})_{m_1} \dots (\mathbf{X}^{(k)})_{m_k}$ and $m = m_1 + \cdots + m_k$. Then, by decomposing $(\vec{\mathbf{X}})_{\vec{m}}$ into a sum of indicator random variables, we can write

$$\mathbb{E}(\vec{\mathbf{X}})_{\vec{m}} = \sum_{\vec{H}} \mathbb{P}\left(H_1^{(1)}, \dots, H_{m_1}^{(1)}, \dots, H_1^{(k)}, \dots, H_{m_k}^{(k)} \subseteq \mathbf{G}_1\right), \tag{7}$$

where the sum is over all m-tuples

$$\vec{H} = \left(H_1^{(1)}, \dots, H_{m_1}^{(1)}, \dots, H_1^{(k)}, \dots, H_{m_k}^{(k)}\right) \tag{8}$$

such that $H_1^{(i)}, \dots, H_{m_i}^{(i)}$ are distinct copies of $H^{(i)}$ in the complete graph K_n for every $1 \le i \le k$. Let us divide the sum in the right-hand side of (7) into two parts: $\mathbb{E}(\vec{\mathbf{X}})_{\vec{m}} = S_1 + S_2$ where S_1 is the sum over the m-tuples $\vec{H}$ which do not share any vertices with each other, and S_2 is the remaining part of the sum.

We first estimate S_1. Since $H^{(1)}, \dots, H^{(k)}$ are asymmetric, it is easy to see that

$$S_1 = (n)_{mv} p^{me} \sim n^{mv} p^{me} = (n^{v-\alpha e})^m.$$

By the definition of v, e, we have $|\alpha - v/e| \le 1/g(e)$, which can also be written as $|v - \alpha e| \le e/g(e)$. Since $n/g(n)$ decreases and $v < e$, we get $|v - \alpha e| \le v/g(v) = f(n)/n$. Since $f = o(\sqrt{\ln n}/\ln\ln n)$,

$$S_1 \sim \exp\left(\ln n \cdot O\left(\frac{f(n)}{n}\right)\right) = \exp(o(1)) = 1 + o(1).$$

It remains to prove $S_2 = o(1)$. Let $\mathcal{F}^*$ be the family of all graphs (up to isomorphism) which are the result of a union of m_i distinct copies of $H^{(i)}$ for every $1 \le i \le m$, with at least one shared vertex. For every $F \in \mathcal{F}^*$, let S_F be the sum from (7) but only over the m-tuples $\vec{H}$ whose union is isomorphic to F. Then $S_2 = \sum_{F\in\mathcal{F}^*} S_F$.

Fix $F \in \mathcal{F}^*$. Let C_F denote the number of m-tuples $\vec{H}$ as in (8) such that $H_1^{(i)}, \dots, H_{m_i}^{(i)}$ are distinct copies of $H^{(i)}$ and $H_1^{(1)} \cup \cdots \cup H_{m_k}^{(k)} = F$. Then

$$S_F = \binom{n}{v(F)} C_F p^{e(F)} \le C_F n^{v(F)} p^{e(F)} = C_F n^{v(F)-\alpha e(F)}.$$

Since $v(H^{(i)}) - \alpha e(H^{(i)}) < 0$ for all $i \in [k]$, then following the usual argument that is used to prove that the union of intersecting strictly balanced graphs with the same density has higher density, we derive the following claim.

CLAIM 4.5. *For every $F \in \mathcal{F}^*$, the inequality $v(F) - \alpha e(F) \le -\frac{c}{v}$ holds for any constant $c < \alpha$.*

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

PROOF. For a graph F we define $\xi(F) = v(F) - \alpha e(F)$. We begin with a few simple observations.

Observation 1. If H is isomorphic to one of $H^{(1)}, \dots, H^{(k)}$, we have $\xi(H) = v - \alpha e < 0$. This is because $\frac{v}{e} < \alpha$ by definition.

Observation 2. For every two graphs F_1, F_2,

$$\xi(F_1 \cup F_2) = \xi(F_1) + \xi(F_2) - \xi(F_1 \cap F_2).$$

Observation 3. Suppose H is isomorphic to one of $H^{(1)}, \dots, H^{(k)}$ and let $H_0 \subset H$ be a proper subgraph with v_0 vertices and e_0 edges. Since H is strictly balanced, we have $\frac{e_0}{v_0} < \frac{e}{v}$. Furthermore,

$$\frac{e}{v} - \frac{e_0}{v_0} = \frac{ev_0 - e_0 v}{v_0 v} \geq \frac{1}{v_0 v}.$$

Denote $\varepsilon = \frac{e}{v} - \frac{1}{\alpha}$ (which is positive). Then

$$\begin{aligned} \frac{1}{\alpha} + \varepsilon - \frac{e_0}{v_0} &\geq \frac{1}{v_0 v}, \\ v_0 + \alpha\varepsilon v_0 - \alpha e_0 &\geq \frac{\alpha}{v}, \\ \xi(H_0) = v_0 - \alpha e_0 &\geq \frac{\alpha}{v} - \alpha\varepsilon v_0. \end{aligned}$$

Since $\varepsilon = O\left(\frac{1}{g(v)}\right)$, we get that $\xi(H_0) \geq \frac{c}{v}$ for any constant $c < \alpha$ (of course, this becomes true when the underlying t is sufficiently large).

With these observations, we now prove Claim 4.5 by induction on $m \geq 2$.

We start with the induction base $m = 2$. Suppose $F = H_1 \cup H_2$ where H_1, H_2 are both from $\mathcal{H}$ and share a vertex. Then

$$\xi(F) = \xi(H_1) + \xi(H_2) - \xi(H_1 \cap H_2) <- \xi(H_1 \cap H_2).$$

$H_1 \cap H_2$ is isomorphic to some proper subgraph H_0 of H_1, therefore,

$$\xi(F) <- \xi(H_0) \leq- \frac{c}{v}.$$

Now, suppose the statement is true for $m \geq 2$ and prove it for $m+1$. Let $F = \bigcup_{i=1}^{m+1} H_i$ be the union of certain copies of graphs from $\mathcal{H}$ such that, without loss of generality, H_1, H_2 share a vertex. Let $F' = \bigcup_{i=1}^{m} H_i$. Then

$$\xi(F) = \xi(F') + \xi(H_{m+1}) - \xi(F' \cap H_{m+1}).$$

$F' \cap H_{m+1}$ is isomorphic to a subgraph H_0 of H_{m+1} (not necessarily proper this time). If $H_0 = H_{m+1}$ then $\xi(F) = \xi(F')$. If $H_0 \subset H_{m+1}$ then the third observation shows that $\xi(H_0) > 0$, but $\xi(H_{m+1}) < 0$ and therefore $\xi(F) < \xi(F')$. From the inductive assumption, in both cases $\xi(F) \leq \xi(F') <- \frac{c}{v}$ and that finishes the proof. ⊣

From Claim 4.5, we can write $S_F \leq C_F n^{-\frac{c}{v}}$ for any constant $c < \alpha$. Finally, $\sum_{F \in \mathcal{F}^*} C_F$ is trivially bounded by $\left((mv)_v\right)^m$. Overall,

$$S_2 \leq (mv)^{mv} n^{-\frac{c}{v}} = \exp\left(\Theta(v \ln v) - \Theta\left(\frac{\ln n}{v}\right)\right) = o(1),$$

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

where the last estimation follows from the assumption $v = f(n) = o\left(\frac{\sqrt{\ln n}}{\ln \ln n}\right)$. That finishes the proof.

From Lemma 4.4 and the method of moments [11, Theorem 6.2] we get the following corollary.

COROLLARY 4.6. *Fix $k \in \mathbb{N}$; then $\vec{\mathbf{X}} = (\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(k)})$ converges in distribution to $\vec{\mathbf{P}} = (\mathbf{P}^{(1)}, \dots, \mathbf{P}^{(k)})$ where $\mathbf{P}^{(1)}, \dots, \mathbf{P}^{(k)} \sim \mathrm{Pois}(1)$ are independent.*

We are now ready to prove Proposition 4.3.

PROOF OF PROPOSITION 4.3. Let $\vec{\mathbf{I}} = (\mathbb{1}_{\mathbf{X}^{(1)} \geq 1}, \dots, \mathbb{1}_{\mathbf{X}^{(\ell)} \geq 1})$ and $\vec{\mathbf{J}} = (\mathbb{1}_{\mathbf{Y}^{(1)} \geq 1}, \dots, \mathbb{1}_{\mathbf{Y}^{(\ell)} \geq 1})$. We need to prove $\mathbb{P}(\vec{\mathbf{I}} = \vec{\mathbf{J}}) = o(1)$. For every fixed $k \in \mathbb{N}$, let $\vec{\mathbf{I}}_k$ and $\vec{\mathbf{J}}_k$ consist of the first k coordinates of $\vec{\mathbf{I}}$ and $\vec{\mathbf{J}}$, respectively. Then $\mathbb{P}(\vec{\mathbf{I}} = \vec{\mathbf{J}}) \leq \mathbb{P}(\vec{\mathbf{I}}_k = \vec{\mathbf{J}}_k)$. From Corollary 4.6, $\mathbb{P}(\vec{\mathbf{I}}_k = \vec{\mathbf{J}}_k) \to \lambda^k$, where

$$\lambda = \mathbb{P}\left(\mathrm{Pois}(1) = 0\right)^2 + \mathbb{P}\left(\mathrm{Pois}(1) \geq 1\right)^2 = \mathrm{e}^{-2} + \left(1 - \mathrm{e}^{-1}\right)^2.$$

Since this is true for every $k \in \mathbb{N}$, we deduce $\mathbb{P}(\vec{\mathbf{I}} = \vec{\mathbf{J}}) = o(1)$ and that finishes the proof. ⊣

§5. A lower bound for almost every α. In this section we prove Theorem 1.4. That is, we fix $d_0 > 2$ and an irrational $\alpha \in (0, 1)$ with Liouville–Roth irrationality measure strictly smaller than d_0 and show that w.h.p. there is no FO-sentence of quantifier depth less than $\frac{1}{\ln d_0} \ln \ln \ln n$ distinguishing between two independent copies $\mathbf{G}_n^1, \mathbf{G}_n^2 \sim G(n, n^{-\alpha})$. To prove this, we present a generalization of Theorem 1.1. The proof of the zero-one law for irrational α relies on the existence of a winning strategy of the second player in the Ehrenfeucht–Fraïssé game with a bounded number of rounds. We show that a similar winning strategy can be used by the second player when the number of rounds grows with n sufficiently slowly, provided that α is not too well-approximable by rational numbers.

From now on we fix $\alpha \in (0, 1)$ with Liouville–Roth irrationality measure strictly smaller than d_0. Let $d \in (2, d_0)$ be such that α is not q^{-d}-approximable, that is, $\left|x - \frac{p}{q}\right| \geq \frac{1}{q^d}$ for all but finitely many $\frac{p}{q}$.

5.1. The Ehrenfeucht–Fraïssé game. We begin with a brief overview of the Ehrenfeucht–Fraïssé game (see Chapter 2 of [22] for a more detailed exposition). Given two graphs G_1, G_2 (with disjoint vertex sets) and $k \in \mathbb{N}$, the Ehrenfeucht–Fraïssé game $\mathrm{EF}(G_1, G_2; k)$ is described as follows. The game has two players, called Spoiler and Duplicator, and consists of k rounds. The graphs G_1, G_2 are the "board" on which the players make their moves. In the i-th round, Spoiler selects a vertex in either graph (to his choice) and marks it i. Duplicator responds by selecting a vertex in the other graph, and also marks it i. At the end of the game, let $x_1, \dots, x_k$ be the vertices of G_1 marked $1, \dots, k$ (regardless of who marked them) and let $y_1, \dots, y_k$ be the vertices of G_2 marked $1, \dots, k$. Duplicator wins if there exists a partial isomorphism from G_1 to G_2 which maps x_i to y_i for every $1 \leq i \leq k$. That is, Duplicator wins if $x_i = x_j \iff y_i = y_j$ and $x_i \sim x_j \iff y_i \sim y_j$ for

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

every $1 \le i, j \le k$. The importance of the Ehrenfeucht–Fraïssé game comes from the following key result, relating it to the FO distinguishability.

THEOREM 5.1 (Ehrenfeucht [6]). *Let G_1, G_2 be two graphs and let $k \in \mathbb{N}$. Then Duplicator has a winning strategy in* $\mathrm{EF}(G_1, G_2; k) \iff k(G_1, G_2) > k$.

From this result we see that, in order to prove Theorem 1.4, it suffices to prove that w.h.p. Duplicator has a winning strategy in $\mathrm{EF}\left(\mathbf{G}_n^1, \mathbf{G}_n^2; k = \frac{1}{\ln d_0} \ln\ln\ln n\right)$. Below, we show that the winning strategy of Duplicator introduced by Shelah and Spencer (*look-ahead strategy*) can be used for k rounds as well. We present the strategy and explain the main properties of the random graph that allow Duplicator to use it. Proofs of these properties are postponed to Appendix B since they resemble the proof in the case of constant number of rounds. For the original argument, we refer to [22, Chapters 4–6].

5.2. Extensions and the closure. The look-ahead strategy is based on the validity of certain $\forall\exists$-sentences known as extension statements. The original argument considers statements with a constant number of variables, but now we allow it to grow with n. We begin by generalizing the concept of safe extensions and their main feature: they w.h.p. exist over any tuple of root vertices.

A *rooted graph* is a graph H with a designated subset of root vertices R. It is denoted by the pair (R, H). We allow $R = \emptyset$ but not $R = V(H)$. Rooted graphs can be used to define graph extensions as follows. Let (R, H) be a rooted graph and label its vertices $a_1, \dots, a_r, b_1, \dots, b_v$ where a_i are the roots and b_j are the non-roots. Let G be any graph and let $\vec{x} = (x_1, \dots, x_r)$ be an r-tuple of distinct vertices of G. An (R, H)-*extension* of $\vec{x}$ is defined as a v-tuple $\vec{y} = (y_1, \dots, y_v)$ of distinct vertices of G such that: (1) $x_i \sim y_j$ in G whenever $a_i \sim b_j$ in H; (2) $y_i \sim y_j$ in G whenever $b_i \sim b_j$ in H. Note that this definition does not consider edges between root vertices, and allows for additional edges except those specified by H. Also note that the definition implicitly assumes an underlying labeling of the vertices of H. From now on, whenever a rooted graph (R, H) is introduced, we always implicitly assume that it is equipped with a predetermined labeling.

A rooted graph has three parameters: the number of root vertices r, the number of non-root vertices v, and the number of edges (excluding edges between root vertices) e. We call (v, e) the *type* of the rooted graph. The following definitions depend on the fixed α, and are designed to be used in studying extensions in $\mathbf{G} \sim G(n, n^{-\alpha})$.

DEFINITION 5.2. Let (R, H) be a rooted graph of type (v, e). If $v - \alpha e > 0$ we say that (R, H) is *sparse*. If $v - \alpha e < 0$ we say that (R, H) is *dense*.

Since α is irrational, every rooted graph is either sparse or dense.

DEFINITION 5.3. Let (R, H) be a rooted graph and let $R \subseteq S \subseteq V(H)$. If $R \neq S$, then we call $(R, H[S])$ a *subextension* of (R, H). If $S \neq V(H)$, then we call (S, H) a *nailextension*of (R, H).

Note that (R, H) is always a subextension and a nailextension of itself.

DEFINITION 5.4. A rooted graph (R, H) is called *safe* if all its subextensions are sparse, and *rigid* if all its nailextensions are dense.

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

The next step is to generalize the Finite Closure Theorem [22, Theorem 4.3.2] (we rename it as Finite Closure Lemma). We first recall the definition of the *t-closure* of a set of vertices.

DEFINITION 5.5. For a graph G, a subset $U \subseteq V(G)$, and $t \in \mathbb{Z}_{\geq 0}$, we define *the t-closure of* U, denoted $\mathrm{cl}_t(U)$, as follows: $\mathrm{cl}_t(U)$ is the minimum set of vertices which contains U and is closed with respect to taking rigid extensions with at most t non-roots.

Now consider $\mathbf{G} \sim G(n, n^{-\alpha})$. In what follows, we deal with rooted graphs (R, H) of type (v, e) with r roots. Importantly, we now allow (R, H) to depend on n, so $r = r(n)$, $v = v(n)$, and $e = e(n)$. As we shall see, the original argument can be stretched up to r, v, e which grow as $(\ln n)^{1/(10d)}$. Thus, we denote $M = M(n) = (\ln n)^{1/(10d)}$ and always assume that $r, v, e = O(M)$.

LEMMA 5.6 (Bounded Closure Lemma). *There exists a constant* C *such that w.h.p. in* $\mathbf{G}$, *for every* $r, t \leq M$, *the* t*-closure of every* r*-tuple* $\vec{x}$ *has size at most* $r + Crt^d$.

Finally, we generalize [22, Theorem 5.3.1] about the existence of *generic extensions*.

DEFINITION 5.7. Let (R, H) be a rooted graph of type (v, e) with r roots. Fix a non-negative integer t. For tuples $\vec{x} = (x_1, \ldots, x_r)$ and $\vec{y} = (y_1, \ldots, y_v)$, we say that an (R, H)-extension $\vec{y}$ of $\vec{x}$ is *t-generic* if: (1) There are no additional edges between the vertices of $\vec{y}$ or from $\vec{y}$ to $\vec{x}$ other than those specified by H. (2) If any $\vec{z} = (z_1, \ldots, z_s)$ with $s \leq t$ forms a rigid extension over $\vec{x} \cup \vec{y}$, then there are no edges between $\vec{z}$ and $\vec{y}$.

LEMMA 5.8. *W.h.p. in* $\mathbf{G}$, *for every safe rooted graph* (R, H) *with* $r \leq M$ *roots and* $v \leq M$ *non-roots and for every* $t \leq M$, *we have that every* r*-tuple of vertices* $\vec{x}$ *has a* t*-generic* (R, H)*-extension* $\vec{y}$.

5.3. Look-ahead strategy. The look-ahead strategy guarantees a win for Duplicator in the Ehrenfeucht–Fraïssé game. The previous results allow us to apply the look-ahead strategy in $\mathrm{EF}(\mathbf{G}_n^1, \mathbf{G}_n^2, \frac{1}{\ln d_0} \ln \ln \ln n)$ where $\mathbf{G}_n^1, \mathbf{G}_n^2 \sim G(n, n^{-\alpha})$ are independent. Let us recall relevant definitions.

DEFINITION 5.9. Let $\vec{x} = (x_1, \ldots, x_r)$ be a tuple of vertices in a graph G_1 and let $\vec{y} = (y_1, \ldots, y_r)$ be a tuple of vertices in a graph G_2. Also let $t \in \mathbb{Z}_{\geq 0}$. We say that the t-closures $\mathrm{cl}_t(\vec{x})$ and $\mathrm{cl}_t(\vec{y})$ are isomorphic, and write $\mathrm{cl}_t(\vec{x}) \cong \mathrm{cl}_t(\vec{y})$, if there exists a graph isomorphism from $\mathrm{cl}_t(\{x_1, \ldots, x_r\})$ to $\mathrm{cl}_t(\{y_1, \ldots, y_r\})$ that sends x_i to y_i for every $1 \leq i \leq r$. The t-type of a tuple $\vec{x}$ is the isomorphism class of $\mathrm{cl}_t(\vec{x})$ with fixed $\vec{x}$.

DEFINITION 5.10. Given a sequence of non-negative integers $0 = t_0, t_1, \ldots, t_{k-1}$, a *look-ahead* $(t_0, t_1, \ldots, t_{k-1})$*-strategy* for Duplicator in $\mathrm{EF}(G_1, G_2, k)$ is a strategy that satisfies the following condition. For every $0 \leq i \leq k-1$, when there are i rounds remaining in the game, the t_i-types of the already marked vertices are the same in both graphs.

The following lemma (see [22, Chapter 6]) summarizes the deterministic part of the original argument about the existence of a look-ahead strategy.

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

LEMMA 5.11 [22]. *Let G_1, G_2 be two graphs and let ℓ, t be non-negative integers. Let $x_1, \dots, x_\ell \in G_1$ and $y_1, \dots, y_\ell \in G_2$ be two tuples with the same* t*-type. Assume that $u \geq t$ is an integer such that* t*-closure of any $(\ell+1)$-tuple in G_1 has at most $u-1$ non-roots (so at most $\ell+u$ vertices overall). Also assume that for every rooted graph (R,H) with $r \leq \ell+u$ roots and $v \leq \ell+u$ non-roots, every* r*-tuple in G_2 has a* u*-generic (R,H)-extension. Then, for every $x_{\ell+1} \in G_1$ (representing Spoiler's move) there exists $y_{\ell+1} \in G_2$ such that $x_1, \dots, x_{\ell+1} \in G_1$ and $y_1, \dots, y_{\ell+1} \in G_2$ have the same* u*-type.*

We can now prove the existence of a look-ahead strategy in $\mathrm{EF}(\mathbf{G}_n^1, \mathbf{G}_n^2, k)$, and thus complete the proof of Theorem 1.4.

PROOF OF THEOREM 1.4. Let $\mathbf{G}_n^1, \mathbf{G}_n^2 \sim G(n, n^{-\alpha})$ be independent and set $k = \frac{1}{\ln d_0} \ln\ln\ln n$. Let $\mathcal{A}$ be the event that the conclusions of Bounded Closure Lemma and Lemma 5.8 hold in both $\mathbf{G}_n^1$ and $\mathbf{G}_n^2$. Then $\mathbb{P}(\mathcal{A}) = 1 - o(1)$. We show that, given the event $\mathcal{A}$, Duplicator has a winning strategy in $\mathrm{EF}(\mathbf{G}_n^1, \mathbf{G}_n^2, k)$.

Construct a sequence $t_0 \leq t_1 \leq \cdots \leq t_{k-1}$ inductively as follows: $t_0 = 0$, and given t_i, we take $t_{i+1} \geq t_i$ such that the t_i-closure of any $(k-i)$-tuple in both $\mathbf{G}_n^1$ and $\mathbf{G}_n^2$ has at most $t_{i+1}-1$ non-roots. Note that these are the conditions from Lemma 5.11 with $t = t_i$, $\ell = k-i-1$ and $u = t_{i+1}$. From the Bounded Closure Lemma, a suitable t_{i+1} exists, and can be taken to be $t_{i+1} = C(k-i)t_i^d + 1$ where $C = C_\alpha$ is a constant (note that it also promises $t_{i+1} \geq t_i$ as long as we take $C \geq 1$). From Theorem 5.8 (with M replaced by $2M$), for every rooted graph (R,H) with $r \leq 2M$ roots and $v \leq 2M$ non-roots and for every $t \leq 2M$, we have that in both $\mathbf{G}_n^1$ and $\mathbf{G}_n^2$, every r-tuple has a t-generic (R,H)-extension.

Note that if $t_{i+1} \leq M$ then the assumptions of Lemma 5.11 apply with t, ℓ, u as above. Indeed, we have $\ell \leq k$ and $k = o(M)$, so $\ell + u \leq 2M$ and $u \leq M$, and so for every rooted graph (R,H) with $r \leq \ell+u$ roots and $v \leq \ell+u$ non-roots, every r-tuple has a u-generic (R,H)-extension. We deduce that Duplicator can follow a look-ahead $(t_0, t_1, \dots, t_{k-1})$-strategy, provided that $t_{k-1} \leq M$. To show that $t_{k-1} \leq M$ we consider the recurrence relation defining t_i. We have $t_0 = 0$, $t_1 = 1$, and for every $1 \leq i \leq k-1$,

$$t_{i+1} = C(k-i)t_i^d + 1 \leq Ckt_i^d + 1 \leq \tilde{C}kt_i^d,$$

where $\tilde{C} = 2C$. We deduce $t_i \leq (\tilde{C}k)^{n_i}$ where $n_1 = 0$ and $n_{i+1} = 1 + dn_i$, implying $n_i \leq d^i$. Overall we have

$$\begin{aligned} t_{k-1} &\leq \left(\tilde{C}k\right)^{d^k} = \exp\left(d^k \ln\left(\tilde{C}k\right)\right) \\ &= \exp\left(\exp\left(\frac{\ln d}{\ln d_0} \ln\ln\ln n\right) \cdot (1+o(1)) \ln\ln\ln\ln n\right) \\ &= \exp\left(o(\ln\ln n)\right) = o(M), \end{aligned}$$

where we used the assumption $d < d_0$. That finishes the proof. ⊣

§6. Asymmetry: Sparse case. In this section we prove the asymmetry of the random balanced graph for the sparse case $m < \frac{3}{2}n$, thus completing the proof of Theorem 2.15.

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

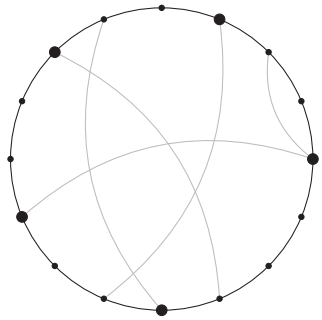


FIGURE 1. A sample of $\mathbf{H}$ with $n = 16$ and $r = 5$. Vertices of R are marked by big black circles.

Fix $m = m(n)$ with $n + 2 \leq m < \frac{3}{2}n$ and let $\mathbf{H} = \mathbf{H}(n, m)$. Write $m = 2 \cdot \frac{n}{2} + r$ where $2 \leq r < \frac{n}{2}$. The regular component of $\mathbf{H}$ is therefore empty and $\mathbf{H} = \mathbf{H}_h + \mathbf{H}_b$. That is, $\mathbf{H}$ is a random n-cycle with additional r random balancing edges. Note that the lower bound $m \geq n + 2$ cannot be improved: a cycle with one additional edge has a single non-trivial isomorphism.

We first expose the edges of $\mathbf{H}_h$. Due to symmetry, it is sufficient to prove that w.h.p. $\mathbf{H}$ is asymmetric subject to $\mathbf{H}_h = H_h := (1, 2, \dots, n, 1)$. Let $R = R(H_h, r)$ be the set of almost equidistributed vertices. The only random component is therefore the endpoints $\{\mathbf{y}[x] : x \in R\}$, which determine the balancing edges $\{x, \mathbf{y}[x]\}$.

For a more convenient description of the proof, let us color the Hamiltonian edges in black and the balancing edges in gray. Since we are working with the simple graph $\mathbf{H}$, every edge is assigned with exactly one color. The black edges are deterministic while the gray edges are random (see Figure 1).

The rest of the section is organized as follows. In Section 6.1 we list several configurations which occur in $\mathbf{H}$ with probability $o(1)$. A key concept in the proof is that of an *alternating cycle*, which is a cycle with edges of alternating colors. We show that w.h.p. $\mathbf{H}$ contains no two large alternating cycles with equal degree sequences, no two small alternating cycles which are connected by an alternating path, no small alternating cycles with additional internal edges, and does not contain some other rare configurations. In Section 6.2 we first show deterministically that every automorphism that does not preserve colors generates a rare configuration and then we show that w.h.p. $\mathbf{H}$ does not have non-trivial color-preserving automorphisms. This completes the proof as Section 6.1 establishes the absence of the rare configurations w.h.p.

6.1. Rare configurations.

DEFINITION 6.1. An *alternating walk* in $\mathbf{H}$ is a walk $v_1, v_2, \dots, v_k$ such that colors of the edges $\{v_1, v_2\}, \{v_2, v_3\}, \dots, \{v_{k-1}, v_k\}$ alternate between black and gray. An *alternating path* is an alternating walk without repeated vertices. An *alternating cycle* is an alternating walk with $v_1 = v_k$ and no other repetitions. Note that an odd alternating cycle contains two adjacent edges of the same color.

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

Let us call labeled cycles $C = (x_1, \dots, x_\ell)$, $C' = (x'_1, \dots, x'_\ell)$ *equipotent* in $\mathbf{H}$, if, for every i, vertices x_i and x'_i have equal degrees, and for every $i \in [\ell]$, edges $\{x_i, x_{i+1}\}$ and $\{x'_i, x'_{i+1}\}$ have different colors. Here, $\ell + 1 := 1$.

PROPOSITION 6.2. *W.h.p.* $\mathbf{H}$ *does not contain two different (though their sets of vertices may coincide) equipotent alternating cycles of length at least* $\ln^2 n$. *In particular, if* $r = o(n)$, *then w.h.p. there are no alternating cycles.*

REMARK 6.3. We make two simple observations:

- Since $r < \frac{n}{2}$, every black edge has at most one endpoint from R.
- By definition, every gray edge has at least one endpoint from R.

PROOF OF PROPOSITION 6.2. Let us first assume that $0 \le r \le \xi n$ for a positive constant $\xi < \frac{1}{2}$ and prove an even stronger statement: either $r = o(n)$, and then w.h.p. $\mathbf{H}$ contains no alternating cycles, or $r = \Theta(n)$, and then w.h.p. $\mathbf{H}$ contains no alternating cycles of length greater than $\ln n$.

We apply Markov's inequality. Let $\mathbf{X}_\ell$ be the number of alternating cycles of length ℓ in $\mathbf{H}$. Write it as the sum of indicators: $\mathbf{X}_\ell = \sum_C \mathbb{1}_{\{C \subseteq \mathbf{H}\}}$ where the sum is over all possible alternating cycles of length ℓ.

First, assume that $\ell \ge 4$ is even. An alternating cycle C of length ℓ must consist of $\frac{\ell}{2}$ black edges and $\frac{\ell}{2}$ gray edges. From Remark 6.3 it follows that C contains exactly $\frac{\ell}{2}$ vertices from R. Every edge in C has exactly one endpoint from R. It can be therefore written as

$$u_1 \sim w_1 \sim u_2 \sim w_2 \sim \dots \sim u_{\frac{\ell}{2}} \sim w_{\frac{\ell}{2}} \sim u_1, \tag{9}$$

where $u_i \in R$ and $w_i \in R^c$.

CLAIM 6.4. *Given* $\mathbf{R}$, *the number of possible alternating cycles of length* ℓ *is at most* $2^{\frac{\ell}{2}} (r)_{\frac{\ell}{2}}$.

PROOF. Consider the following procedure:

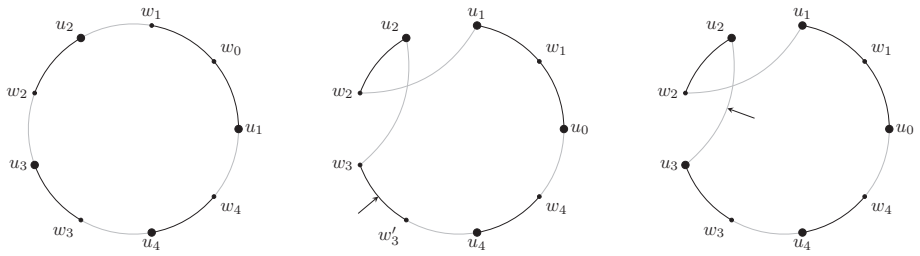
1. Choose a sequence $u_1, u_2, \dots, u_{\frac{\ell}{2}}$ of distinct vertices from R.
2. For every $1 \le i \le \frac{\ell}{2}$, choose one of the two black edges incident to u_i, denote it $\{u_i, w_i\}$.

This procedure defines a single possible alternating cycle (9) of length ℓ. Every possible alternating cycle can be obtained from the procedure (perhaps in more than one way). The number of possible choices is $2^{\frac{\ell}{2}} (r)_{\frac{\ell}{2}}$. This proves the claim. ⊣

Given a possible alternating cycle C, the probability of $C \subseteq \mathbf{H}$ is the probability that given $\frac{\ell}{2}$ edges, each with exactly one endpoint from R, appear as gray edges in $\mathbf{H}$. This event can be interpreted as determining the values of $\frac{\ell}{2}$ out of the r random vertices $\{\mathbf{y}[x] : x \in R\}$. In $\mathbf{H}^*$, the probability of this event is $\frac{1}{(n)_{\frac{\ell}{2}}}$. Therefore, by Proposition 2.10, in $\mathbf{H}$ the probability of this event is at most $\frac{p_0}{(n)_{\frac{\ell}{2}}}$ where p_0 is a constant. Overall,

$$\mathbb{E}(\mathbf{X}_\ell) \le p_0 2^{\frac{\ell}{2}} \frac{(r)_{\frac{\ell}{2}}}{(n)_{\frac{\ell}{2}}} \le p_0 \left(\frac{2r}{n}\right)^{\frac{\ell}{2}}. \tag{10}$$

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press



(a) Type 1. (b) Type 2. (c) Type 3.

FIGURE 2. Three types of b-cycles, $\ell = 9$. In types 2 and 3, an arrow points at the "special edge" whose choice adds a factor of $\Theta(\ell)$ to the bound. The cycles are drawn such the black edges are fixed, thus emphasizing the process of choosing the gray edges.

Now assume that $\ell \geq 3$ is odd. Write $\mathbf{X}_\ell = \mathbf{X}_\ell^{[b]} + \mathbf{X}_\ell^{[r]}$ where $\mathbf{X}_\ell^{[b]}$ is the number of alternating cycles with $\frac{\ell+1}{2}$ gray edges (*g-cycles*) and $\mathbf{X}_\ell^{[r]}$ is the number of alternating cycles with $\frac{\ell+1}{2}$ black edges (*b-cycles*).

We start with $\mathbf{X}_\ell^{[b]}$. Let C be a possible g-cycle of length ℓ. It has a single vertex u_0 incident to two (potentially) gray edges. From Remark 6.3 it follows that C contains exactly $\frac{\ell+1}{2}$ vertices from R. Without loss of generality, we can express any g-cycle of length ℓ as

$$u_0 \sim u_1 \sim w_1 \sim u_2 \sim w_2 \sim \ldots \sim u_{\frac{\ell-1}{2}} \sim w_{\frac{\ell-1}{2}} \sim u_0,$$

where $u_i \in R$, $w_i \in R^c$. Indeed, since all $y[x]$ have to be distinct, the only common vertex u_0 of two gray edges have to be in R. Then, exactly one of the two gray edges adjacent to u_0 must have both endpoints from R. Now we can similarly claim that the number of possible g-cycles of length ℓ is at most $2^{\frac{\ell-1}{2}}(r)_{\frac{\ell+1}{2}}$. Moreover, the event $C \subseteq \mathbf{H}$ now determines the values of $\frac{\ell+1}{2}$ random vertices $\mathbf{y}[x]$, therefore its probability in $\mathbf{H}$ is $\leq \frac{p_0}{(n)_{\frac{\ell+1}{2}}}$. Overall,

$$\mathbb{E}\left(\mathbf{X}_\ell^{[b]}\right) \leq p_0 2^{\frac{\ell-1}{2}} \frac{(r)_{\frac{\ell+1}{2}}}{(n)_{\frac{\ell+1}{2}}} \leq \frac{p_0}{2}\left(\frac{2r}{n}\right)^{\frac{\ell+1}{2}}.$$

It remains to estimate the expectation of $\mathbf{X}_\ell^{[r]}$. Let C be a possible b-cycle of length ℓ. From Remark 6.3 it follows that the number of R-vertices in C is either $\frac{\ell-1}{2}$ or $\frac{\ell+1}{2}$. Moreover, C contains a black 2-path, and at least one of its endpoints must be from R (otherwise there cannot be at least $\frac{\ell-1}{2}$ R-vertices in C). Let xyz be the black 2-path, where $x \in R$. It is possible that $z \in R$ as well. If $z \notin R$ then necessarily C contains exactly $\frac{\ell-1}{2}$ R-vertices. Overall we have three different types of possible b-cycles; we bound the expected number of cycles of each type separately. The three types are demonstrated in Figure 2.

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

Type 1. There are $\frac{\ell-1}{2}$ R-vertices and $z \notin R$. Cycles of these type behave exactly like even alternating cycles, with one black edge is replaced by a black 2-path. In this case we relabel $x = u_1, y = w_0, z = w_1$, and then C can be written as

$$u_1 \sim w_0 \sim w_1 \sim u_2 \sim w_2 \sim \ldots \sim u_{\frac{\ell-1}{2}} \sim w_{\frac{\ell-1}{2}} \sim u_1,$$

where $u_i \in R$ and $w_i \in R^c$. The number of possible cycles of this type is bounded by $2^{\frac{\ell-1}{2}} (r)_{\frac{\ell-1}{2}}$. Since $\mathbb{P}(C \subseteq \mathbf{H}) \leq \frac{p_0}{(n)_{\frac{\ell-1}{2}}}$, the expected number of b-cycles of type 1 is at most $p_0 \left(\frac{2r}{n}\right)^{\frac{\ell-1}{2}}$.

Type 2. There are $\frac{\ell-1}{2}$ R-vertices and $z \in R$. Then there exists a single black edge which has *no* endpoints in R. Let $e_0, e_1, e_2, \ldots, e_{\frac{\ell-1}{2}}$ be the sequence of black edges in C, starting from $e_0 = \{x, y\}, e_1 = \{y, z\}$, and going in the cyclic order. Let $2 \leq j \leq \frac{\ell-1}{2}$ be such that e_j is the black edge with no endpoint from R. Write $e_j = \{w_j, w_j'\}$. For $1 \leq i \leq \frac{\ell-1}{2}$, $i \neq j$ write $e_i = \{u_i, w_i\}$ where $u_i \in R, w_i \in R^c$ (so in particular $y = w_1$ and $z = u_1$) and also let $x = u_0$. The gray neighbors of w_j, w_j' in C must be u_{j-1}, u_{j+1}; without loss of generality, assume that the gray edges are $\{u_{j-1}, w_j\}$ and $\{w_j', u_{j+1}\}$. We get the following cycle:

$$\begin{aligned} &u_0 \sim w_1 \sim u_1 \sim w_2 \sim u_2 \sim \ldots \sim w_{j-1} \sim u_{j-1} \\ &\sim w_j \sim w_j' \sim u_{j+1} \sim w_{j+1} \sim \ldots \sim u_{\frac{\ell-1}{2}} \sim w_{\frac{\ell-1}{2}} \sim u_0. \end{aligned} \tag{11}$$

The conclusion of this analysis is that the number of possible cycles of type 2 is at most the number of choices in the following procedure:

1. Choose an index $2 \leq j \leq \frac{\ell-1}{2}$.
2. Choose a sequence of $\frac{\ell-3}{2}$ distinct vertices from R and denote them u_i for $i \in \{0\} \cup \left\{2, 3, \ldots \frac{\ell-1}{2}\right\} \setminus \{j\}$.
3. For u_0 choose one of the two black 2-paths starting from it, denote it $\{u_0, w_1, u_1\}$. If $u_1 \notin R$, halt.
4. For every other u_i choose one of the two black edges incident to it, denote it $\{u_i, w_i\}$.
5. Choose an additional black edge with no endpoints in R. Also choose one of its vertices to be denoted w_j, and denote the second one w_j'.

The procedure indeed uniquely defines a possible cycle of type 2: the one described in (11). As explained, every possible cycle of type 2 can be obtained from it. The number of choices in the procedure, which bounds the number of possible cycles of type 2, is at most $\ell \cdot (r)_{\frac{\ell-3}{2}} \cdot 2^{\frac{\ell-3}{2}} \cdot 2n$. Like before, we still have $\mathbb{P}(C \subseteq \mathbf{H}) \leq \frac{p_0}{(n)_{\frac{\ell-1}{2}}}$. In conclusion, the expected number of b-cycles of type 2 is at most

$$2p_0\ell \cdot 2^{\frac{\ell-3}{2}} \frac{n(r)_{\frac{\ell-3}{2}}}{(n)_{\frac{\ell-1}{2}}} = 2p_0\ell \cdot 2^{\frac{\ell-3}{2}} \frac{(r)_{\frac{\ell-3}{2}}}{(n-1)_{\frac{\ell-3}{2}}} \leq 2p_0\ell \cdot \left(\frac{2r}{n-1}\right)^{\frac{\ell-3}{2}}.$$

Type 3. There are $\frac{\ell+1}{2}$ R-vertices and $z \in R$. In this case all black edges have exactly one endpoint from R. Moreover, there is a single gray edge which has both endpoints from R. The analysis is now similar to the previous type. Again, let $e_0, e_1, e_2, \ldots, e_{\frac{\ell-1}{2}}$ be the sequence of black edges in C, starting from $e_0 = \{x, y\}, e_1 = \{y, z\}$, and

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

going in cyclic order. For $1 \le i \le \frac{\ell-1}{2}$ write $e_i = \{u_i, w_i\}$ where $u_i \in R, w_i \in R^c$, and also let $x = u_0$. There exists $1 \le j \le \frac{\ell-1}{2}$ such that the gray edge connecting e_j with e_{j+1} has both endpoints from R (here we let $e_{\frac{\ell+1}{2}} = e_0$). We get the following cycle:

$$\begin{aligned} &u_0 \sim w_1 \sim u_1 \sim w_2 \sim u_2 \sim \ldots \sim w_j \sim u_j \\ &\sim u_{j+1} \sim w_{j+1} \sim \ldots \sim u_{\frac{\ell-1}{2}} \sim w_{\frac{\ell-1}{2}} \sim u_0. \end{aligned} \tag{12}$$

So the number of possible cycles of type 3 is at most $\ell \cdot (r)_{\frac{\ell-1}{2}} \cdot 2^{\frac{\ell-1}{2}}$. As for $\mathbb{P}(C \subseteq \mathbf{H})$, it is now bounded by $\frac{2p_0}{(n)_{\frac{\ell-1}{2}}}$; the extra factor of 2 comes from the gray edge $\{u_j, u_{j+1}\}$, which could come from $u_{j+1} = y[u_j]$ or from $u_j = y[u_{j+1}]$. In conclusion, the expected number of b-cycles of type 3 is at most

$$2p_0\ell \cdot 2^{\frac{\ell-1}{2}} \frac{(r)_{\frac{\ell-1}{2}}}{(n)_{\frac{\ell-1}{2}}} \le 2p_0\ell \cdot \left(\frac{2r}{n}\right)^{\frac{\ell-1}{2}}.$$

In summary, we have proved the following bound for every $\ell \ge 3$:

$$\mathbb{E}(\mathbf{X}_\ell) = O\left(\ell \left(\frac{2r}{n}\right)^{\frac{\ell-3}{2}}\right). \tag{13}$$

Assume $r = \Theta(n)$. Then the expected number of alternating cycles of length $\ge \ln n$ is

$$\sum_{\ell \ge \ln n} \mathbb{E}(\mathbf{X}_\ell) = O\left(\sum_{\ell \ge \ln n} \ell \left(\frac{2r}{n}\right)^{\frac{\ell-3}{2}}\right) = O\left(\sum_{\ell \ge \ln n} \ell (2\xi)^{\frac{\ell-3}{2}}\right) = o(1)$$

since $2\xi < 1$. Then we are done by Markov's inequality.

Now assume $r = o(n)$. Then b-cycles of type 2 or 3 are impossible, because there is no black path of length 2 with both endpoints in R. In that case we can improve (13) and write $\mathbb{E}(\mathbf{X}_\ell) = O\left(\ell \left(\frac{2r}{n}\right)^{\frac{\ell-1}{2}}\right)$. Then the expected number of alternating cycles is

$$O\left(\sum_{\ell \ge 3} \ell \left(\frac{2r}{n}\right)^{\frac{\ell-1}{2}}\right) = o(1).$$

Again, we apply Markov's inequality and complete the proof in the case when r is bounded away from $n/2$.

Finally, assume that $r < n/2$ and $r = (\frac{1}{2} - o(1))n$. We apply the bound (13), noting that its proof does not use the fact that r is bounded away from $\frac{n}{2}$. The bound can now be written more neatly as

$$\mathbb{E}(\mathbf{X}_\ell) = O\left(\ell \left(\frac{2r}{n}\right)^{\frac{\ell}{2}}\right). \tag{14}$$

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

Let us first prove that w.h.p. there are no alternating cycles of size at least $2n/7$. It immediately follows from the union bound and (14). Note that the power of $(2r/n)$ in this bound is, in particular, due to the inequality $(r)_{\lfloor \ell/2 \rfloor}/(n)_{\lfloor \ell/2 \rfloor} \leq (r/n)^{\lfloor \ell/2 \rfloor}$. When ℓ is linear n a much stronger inequality holds true:

$$\begin{aligned}\frac{(r)_{\lfloor \ell/2 \rfloor}}{(n)_{\lfloor \ell/2 \rfloor}} &= \frac{r!\,(n-\lfloor \ell/2 \rfloor)!}{n!\,(r-\lfloor \ell/2 \rfloor)!} \lesssim \frac{r^r (n-\lfloor \ell/2 \rfloor)^{n-\lfloor \ell/2 \rfloor}}{n^n (r-\lfloor \ell/2 \rfloor)^{r-\lfloor \ell/2 \rfloor}} \\ &= \left(\frac{r}{n}\right)^{\lfloor \ell/2 \rfloor} \left(1+\frac{\lfloor \ell/2 \rfloor}{r-\lfloor \ell/2 \rfloor}\right)^{r-\lfloor \ell/2 \rfloor} \left(1-\frac{s}{n-s}\right)^{n-\lfloor \ell/2 \rfloor} \\ &= \left(\frac{r}{n}\left[\left(\frac{1}{1-2x}\right)^{(1-2x)/(2x)} (1-x)^{(1-x)/x}\right]\right)^{\lfloor \ell/2 \rfloor},\end{aligned}$$

where $x = \lfloor \ell/2 \rfloor / n \in [1/7, 1/2)$. Note that the function of x in the above bound decreases with x, and so it is at most $(7/5)^{5/2}(6/7)^6 := \alpha < 1$. Eventually, we get the following refined bound:

$$\mathbb{E}\,(\mathbf{X}_\ell) = O\left(\ell \left(\alpha \frac{2r}{n}\right)^{\frac{\ell}{2}}\right),$$

and then Markov's inequality and the union bound over $\ell \geq \frac{2}{7}n$ implies that indeed w.h.p. there are no alternating cycles of length at least $2n/7$. This implies that w.h.p. any union of two alternating cycles has at most $\frac{3}{7}n + O(1)$ vertices from R.

Assume that different alternating $C = (x_1, \dots, x_\ell)$ and $C' = (x'_1, \dots, x'_\ell)$ are equipotent in $\mathbf{H}$. Let us assume that, for a certain $i \in [\ell]$, $x_i = x'_i \in R$ and, at the same time, $x_{i+2} = x'_{i+2} \in R$ as well. Due to the description of types of alternating cycles, it may only happen when x_i and x_{i+2} are joined in C by the path $x_i x_{i+1} x_{i+2}$ with edges of different color, and the same is true in C'. Indeed, otherwise, since colors in equipotent cycles are swapped, without loss of generality $x_i x_{i+1} x_{i+2}$ is entirely gray, but then might have two consecutive vertices from R, and the third vertex should not belong to R. Moreover, both x_i and x_{i+2} should be incident to at least one gray edge, both in C and C'. Since colors of $\{x_i, x_{i+1}\}$ and $\{x'_i, x'_{i+1}\}$ are different, $x_{i+1} \neq x'_{i+1}$. This may only happen when either x_i or x_{i+2} belongs to two gray edges in $C \cup C'$. But then one of these two gray edges joins two vertices from R. There can be only constantly many such i due to the description of types of alternating cycles.

Then there exists a set $\mathcal{J} \subset [\ell]$ of size at least $\frac{1-o(1)}{4}\ell$ such that, for every $i \in \mathcal{J}$, $x_i \in R$, and $x_i \neq x'_i$. From this it immediately follows that there exists $\tilde{\mathcal{J}} \subset \mathcal{J}$ of size at least $\frac{1-o(1)}{8}\ell$ such that sets $\mathcal{R} := \{x_i, i \in \tilde{\mathcal{J}}\}$ and $\mathcal{R}' := \{x'_i, i \in \tilde{\mathcal{J}}\}$ are disjoint and $\mathcal{R} \subset R$.

Let us now finish the proof. Consider two (not necessarily disjoint) sets $\{x_1, \dots, x_\ell\}$ and $\{x'_1, \dots, x'_\ell\}$. Assume that the event saying that $C = (x_1, \dots, x_\ell)$ and $C' = (x'_1, \dots, x'_\ell)$ are alternating cycles in $\mathbf{H}$ holds. We may assume that at least $(1 - o(1))\frac{n}{14}$ vertices of R do not belong to $C \cup C'$. Denote the set of these vertices of R that do not belong to any of the cycles by R'. Almost all (but constantly many) vertices of R that belong to the union of the cycles still can play the role of $y[x]$ for $x \in R'$ (unless they do not belong to gray edges with both endpoints from R). We let $\tilde{R}$

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

be the set of such vertices from $R \cap (V(C) \cup V(C'))$. Moreover, let $\tilde{\mathcal{R}} = \tilde{R} \cap \mathcal{R}$, $\tilde{\mathcal{R}}' = \tilde{R} \cap \mathcal{R}'$. Recall that $|\tilde{\mathcal{R}}| = \frac{1-o(1)}{8}\ell$.

Note that the set $\mathcal{Y}$ of $y[x]$, $x \in R'$, is a uniformly random subset of $[n] \setminus (V(C) \cup V(C')) \cup \tilde{R}$ of size $|R'|$. Let X be the number of vertices in $\tilde{\mathcal{R}} \cap \mathcal{Y}$. Due to the Hoeffding tail bounds for the hypergeometric distribution (see, e.g., [11, Theorem 2.10]), X is at least $\frac{1-o(1)}{112}\ell$ with probability $1 - \exp(-\Omega(\ell))$. Let φ be the bijection from $\mathcal{R}$ to $\mathcal{R}'$ that sends x_i to x'_i. If, for a certain $x \in \mathcal{Y} \cap \tilde{\mathcal{R}}$, $\varphi(x) \notin \tilde{\mathcal{R}}'$, then cycles C and C' cannot be equipotent. Thus, we get that the event that C and C' are equipotent implies that $\varphi|_{\mathcal{Y} \cap \tilde{\mathcal{R}}}$ is a bijection between $\mathcal{Y} \cap \tilde{\mathcal{R}}$ and $\mathcal{Y} \cap \tilde{\mathcal{R}}'$. Since the probability that $\mathcal{Y} \cap \tilde{\mathcal{R}}'$ coincides with a fixed subset of $\tilde{\mathcal{R}}'$ is $\exp(-\Omega(\ell))$, we get the statement of Proposition 6.2 due to the bound (14) and the union bound over the choices of ℓ, x_1, x'_1, and the directions in both cycles C, C'.

In the case $r = o(n)$, Proposition 6.2 rules out the existence of alternating cycles, which is already sufficient for a proof that $\mathbf{H}$ is asymmetric w.h.p. (as would be clear from the rest of the proof). In the case $r = \Theta(n)$, however, we must consider more specific configurations. Thus, for the rest of this section, let us assume that $r \geq cn$ where c is a positive constant.

DEFINITION 6.5. For two non-empty sets $U_1, U_2 \subseteq [n]$, the *black distance* between them is the minimal length of a black path connecting a vertex from U_1 to a vertex from U_2.

DEFINITION 6.6. For a set $U \subseteq [n]$, $|U| \geq 2$, a *black segment* of U of length t is a black path with endpoints $x_0, x_t \in U$ and internal vertices $x_1, \ldots, x_{t-1} \notin U$.

PROPOSITION 6.7. *W.h.p. in* $\mathbf{H}$, *there are no alternating cycles of length at most* $\ln^2 n$ *satisfying at least one of the following properties*:

1. *with an additional alternating path of length between* 2 *and* $\ln^2 n$ *connecting two* (*not necessarily distinct*) *vertices of the cycle*;
2. *with additional internal edges*;
3. *with another alternating cycle of length at most* $\ln^2 n$ *at black distance less than* $\sqrt{n}$;
4. *with a black segment of length between* 2 *and* $\sqrt{n}$.

We will need the following claim.

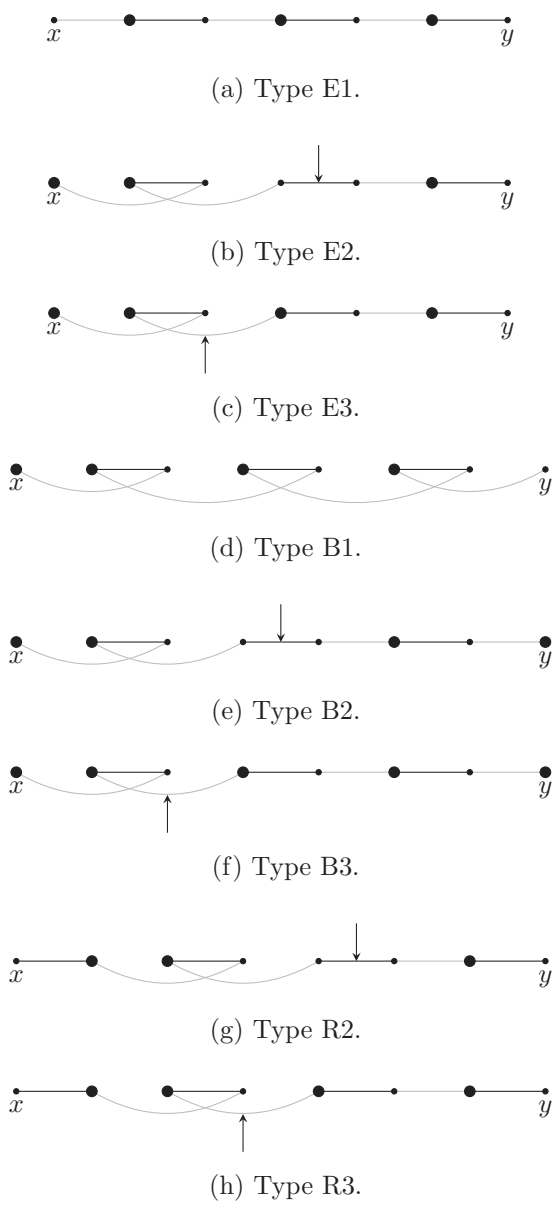
CLAIM 6.8. *Fix two distinct vertices* x, y *and let* $t \geq 2$. *Let* $\mathbf{X}_t[x, y]$ *be the number of alternating paths of length* t *which have* x, y *as the two endpoints. Then*

$$\mathbb{E}(\mathbf{X}_t[x, y]) = O\left(\frac{1}{n}\right) \cdot t \left(\frac{2r}{n}\right)^{\frac{t}{2}}.$$

PROOF. The proof involves case analysis very similar to that of the previous proof. We avoid a repetition of the same details, and instead only briefly describe the different possible types of alternating paths (see also Figure 3).

Again, we distinguish between three cases: even paths, odd g-paths, and odd b-paths. Each case can be subdivided into types analogous to the types of b-cycles from the previous proof.

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press



(a) Type E1.

(b) Type E2.

(c) Type E3.

(d) Type B1.

(e) Type B2.

(f) Type B3.

(g) Type R2.

(h) Type R3.

FIGURE 3. All possible types of alternating paths connecting a given pair of vertices x, y. Vertices of R are marked by big black circles (excluding the endpoints in the black cases, as described in the proof). An arrow marks the "special edge," which is again either a black edge with no R-endpoints or a gray edge with two R-endpoints.

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

Even path P consists of $\frac{t}{2}$ black edges and $\frac{t}{2}$ gray edges. Without loss of generality x is incident to a gray edge.
Type E1. $x \notin R$ and $|V(P) \cap R| = \frac{t}{2}$.
Type E2. $x \in R$ and $|V(P) \cap R| = \frac{t}{2}$.
Type E3. $x \in R$ and $|V(P) \cap R| = \frac{t}{2} + 1$.

G-path P consists of $\frac{t-1}{2}$ black edges and $\frac{t+1}{2}$ gray edges. At least one endpoint of P must be in R, due to Remark 6.3. Without loss of generality assume $x \in R$.
Type B1. $y \notin R$ and $|V(P) \cap R| = \frac{t+1}{2}$.
Type B2. $y \in R$ and $|V(P) \cap R| = \frac{t+1}{2}$.
Type B3. $y \in R$ and $|V(P) \cap R| = \frac{t+3}{2}$.

B-path P consists of $\frac{t+1}{2}$ black edges and $\frac{t-1}{2}$ gray edges. In this case we exclude the endpoints of P from the count of R-vertices, since they are not incident to any gray edge in the path.
Type R2. $\left|\left(V(P) \cap R\right) \setminus \{x, y\}\right| = \frac{t-1}{2}$.
Type R3. $\left|\left(V(P) \cap R\right) \setminus \{x, y\}\right| = \frac{t+1}{2}$.

For all types, the expected number of paths is determined (up to a $\Theta(1)$-factor) by combining a factor of $2r$ for every *internal* black edge, a factor of $\frac{1}{n}$ for every gray edge, and an additional factor of t in types numbered 2 and 3 (coming from the choice of one "special edge," indicated by an arrow in Figure 3). ⊣

Proof of Proposition 6.7.1. Let $\mathbf{X}_{\ell,t}$ be the number of pairs (C, P), where C is an alternating cycle of length $\ell \leq \ln^2 n$ and P is an alternating path of length $t \leq \ln^2 n$ which shares its two endpoints with C (internally disjoint from C). A pair (C, P) can be chosen by first choosing an alternating cycle C, then two vertices x, y from C, and then an alternating path connecting x, y. From (14) and Claim 6.8, we get

$$\mathbb{E}\left(\mathbf{X}_{\ell,t}\right) = O\left(\ell \left(\frac{2r}{n}\right)^{\frac{\ell}{2}} \ell^2 \frac{1}{n} t \left(\frac{2r}{n}\right)^{\frac{t}{2}}\right) = \frac{1}{n} O\left(\ell^3 \left(\frac{2r}{n}\right)^{\frac{\ell}{2}} t \left(\frac{2r}{n}\right)^{\frac{t}{2}}\right).$$

Summing over $3 \leq \ell \leq \ln^2 n$ and $t \leq \ln^2 n$, and using the inequality $\frac{2r}{n} < 1$, we deduce $\sum_{\ell,t} \mathbb{E}\left(\mathbf{X}_{\ell,t}\right) = o(1)$. We then apply Markov's inequality and finish the proof. ⊣

Proof of Proposition 6.7.2. For every $4 \leq \ell \leq \ln^2 n$, let $\mathbf{X}_\ell^{[r]}$ be the number of pairs (C, e) where C is an alternating cycle of length ℓ and e is an additional internal black edge. Similarly let $\mathbf{X}_\ell^{[b]}$ be the number of (C, e) where now e is an additional internal gray edge. The idea is that an additional black edge reduces the number of possible choices by a factor of ℓ/r, and an additional gray edge reduces the probability of the existence in $\mathbf{H}$ by a factor of ℓ/n.

We begin with $\mathbf{X}_\ell^{[b]}$. First notice that an additional gray edge in C is possible only when ℓ is odd and C is a b-cycle of type 3 (as defined in the proof of Proposition 6.2). Indeed, in all other cases, the number of gray edges in C equals the number of R-vertices in C, and therefore no additional gray edges may exist. The general form of a b-cycle of type 3 is described in (12). Using that notation, the additional gray edge e must have either u_j or u_{j+1} as one of its endpoints. The number of possible choices

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

for (C, e) is therefore at most $2\ell \cdot \ell \cdot (r)_{\frac{\ell-1}{2}} \cdot 2^{\frac{\ell-1}{2}}$ (an extra factor of 2ℓ is added due to e). However, the probability that (C, e) exists in $\mathbf{H}$ is now $\frac{1}{(n)_{\frac{\ell+1}{2}}}$. Overall,

$$\mathbb{E}\left(\mathbf{X}_\ell^{[b]}\right) = O\left(\ell^2 \cdot 2^{\frac{\ell-1}{2}} \frac{(r)_{\frac{\ell-1}{2}}}{(n)_{\frac{\ell+1}{2}}}\right) = O\left(\frac{\ell^2}{n} \cdot \left(\frac{2r}{n}\right)^{\frac{\ell}{2}}\right).$$

Now consider $\mathbf{X}_\ell^{[r]}$. We focus on the case where ℓ is even; the odd case is handled similarly and gives an additional factor of ℓ.

In any pair (C, e) the cycle can be labelled as in C as in (9) in a way such that e has u_1 as an endpoint. There are at most ℓ ways to choose the other endpoint of e. Moreover, e determines a black path of length 3 which contains two black edges of C, and reduces the number of choices by a factor of r. Therefore, a bound on $\mathbb{E}\left(\mathbf{X}_\ell^{[r]}\right)$ can be obtained by multiplying (10) by $\frac{\ell}{r}$, which yields

$$\mathbb{E}\left(\mathbf{X}_\ell^{[r]}\right) = O\left(\frac{\ell}{r} \cdot \left(\frac{2r}{n}\right)^{\frac{\ell}{2}}\right) = O\left(\frac{\ell}{n} \cdot \left(\frac{2r}{n}\right)^{\frac{\ell}{2}}\right).$$

Recalling that $\frac{2r}{n} < 1$ and combining the above,

$$\sum_{4 \le \ell \le \ln^2 n} \mathbb{E}\left(\mathbf{X}_\ell^{[b]} + \mathbf{X}_\ell^{[r]}\right) = O\left(\frac{1}{n}\right) \sum \ell^2 \left(\frac{2r}{n}\right)^{\frac{\ell}{2}} = O\left(\frac{\ln^5 n}{n}\right).$$

We apply Markov's inequality and finish the proof. ⊣

Proof of Proposition 6.7.3. We have already proved that there are no two overlapping alternating cycles w.h.p. Fix $3 \le \ell, \ell' \le \ln^2 n$, and $t \le \sqrt{n}$. Let $\mathbf{X}_{\ell,\ell',t}$ be the number of triplets (C, C', P) where C, C' are disjoint alternating cycles of lengths ℓ, ℓ', respectively, and P is a black path connecting them. The idea is that the black path P eliminates one degree of freedom in the choice of cycles, therefore reduces the number of choices by a factor of r. Let us handle the case when ℓ, ℓ' are both even; the odd case is handled similarly, and gives an additional factor of $\ell\ell'$.

Note that P together with two black edges from C and C' produce a black path P' of length $t + 2$, with two vertices from $R \cap (V(C) \cup V(C'))$ in it. There are $O(r)$ ways to choose the path P', and this choice determines a black edge of C and a black edge of C'. The number of choices of C and of C' are then reduced by a factor of r each. Combining these factors with (10), we obtain the bound

$$\mathbb{E}\left(\mathbf{X}_{\ell,\ell',t}\right) = O\left(\frac{1}{n} \cdot \left(\frac{2r}{n}\right)^{\frac{\ell}{2}+\frac{\ell'}{2}}\right).$$

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

Summing over ℓ, ℓ', t, and adding a factor of $\ell\ell'$ for the odd case, we have

$$\sum_{\ell,\ell' \le \ln^2 n,\ t \le \sqrt{n}} \mathbb{E}\left(\mathbf{X}_{\ell,\ell',t}\right) = \sqrt{n} \cdot O\left(\frac{1}{n}\right) \cdot \sum_{\ell,\ell' \le \ln^2 n} \ell\ell' (2\xi)^{\frac{\ell}{2}+\frac{\ell'}{2}}$$
$$= O\left(\frac{\ln^6 n}{\sqrt{n}}\right) = o(1).$$

⊣

Proof of Proposition 6.7.4. Fix $\ell \le \ln^2 n$ and $2 \le t \le \sqrt{n}$. Let $\mathbf{X}_{\ell,t}$ count pairs (C, P) where C is an alternating cycle of length ℓ and P is a black segment of C of length t. As above, we only consider the case when ℓ is even; the odd case is handled similarly, and gives an additional factor of ℓ.

Note that the endpoints of P are not adjacent by a black edge—otherwise this edge would form with P the entire Hamilton cycle. We apply the usual argument: P together with the two black edges from C produce a black path P' of length $t+2$ with two vertices from $R \cap V(C)$ in it. There are $O(r)$ ways to choose the path P', and this choice determines two black edges of C. There are $O(\ell)$ ways to choose their relative position in C. The number of choices of C is therefore multiplied by a factor of $\frac{\ell}{r^2}$. Combining with (10), we obtain the bound

$$\mathbb{E}\left(\mathbf{X}_{\ell,t}\right) = O\left(r \cdot \frac{\ell}{r^2}\left(\frac{2r}{n}\right)^{\frac{\ell}{2}}\right) = O\left(\frac{\ell}{n} \cdot \left(\frac{2r}{n}\right)^{\frac{\ell}{2}}\right).$$

Summing over ℓ, t, and adding a factor of $\ell\ell'$ for the odd case, we have

$$\sum_{3 \le \ell \le \ln^2 n,\ 2 \le t \le \sqrt{n}} \mathbb{E}\left(\mathbf{X}_{\ell,t}\right) = \sqrt{n} \cdot O\left(\frac{1}{n}\right) \cdot \sum_{3 \le \ell \le \ln^2 n} \ell^2 (2\xi)^{\frac{\ell}{2}}$$
$$= O\left(\frac{\ln^5 n}{\sqrt{n}}\right) = o(1).$$

⊣

Finally, we will need the following property.

Proposition 6.9. *W.h.p. in* $\mathbf{H}$, *there exist no two disjoint black paths* $P = (x_1, x_2, \ldots, x_t)$ *and* $P' = (x'_1, x'_2, \ldots, x'_t)$ *of length* $t \ge \sqrt{n}$ *such that* $\deg_{\mathbf{H}}(x_i) = \deg_{\mathbf{H}}(x'_i)$ *for every* $1 \le i \le t$.

Proof. We prove the proposition in the random multigraph $\mathbf{H}^*$. Then, from Proposition 2.10 it is also true in $\mathbf{H}$. Moreover, it is sufficient to prove the proposition only for $t = \sqrt{n}$ due to monotonicity.

The number of disjoint labelled black paths $P = (x_1, x_2, \ldots, x_t)$ and $P' = (x'_1, x'_2, \ldots, x'_t)$ is at most $(2n)^2 = 4n^2$. Let us fix such a pair (P, P') and bound $\mathbb{P}(\mathcal{D})$, where

$$\mathcal{D} = \mathcal{D}_{P,P'} = \{\deg(x_i) = \deg(x'_i) \quad \forall 1 \le i \le t\}.$$

Note that $\deg(x)$ refers to the degree in $\mathbf{H}^*$.

We begin by defining some notation. Let $\mathbf{Y} = \{y[x] : x \in R\}$. In $\mathbf{H}^*$, by definition, $\mathbf{Y}$ is distributed uniformly over all subsets of $[n]$ of size r. For every $v \in [n]$, $\deg(v) = 2 + \mathbb{1}_{v \in \mathbf{Y}} + \mathbb{1}_{v \in R}$. Let $\mathbf{z} = \left(\mathbb{1}_{x_1 \in \mathbf{Y}}, \ldots, \mathbb{1}_{x_t \in \mathbf{Y}}\right)$ be a random binary vector, whose value

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

precisely determines the intersection of P with $\mathbf{Y}$. We bound $\mathbb{P}(\mathcal{D})$ by conditioning by the value of $\mathbf{z}$.

So, let $z = \left(I_{x_1}, I_{x_2}, \dots, I_{x_t}\right)$ be a binary vector and consider $\mathbb{P}\left(\mathcal{D} \mid \mathbf{z} = z\right)$. Let $j = \sum_{i=1}^{t} I_{x_i}$ (the number of 1-entries in z). Then, given $\mathbf{z} = z$, the subset $\tilde{\mathbf{Y}} = \mathbf{Y} \setminus P$ is distributed uniformly over all subsets of $[n] \setminus P$ of size $r - j$. The event $\mathcal{D}$ is saying that $\mathbb{1}_{x'_i \in \mathbf{Y}} = I_{x_i} + \mathbb{1}_{x_i \in R} - \mathbb{1}_{x'_i \in R}$ for every $1 \le i \le t$. If there exists i such that $\mathbb{1}_{x'_i \in \mathbf{Y}} = I_{x_i} + \mathbb{1}_{x_i \in R} - \mathbb{1}_{x'_i \in R} \notin \{0, 1\}$, then of course $\mathbb{P}\left(\mathcal{D} \mid \mathbf{z} = z\right) = 0$. Otherwise, we can partition the vertices of P' into

$$B = \left\{x'_i : I_{x_i} + \mathbb{1}_{x_i \in R} - \mathbb{1}_{x'_i \in R} = 1\right\},$$
$$C = \left\{x'_i : I_{x_i} + \mathbb{1}_{x_i \in R} - \mathbb{1}_{x'_i \in R} = 0\right\}.$$

Then $\mathcal{D}$ is the event that $B \subseteq \tilde{\mathbf{Y}}$ and $\tilde{\mathbf{Y}} \cap C = \emptyset$:

$$\mathbb{P}(\mathcal{D} \mid \mathbf{z} = z) = \frac{\binom{n-2t}{r-j-t}}{\binom{n-t}{r-j}} \le \frac{\binom{n-2t}{r-j}}{\binom{n-t}{r-j}} = \frac{(n-2t)_{r-j}}{(n-t)_{r-j}} \le \frac{(n-t)_{r-j}}{(n)_{r-j}}$$
$$= O\left(\prod_{i=0}^{r-1}\left(1 - \frac{t}{n-i}\right)\right) = O\left(\exp\left(-\sum_{i=0}^{r-1}\frac{t}{n-i}\right)\right).$$

Now, since $r \ge cn$ for a positive c,

$$\sum_{i=0}^{r-1}\frac{1}{n-i} = \frac{1}{n}\sum_{i=0}^{r-1}\frac{1}{1-\frac{i}{n}} \ge (1 + o(1))\int_0^c \frac{\mathrm{d}x}{1-x}$$
$$= (1 + o(1))\ln\left(\frac{1}{1-c}\right).$$

We deduce

$$\mathbb{P}(\mathcal{D} \mid \mathbf{z} = z) \le \exp\left(-\Theta(t)\right) = \exp\left(-\Theta\left(\sqrt{n}\right)\right).$$

This bound is true for every value z (uniformly), therefore it also holds for $\mathbb{P}(\mathcal{D})$. Now, taking a union bound over the pairs (P, P'), we have

$$\mathbb{P}\left(\bigcup_{(P,P')} \mathcal{D}_{P,P'}\right) \le 4n^2 \cdot \exp\left(-\Theta\left(\sqrt{n}\right)\right) = o(1)$$

and that finishes the proof. ⊣

6.2. Proof of asymmetry. Let Σ_n be the group of permutations of $[n]$. We need to prove that w.h.p., for every $\sigma \in \Sigma_n \setminus \{\mathrm{id}\}$, σ is not an automorphism of $\mathbf{H}$.

Let $\Sigma_n^{[h]} \subset \Sigma_n$ be the subgroup of $2n$ automorphisms of $H_h = (1, 2, \dots, n, 1)$. Note that these are exactly the permutations which preserve edge colors. The first part of the argument here is deterministic: we prove that if $\sigma \in \Sigma_n \setminus \Sigma_n^{[h]}$ is an automorphism of $\mathbf{H}$, then $\mathbf{H}$ has a rare configuration, excluded by one of the propositions from the previous section.

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

Suppose there exists $\sigma \in \Sigma_n \setminus \Sigma_n^{[h]}$ which is an automorphism of $\mathbf{H}$. Then there exists a black edge $\{x_1, x_2\}$ such that $\sigma(\{x_1, x_2\})$ is gray. Consider the two black edges incident to $\sigma(x_2)$; write them as $\{\sigma(x_2), \sigma(x_3)\}$ and $\{\sigma(x_2), \sigma(x_3')\}$. Since x_2 is also incident to exactly two black edges, and $\{x_1, x_2\}$ is black, one of $\{x_2, x_3\}, \{x_2, x_3'\}$ must be gray. Assuming that $\{x_2, x_3\}$ is gray, and applying the same argument but for σ^{-1} applied to $\{\sigma(x_2), \sigma(x_3)\}$, we would get that there exists a black edge $\{x_3, x_4\}$ such that $\{\sigma(x_3), \sigma(x_4)\}$ is gray. Proceeding by induction, we obtain an alternating walk such that its image under σ is also an alternating walk, with colors switched. Since the graph is finite, at some point the walk meets itself, and we would get an alternating cycle C such that $\sigma(C)$ is also an alternating cycle, with edge colors switched by σ. If $r = o(n)$, we are done: from Proposition 6.2 we know that w.h.p. $\mathbf{H}$ contains no alternating cycles. So assume $r \geq cn$ for a positive constant c. Since labeled C and $\sigma(C)$ are different, due to Proposition 6.2, we may assume that C has length at most $\ln^2 n$.

Let u_1 be a vertex in C, incident to a black edge and a gray edge in C, such that $\sigma(u_1) \neq u_1$. It is obvious that such a vertex exists when C is even. If C is odd, then it has a unique vertex adjacent to two edges of the same color. If all other vertices are fixed, then there exists an edge in C that maps to itself, and so its image has the same color—a contradiction.

Let $\{u_1, v\}$ be the black edge incident to u_1 in C. Let $\{u_1, u_2\}$ be the second black edge incident to u_1. From Proposition 6.7, we may assume that $u_2 \notin C$, $\{\sigma(u_1), \sigma(u_2)\}$ is black, and all four vertices $u_1, u_2, \sigma(u_1), \sigma(u_2)$ must be distinct. We can now continue inductively, constructing two disjoint black paths $(u_1, u_2, u_3, \dots)$ and $(\sigma(u_1), \sigma(u_2), \sigma(u_3), \dots)$. Since Proposition 6.7 forbids any black distances or black segments of length at most $\sqrt{n}$, in particular we can proceed until the black paths are of length $t \geq \sqrt{n}$. We get two directed disjoint black paths of length at least $\sqrt{n}$ with the same sequence of degrees, so Proposition 6.9 applies. We see that the assumption that $\mathbf{H}$ is σ-symmetric leads to the existence of one of the rare configurations from the previous section in $\mathbf{H}$, which happens with probability $o(1)$.

Finally, it remains to show that w.h.p., no permutation $\sigma \in \Sigma_n^{[h]} \setminus \{\mathrm{id}\}$ is an automorphism of $\mathbf{H}$. Recall that these are the permutations which preserve edge colors. This time we separately consider the cases $r \geq 30 \ln n$ and $r \leq 30 \ln n$.

Start with $r \geq 30 \ln n$. We show that for a given $\sigma \in \Sigma_n^{[h]} \setminus \{\mathrm{id}\}$, the probability that it is an automorphism of $\mathbf{H}$ is $o(1/n)$. Then, it would only remain to take a union bound over $2n - 1$ possible σ.

We construct a set of vertices S (depending only on σ) with $|S| = \lfloor \frac{n}{3} \rfloor$ and $|\sigma(S) \cap R| \geq \lfloor \frac{r}{3} \rfloor$ and $S \cap \sigma(S) = \emptyset$. The construction is done with the following algorithm:

1. Start with $S = \emptyset$.
2. Choose a vertex $x \in \sigma^{-1}(R)$ such that $\sigma(x) \neq x$ and $x \notin S \cup \sigma(S) \cup \sigma^{-1}(S)$. Update $S \leftarrow S \cup \{x\}$. Repeat this step $\lfloor \frac{r}{3} \rfloor$ times.
3. Choose a vertex $x \in [n]$ such that $\sigma(x) \neq x$ and $x \notin S \cup \sigma(S) \cup \sigma^{-1}(S)$. Update $S \leftarrow S \cup \{x\}$. Repeat this step $\lfloor \frac{n}{3} \rfloor - \lfloor \frac{r}{3} \rfloor$ times.

Since σ is either a rotation or a reflection of the Hamiltonian cycle H_h, it has at most 2 fixed points. Therefore, at the i-th step there are at most $2 + 3(i - 1) = 3i - 1$ forbidden vertices. This shows that indeed we can repeat the second step

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

$\lfloor \frac{r}{3} \rfloor$ times, and then repeat the third step $\lfloor \frac{n}{3} \rfloor - \lfloor \frac{r}{3} \rfloor$ times. By construction $S \cap \sigma(S) = \emptyset$, $|\sigma(S) \cap R| \geq \lfloor \frac{r}{3} \rfloor \geq \lfloor 10 \ln n \rfloor$ and $|S| = \lfloor \frac{n}{3} \rfloor$. Also let $T \subseteq S$ such that $\sigma(T) \subseteq \sigma(S) \cap R$ and $|T| = \lfloor 10 \ln n \rfloor$.

Let $\mathbf{B}[T, S]$ denote the set of gray edges with an endpoint in T and the other endpoint in S. If σ is an automorphism of $\mathbf{H}$ then the event $\sigma(\mathbf{B}[T, S]) = \mathbf{B}[\sigma(T), \sigma(S)]$ holds. So it remains to show that, for every possible value B of $\mathbf{B}[T, S]$,

$$\mathbb{P}\left(\mathbf{B}[\sigma(T), \sigma(S)] = \sigma(B) \mid \mathbf{B}[T, S] = B\right) = o(1/n)$$

uniformly over B.

Consider the vertices $\{\mathbf{y}[x] : x \in \sigma(T)\}$. Given $\mathbf{B}[T, S] = B$, each $\mathbf{y}[x]$ can still be any vertex from $\sigma(S)$ (since $S \cap \sigma(S) = \emptyset$). However, the event $\mathbf{B}[\sigma(T), \sigma(S)] = \sigma(B)$ restricts $\mathbf{y}[x]$ to at most 2 possible values from $\sigma(S)$ (since there are at most two gray edges between x and S). Recall that $|S| = \lfloor \frac{n}{3} \rfloor$ and that $|T| = \lfloor 10 \ln n \rfloor$, so given $\mathbf{B}[T, S] = B$, the event $\mathbf{B}[\sigma(T), \sigma(S)] = \sigma(B)$ restricts the number of values of each $\mathbf{y}[x]$ by a factor of $\frac{2}{3} + o(1)$. We therefore get the following coarse bound:

$$\mathbb{P}\left(\mathbf{B}[\sigma(S)] = \sigma(B) \mid \mathbf{B}[S] = B\right) \leq \left(2/3 + o(1)\right)^{|T|} = o(1/n).$$

In the case $r \leq 30 \ln n$ we use a different argument. Fix $\sigma \in \Sigma_n^{[h]} \setminus \{\mathrm{id}\}$. If σ is an automorphism of $\mathbf{H}$, one of the following events must hold:

1. $E_\sigma^{(1)}$, the event that $\sigma(e) = e$ for every gray edge e.
2. $E_\sigma^{(2)}$, the event that there exist two gray edges $e_1 \neq e_2$ such that $\sigma(e_1) = e_2$.

The first event fully determines the values $\mathbf{y}[x]$ for every $x \in R$ (necessarily $\mathbf{y}[x] = \sigma(x)$, except when σ is a reflection and there are two fixed points x_1, x_2, in which case $\mathbf{y}[x_1] = x_2$ and $\mathbf{y}[x_2] = x_1$). Therefore $\mathbb{P}\left(E_\sigma^{(1)}\right) \leq p_0/(n)_r$.

The second event implies the following event E (which is not dependent on σ). For every $x \in R$ let $\mathbf{d}(x)$ measure the black distance between x and $\mathbf{y}[x]$. Let E be the event that there exist two distinct vertices $x_1, x_2 \in R$ such that $\mathbf{d}(x_1) = \mathbf{d}(x_2)$. It is easy to see that $\mathbb{P}(E) = \binom{r}{2} \cdot O(1/n)$. Overall, by the union bound, the probability that $\mathbf{H}$ is σ-symmetric in this case is $o(1)$ since $r \geq 2$ and $r = O(\ln n)$. That finishes the proof of Theorem 2.15 for the case $m < \frac{3}{2}n$.

§7. Final remarks. In this article, we study distinguishability of sparse random graphs by FO sentences. We proved a general upper bound on the minimum quantifier depth of a sentence that distinguishes between independent samples of $G(n, n^{-\alpha})$ that answers the question from [4], developed a new tool—random balanced graph—and studied its properties. We observed a new phenomenon: the FO distinguishability between independent samples of $G(n, n^{-\alpha})$ depends on how well α can be approximated by rational numbers. In particular, from Theorem 1.4 it follows that the lower bound $\mathbf{k}_\alpha \geq \frac{1-o(1)}{\ln 2} \ln \ln \ln n$ w.h.p. holds true for almost all $\alpha \in (0, 1)$—in particular, for all algebraic irrational $\alpha \in (0, 1)$.

We believe that the general upper bound on $\mathbf{k}_\alpha$ in Theorem 1.2 can be improved. A possible way to do it is to consider existential sentences that describe existence of a tuple of graphs with a common set of roots defined in such a way that, as

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

soon as, intersections between these graphs become large, the density of their union becomes non-typical. We suspect that a further increase of the width of syntax trees of existential sentences by branching would improve the bound up to $O(\ln \ln n)$, but we do not see a promising way to reach the lower bound in Theorem 1.4.

A more general and involved problem that is very interesting to address is to get a tight dependency between typical values of $\mathbf{k}_\alpha$ and an accuracy of approximations of α by rational numbers.

A. Asymmetry: Dense case In this section we prove Theorem 2.16. From now on, suppose that $\{\mathbf{G}_n\}_{n=1}^{\infty}$ is a sequence of random graphs with the following three properties:

(P1) $\delta(\mathbf{G}_n) \geq 3$.
(P2) There exists a constant $\Delta \geq 3$ such that $\Delta(\mathbf{G}_n) \leq \Delta$ for every n.
(P3) There exists a constant $c_0 > 0$ such that $\mathbb{P}\left(E_0 \subseteq E\left(\mathbf{G}_n\right)\right) \leq \left(\frac{c_0}{n}\right)^{|E_0|}$ for every $E_0 \subseteq \binom{[n]}{2}$ and every n.

We need to show that $\mathbf{G}_n$ is asymmetric w.h.p.

A.1. Orbits analysis. Let Σ_n denote the group of permutations on $[n]$. For every $\sigma \in \Sigma_n$ let P_σ denote the probability that σ is an automorphism of $\mathbf{G}_n$. Then it suffices to show that

$$\sum_{\sigma \in \Sigma_n \setminus \{\mathrm{id}\}} P_\sigma = o(1). \tag{A.1}$$

To bound P_σ for a given permutation $\sigma \in \Sigma_n$, we shall analyze the orbits of edges under the action of σ.

Definition A.1. Let $E_n = \binom{[n]}{2}$. We define a group action of Σ_n on E_n as follows: for every $\sigma \in \Sigma_n$ and $\{x, y\} \in E_n$,

$$\sigma\left(\{x, y\}\right) = \{\sigma(x), \sigma(y)\}.$$

Given a permutation $\sigma \in \Sigma_n$ and a pair $e \in E_n$, we define $\mathrm{Orb}_\sigma(e)$ as the orbit of e under the action of the subgroup $\langle\sigma\rangle$ generated by σ. That is, the σ-orbit $\mathrm{Orb}_\sigma(e) = \left\{\sigma^k(e) : k \in \mathbb{Z}\right\} \subseteq E_n$.

Note that for every σ, the quotient set is

$$E_n / \langle\sigma\rangle = \{\mathrm{Orb}_\sigma(e) : e \in E_n\}$$

defines a partition of E_n.

The relation between asymmetry of a graph and orbits is given in the following simple claim.

Claim A.2. *σ as an automorphism of a graph $G = ([n], E) \iff E$ does not split σ-orbits (i.e., for every σ-orbit $\mathbf{O}$, either $O \subseteq E$ or $O \cap E = \emptyset$).*

The σ-orbits in the set of edges E_n are closely related to the σ-orbits in the vertex set $[n]$, which are simply sets of vertices of cycles in the cycle decomposition of σ. The following proposition is straightforward.

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

PROPOSITION A.3. *Let $e = \{x_1, x_2\} \in E_n$, $x_1 \in C_1$ and $x_2 \in C_2$ where C_1, C_2 are cycles from the cycle decomposition of σ. Denote $\ell_1 = |C_1|$ and $\ell_2 = |C_2|$.*

- *If $C_1 \neq C_2$ then $\left|\mathrm{Orb}_\sigma(e)\right| = \mathrm{lcm}(\ell_1, \ell_2)$.*
- *Now suppose $C_1 = C_2$ and let $\ell = \ell_1 = \ell_2$. Then $\left|\mathrm{Orb}_\sigma(e)\right| = \ell$, unless ℓ is even and x, y are opposite in the cycle; in that case $\left|\mathrm{Orb}_\sigma(e)\right| = \frac{\ell}{2}$.*

Proposition A.3 motivates us to classify vertices according to their cycle length in σ. As it turns out, to properly bound P_σ for $\sigma \in \Sigma_n \setminus \{\mathrm{id}\}$ we must consider 2-cycles and 3-cycles separately from the longer cycles.

From now on let us fix a permutation $\sigma \in \Sigma_n$; all the following definitions are with respect to the action of $\langle\sigma\rangle$.

DEFINITION A.4. We define a partition $[n] = S_1 \cup S_2 \cup S_3 \cup S_4$ as follows. S_1, S_2, S_3 are the sets of vertices in 1-cycles (fixed points), 2-cycles and 3-cycles of σ, respectively. S_4 is the set of remaining vertices: vertices in ℓ-cycles for $\ell \geq 4$. We also write $S = S_2 \cup S_3 \cup S_4$, $s_i = |S_i|$ for $i \in \{1, 2, 3, 4\}$ and $s = |S|$.

Note that $\sigma \neq \mathrm{id}$ is equivalent to $S \neq \emptyset$.

DEFINITION A.5. We classify pairs $e \in E_n$ as follows. Suppose $e = \{x, y\}$ where $x \in S_i$ and $y \in S_j$ for some $i, j \in \{1, 2, 3, 4\}$.

- If x, y are not opposite vertices in an even cycle, we say that e is a *pair of type (i,j)* (and also of type (j, i)).
- Now suppose that x, y are opposite vertices in an even cycle of length ℓ. Note that then $i = j = \begin{cases} 2 & \ell = 2 \\ 4 & \ell \geq 4 \end{cases}$. In this case we say that e is a *pair of type (i,0)* (and also of type $(0, i)$).

A pair of type (i, j) is also called an *(i,j)-pair*.

REMARK A.6. We make several remarks about the last definition.

- The types (i, j) and (j, i) are identical.
- Two pairs in E_n from the same σ-orbit have the same type. An orbit of an (i, j)-pair is called an *(i,j)-orbit*.
- There are 12 distinct possible types, arranged in the table below.

$$\begin{array}{cccc} & (0,2), & & (0,4), \\ (1,1), & (1,2), & (1,3), & (1,4), \\ & (2,2), & (2,3), & (2,4), \\ & & (3,3), & (3,4), \\ & & & (4,4). \end{array}$$

DEFINITION A.7. Following the last definitions and remarks, we introduce some additional notations.

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

- For every type (i, j) let T_{ij} be the set of (i, j)-pairs. Also denote $t_{ij} = |T_{ij}|$.
- For every type (i, j), let h_{ij} denote the minimum possible size of an (i, j)-orbit. The values of h_{ij} for all distinct possible types are arranged in the table below.

$$\begin{array}{cccc} & h_{02}=1, & & h_{04}=2, \\ h_{11}=1, & h_{12}=2, & h_{13}=3, & h_{14}=4, \\ & h_{22}=2, & h_{23}=6, & h_{24}=4, \\ & & h_{33}=3, & h_{34}=6, \\ & & & h_{44}=4. \end{array}$$

A.2. Lists of parameters. We shall bound P_σ by taking into account the number of edges of each type in $\mathbf{G}_n$; we call these numbers the *parameters* of the graph. Actually, our definition excludes the trivial type $(1, 1)$ from the list of parameters, since it will not contribute anything to the bound. The following definitions are still with respect to a given $\sigma \in \Sigma_n$.

DEFINITION A.8. Let $\mathcal{I}$ be the set of all 11 non-trivial distinct types:

$$\mathcal{I} = \left\{ \begin{array}{ccc} (0,2), & & (0,4), \\ (1,2), & (1,3), & (1,4), \\ (2,2), & (2,3), & (2,4), \\ & (3,3), & (3,4), \\ & & (4,4) \end{array} \right\}.$$

DEFINITION A.9. Let $G = (V, E)$ be a graph. Its *list of parameters* is the 11-tuple $(k_{ij})_{(i,j)\in\mathcal{I}}$ where $k_{ij} = |E \cap T_{ij}|$ for every $(i, j) \in \mathcal{I}$. That is, k_{ij} is the number of edges of type (i, j) in G.

For convenience, lists of parameters are denoted (k_{ij}). We also let $k_{ji} = k_{ij}$ for every (i, j). At this point we slightly deviate from Bollobás's original definitions, in order to adjust to the non-regular case.

PROPOSITION A.10. *Let $G = (V, E)$ be a graph with the list of parameters (k_{ij}). For every $i \in \{2, 3, 4\}$ let r_i be the average degree of vertices from S_i in* G. *Then the following equations hold*:

$$\begin{cases} r_2 s_2 &= 2k_{20} + k_{21} + 2k_{22} + k_{23} + k_{24}, \\ r_3 s_3 &= k_{31} + k_{32} + 2k_{33} + k_{34}, \\ r_4 s_4 &= 2k_{40} + k_{41} + k_{42} + k_{43} + 2k_{44}. \end{cases} \tag{A.2}$$

PROOF. The three equations come from counting the sum of degrees of S_i for $i = 2, 3, 4$. From the definition of r_i this sum is $r_i s_i$. For the right-hand side, note that it indeed counts all the edges incident to S_i, with edges with both endpoints in S_i counted twice. The counting is done by considering the different possible types of the edges. ⊣

A.3. Bounding P_σ. We are now ready for the main part of the proof: bounding P_σ, the probability that σ is an automorphism of $\mathbf{G}_n$. We shall prove a bound which implies Equation (A.1), and that will finish the proof.

Let $(\mathbf{k}_{ij})$ be the list of parameters of the random graph $\mathbf{G}_n$.

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

DEFINITION A.11. A list (k_{ij}) is called *σ-valid* if

$$P_{\sigma,(k_{ij})} := \mathbb{P}\left(\sigma \text{ is an automorphism of } \mathbf{G}_n \text{ and } (\mathbf{k}_{ij}) = (k_{ij})\right) > 0.$$

By definition, $P_\sigma = \sum_{(k_{ij})} P_{\sigma,(k_{ij})}$ where the sum is over all σ-valid lists. To bound P_σ, we first bound $P_{\sigma,(k_{ij})}$ for a given σ-valid list (k_{ij}).

We fix a σ-valid list (k_{ij}). Then $P_{\sigma,(k_{ij})}$ is the probability that, for every $(i,j) \in \mathcal{I}$, $\mathbf{G}_n$ contains exactly k_{ij} edges of type (i,j), and also does not split (i,j)-orbits by Claim A.2.

- Let f_{ij} denote the number of sets E_{ij} of size k_{ij} which can be written as unions of (i,j)-orbits.
- Let $k = \sum_{(i,j)\in\mathcal{I}} k_{ij}$ be the number of edges of non-trivial types. Equivalently, k is the number of edges incident to S.

PROPOSITION A.12. *Following the above definitions, we have*

$$P_{\sigma,(k_{ij})} \le \left(\frac{c_0}{n}\right)^k \prod_{(i,j)\in\mathcal{I}} f_{ij}$$

where c_0 is the constant from property (P3).

PROOF. There are $\prod_{(i,j)\in\mathcal{I}} f_{ij}$ ways to choose, for every $(i,j) \in \mathcal{I}$, a set $E_{ij} \subseteq T_{ij}$ of size k_{ij} which is a union of (i,j)-orbits. Given such a choice of sets E_{ij}, let $E_0 = \bigcup_{(i,j)\in\mathcal{I}} E_{ij}$. From (P3) we have

$$\mathbb{P}\left(E_0 \subseteq E(\mathbf{G}_n)\right) \le \left(\frac{c_0}{n}\right)^{|E_0|} = \left(\frac{c_0}{n}\right)^k.$$

The union bound completes the proof. ⊣

The following simple combinatorial lemma allows us to bound the terms f_{ij}.

LEMMA A.13. *Let $\mathcal{A}$ be a finite collection of mutually disjoint finite sets, which contains exactly m_i sets of size* i *for every $i \ge 1$. Suppose there exists a positive integer* h *such that $m_i = 0$ for every $i < h$. Let $f(k;\mathcal{A})$ denote the number of sets of size* k *which are unions of some of the sets from $\mathcal{A}$. Then $f(k;\mathcal{A}) \le \binom{\lfloor t/h \rfloor}{\lfloor k/h \rfloor}$ where $t = \sum_i i m_i$ is the total size of all sets from $\mathcal{A}$.*

The proof is fairly straightforward and is given in (5, Equation (5)).

To effectively bound the terms f_{ij} with Lemma A.13, in a way which applies generally for all lists (k_{ij}), we must separate "large" k_{ij} from "small" k_{ij}, which motivates the use of ε_{ij} below.

PROPOSITION A.14. *For every $(i,j) \in \mathcal{I}$ let $\ell_{ij} = \frac{k_{ij}}{h_{ij}}$ and*

$$\varepsilon_{ij} = \begin{cases} 1, & \lfloor \ell_{ij} \rfloor > \frac{s}{22000}; \\ 0, & \lfloor \ell_{ij} \rfloor \le \frac{s}{22000}. \end{cases}$$

Then $\prod_{(i,j)\in\mathcal{I}} f_{ij} \le n^{\sum_{j=2}^4 \varepsilon_{1j}\ell_{1j}} \cdot s^{\sum_{2\le i\le j\le 4} \varepsilon_{ij}\ell_{ij}} \cdot c_1^k \cdot n^{\frac{s}{1000}}$ for some positive constant c_1.

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

PROOF. For every $(i,j) \in \mathcal{I}$, Lemma A.13 applies where $\mathcal{A}$ is the set of (i,j)-orbits, $k = k_{ij}$, $h = h_{ij}$ and $t = t_{ij}$. Therefore,

$$f_{ij} \leq \binom{\lfloor t_{ij}/h_{ij} \rfloor}{\lfloor k_{ij}/h_{ij} \rfloor} = \binom{\lfloor t_{ij}/h_{ij} \rfloor}{\lfloor \ell_{ij} \rfloor} \leq \binom{t_{ij}}{\lfloor \ell_{ij} \rfloor}.$$

When $\varepsilon_{ij} = 0$, we are satisfied with the very coarse bound

$$\binom{t_{ij}}{\lfloor \ell_{ij} \rfloor} \leq t_{ij}^{\lfloor \ell_{ij} \rfloor} \leq (n^2)^{s/22000} = n^{s/11000}.$$

When $\varepsilon_{ij} = 1$, we apply the bound

$$\binom{t_{ij}}{\lfloor \ell_{ij} \rfloor} \leq \left(\frac{\mathrm{e} t_{ij}}{\lfloor \ell_{ij} \rfloor}\right)^{\lfloor \ell_{ij} \rfloor} \leq \left(\frac{\mathrm{e} t_{ij}}{s/22000}\right)^{\ell_{ij}} = \left(c_1 \frac{t_{ij}}{s}\right)^{\ell_{ij}}$$

where $c_1 = 22000\mathrm{e}$. The bound on t_{ij} itself depends on the type (i,j). We have three cases:

- $2 \leq i \leq j \leq 4$. In this case $t_{ij} \leq s_i s_j \leq s^2$ and $f_{ij} \leq (c_1 s)^{\ell_{ij}}$.
- $i = 1$ and $2 \leq j \leq 4$. In this case $t_{1j} \leq s_1 s_j \leq ns$ and $f_{ij} \leq (c_1 n)^{\ell_{ij}}$.
- $i = 0$ and $j \in \{2,4\}$. In this case $t_{0j} = \frac{s_j}{2} \leq \frac{s}{2}$ and $f_{ij} \leq \left(\frac{c_1}{2}\right)^{\ell_{ij}}$.

Overall,

$$\begin{aligned}\prod_{(i,j)\in\mathcal{I}} f_{ij} &\leq n^{\sum_{j=2}^{4} \varepsilon_{1j}\ell_{1j}} \cdot s^{\sum_{2\leq i\leq j\leq 4} \varepsilon_{ij}\ell_{ij}} \cdot c_1^{\sum_{(i,j)\in\mathcal{I}} \ell_{ij}} \cdot n^{\sum_{(i,j)\in\mathcal{I}} \frac{s}{11000}} \\ &\leq n^{\sum_{j=2}^{4} \varepsilon_{1j}\ell_{1j}} \cdot s^{\sum_{2\leq i\leq j\leq 4} \varepsilon_{ij}\ell_{ij}} \cdot c_1^{k} \cdot n^{\frac{s}{1000}}.\end{aligned}$$

⊣

COROLLARY A.15. *For large enough* n,

$$P_{\sigma,(k_{ij})} \leq n^{\frac{s}{100}} n^{-k} \cdot n^{\sum_{j=2}^{4} \varepsilon_{1j}\ell_{1j}} s^{\sum_{2\leq i\leq j\leq 4} \varepsilon_{ij}\ell_{ij}}.$$

PROOF. Due to Proposition A.14, we get

$$P_{\sigma,(k_{ij})} \leq \left(\frac{c_0}{n}\right)^k \prod_{(i,j)\in\mathcal{I}} f_{ij} \leq \left(\frac{c_0}{n}\right)^k \cdot c_1^k \cdot n^{\frac{s}{1000}} \cdot n^{\sum_{j=2}^{4} \varepsilon_{1j}\ell_{1j}} \cdot s^{\sum_{2\leq i\leq j\leq 4} \varepsilon_{ij}\ell_{ij}}.$$

In addition,

$$\begin{aligned}\left(\frac{c_0}{n}\right)^k \cdot c_1^k \cdot n^{\frac{s}{1000}} &= (c_0 c_1)^k \cdot n^{\frac{s}{1000}} n^{-k} \\ &\overset{(*)}{\leq} (c_0 c_1)^{\Delta s} \cdot n^{\frac{s}{1000}} n^{-k} \overset{(**)}{=} c_2^s \cdot n^{\frac{s}{1000}} n^{-k} \leq n^{\frac{s}{100}} n^{-k}\end{aligned}$$

for large enough n. Inequality $(*)$ follows from Property (P2) of the model (and our assumption that (k_{ij}) is σ-valid). Indeed, recall that k is the number of edges incident to S; in a graph with maximum degree at most Δ, this number is at most Δs. In Equality $(**)$ we simply define $c_2 = (c_0 c_1)^{\Delta}$. ⊣

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

The next assertion is the main technical part of the proof.

LEMMA A.16. *For every σ-valid list (k_{ij}),*

$$P_{\sigma,(k_{ij})} \le \frac{1}{n^{(1+\frac{1}{100})s}} s_2^{s_2/2} s_3^{s_3/3}.$$

Note that the bound depends only on σ, and not on (k_{ij}).

PROOF. From Corollary A.15, it suffices to prove

$$\left[n^{\frac{s}{100}} n^{-k} \cdot n^{\sum_{j=2}^4 \varepsilon_{1j}\ell_{1j}} s^{\sum_{2\le i\le j\le 4}\varepsilon_{ij}\ell_{ij}}\right] \cdot n^s s_2^{-s_2/2} s_3^{-s_3/3} \le n^{-\frac{s}{100}}.$$

Let us denote

$$A = n^{s-k} \cdot n^{\sum_{j=2}^4 \varepsilon_{1j}\ell_{1j}} s^{\sum_{2\le i\le j\le 4}\varepsilon_{ij}\ell_{ij}} \cdot s_2^{-s_2/2} s_3^{-s_3/3}.$$

Then we need to prove $n^{\frac{s}{100}} A \le n^{-\frac{s}{100}}$. Equivalently, we need to prove

$$-\frac{\ln A}{\ln n} \ge \frac{2}{100} s. \tag{A.3}$$

Let us bound $-\frac{\ln A}{\ln n}$ from below. Denote $\alpha = \frac{\ln s}{\ln n}$; then

$$-\frac{\ln A}{\ln n} = (k-s) - \sum_{j=2}^4 \varepsilon_{1j}\ell_{1j} - \alpha \sum_{2\le i\le j\le 4} \varepsilon_{ij}\ell_{ij} + \frac{1}{2} s_2 \frac{\ln s_2}{\ln n} + \frac{1}{3} s_3 \frac{\ln s_3}{\ln n}. \tag{A.4}$$

At this point we invoke Equation (A.2). Then

$$\begin{aligned} k - s &= \sum_{(i,j)\in\mathcal{I}} k_{ij} - (s_2 + s_3 + s_4) \\ &= \left(1 - \frac{2}{r_2}\right) k_{20} + \left(1 - \frac{1}{r_2}\right) k_{21} + \left(1 - \frac{2}{r_2}\right) k_{22} \\ &\quad + \left(1 - \frac{1}{r_2} - \frac{1}{r_3}\right) k_{23} + \left(1 - \frac{1}{r_2} - \frac{1}{r_4}\right) k_{24} + \left(1 - \frac{1}{r_3}\right) k_{31} + \left(1 - \frac{2}{r_3}\right) k_{33} \\ &\quad + \left(1 - \frac{1}{r_3} - \frac{1}{r_4}\right) k_{34} + \left(1 - \frac{2}{r_4}\right) k_{40} + \left(1 - \frac{1}{r_4}\right) k_{41} + \left(1 - \frac{2}{r_4}\right) k_{44}. \end{aligned}$$

In addition,

$$\sum_{j=2}^4 \varepsilon_{1j}\ell_{1j} = \varepsilon_{12}\frac{k_{12}}{2} + \varepsilon_{13}\frac{k_{13}}{3} + \varepsilon_{14}\frac{k_{14}}{4},$$

$$\alpha \sum_{2\le i\le j\le 4} \varepsilon_{ij}\ell_{ij} = \alpha\varepsilon_{22}\frac{k_{22}}{2} + \alpha\varepsilon_{23}\frac{k_{23}}{6} + \alpha\varepsilon_{24}\frac{k_{24}}{4} + \alpha\varepsilon_{33}\frac{k_{33}}{3} + \alpha\varepsilon_{34}\frac{k_{34}}{6} + \alpha\varepsilon_{44}\frac{k_{44}}{4}.$$

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

As for $\frac{1}{2}s_2\frac{\ln s_2}{\ln n}$, we have

$$\begin{aligned}\frac{1}{2}s_2\frac{\ln s_2}{\ln n} &= \frac{1}{2\ln n}\left[s_2\ln s - s_2\ln\left(\frac{s}{s_2}\right)\right]\\ &\geq \frac{1}{2\ln n}\left[s_2\ln s - s_2\cdot\frac{s}{s_2}\right]\\ &= \frac{\alpha s_2}{2} - \frac{s}{2\ln n}\end{aligned}$$

and similarly,

$$\frac{1}{3}s_3\frac{\ln s_3}{\ln n} \geq \frac{\alpha s_3}{3} - \frac{s}{3\ln n}.$$

Again, we apply Equation (A.2) to write

$$\begin{aligned}\frac{\alpha s_2}{2} &= \frac{\alpha}{2r_2}\left(2k_{20}+k_{21}+2k_{22}+k_{23}+k_{24}\right),\\ \frac{\alpha s_3}{3} &= \frac{\alpha}{3r_3}\left(k_{31}+k_{32}+2k_{33}+k_{34}\right).\end{aligned}$$

Overall, putting everything back into (A.4), we have the following lower bound on $-\frac{\ln A}{\ln n}$. We use colors to help keeping track of the terms:

$$\begin{aligned}-\frac{\ln A}{\ln n} \geq& -\frac{s}{\ln n}+(k-s)-\sum_{j=2}^{4}\varepsilon_{1j}\ell_{1j}-\alpha\sum_{2\leq i\leq j\leq 4}\varepsilon_{ij}\ell_{ij}+\frac{\alpha s_2}{2}+\frac{\alpha s_3}{3}\\ =& -\frac{s}{\ln n}+\left(1-\frac{2}{r_2}+\frac{\alpha}{r_2}\right)k_{20}+\left(1-\frac{1}{r_2}-\frac{\varepsilon_{21}}{2}+\frac{\alpha}{2r_2}\right)k_{21}\\ &+\left(1-\frac{2}{r_2}-\frac{\alpha\varepsilon_{22}}{2}+\frac{\alpha}{r_2}\right)k_{22}+\left(1-\frac{1}{r_2}-\frac{1}{r_3}-\frac{\alpha\varepsilon_{23}}{6}+\frac{\alpha}{2r_2}+\frac{\alpha}{3r_3}\right)k_{23}\\ &+\left(1-\frac{1}{r_2}-\frac{1}{r_4}-\frac{\alpha\varepsilon_{24}}{4}+\frac{\alpha}{2r_2}\right)k_{24}+\left(1-\frac{1}{r_3}-\frac{\varepsilon_{31}}{3}+\frac{\alpha}{3r_3}\right)k_{31}\\ &+\left(1-\frac{2}{r_3}-\frac{\alpha\varepsilon_{33}}{3}+\frac{2\alpha}{3r_3}\right)k_{33}+\left(1-\frac{1}{r_3}-\frac{1}{r_4}-\frac{\alpha\varepsilon_{34}}{6}+\frac{\alpha}{3r_3}\right)k_{34}\\ &+\left(1-\frac{2}{r_4}\right)k_{40}+\left(1-\frac{1}{r_4}-\frac{\varepsilon_{41}}{4}\right)k_{41}+\left(1-\frac{2}{r_4}-\frac{\alpha\varepsilon_{44}}{4}\right)k_{44}.\end{aligned}$$

We write simpler lower bounds on the terms in the brackets. In negative summands with ε_{ij} we replace it with 1. In addition, we omit some positive black!50!white and black!40!white summands:

$$\begin{aligned}-\frac{\ln A}{\ln n} \geq& -\frac{s}{\ln n}+\left(1-\frac{2}{r_2}\right)k_{20}+\left(1-\frac{1}{r_2}-\frac{1}{2}\right)k_{21}\\ &+\left(1-\frac{2}{r_2}-\frac{\alpha}{2}+\frac{\alpha}{r_2}\right)k_{22}+\left(1-\frac{1}{r_2}-\frac{1}{r_3}-\frac{\alpha}{6}+\frac{\alpha}{2r_2}\right)k_{23}\\ &+\left(1-\frac{1}{r_2}-\frac{1}{r_4}-\frac{\alpha}{4}+\frac{\alpha}{2r_2}\right)k_{24}+\left(1-\frac{1}{r_3}-\frac{1}{3}\right)k_{31}\end{aligned}$$

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

$$+\left(1-\frac{2}{r_3}-\frac{\alpha}{3}+\frac{2\alpha}{3r_3}\right)k_{33}+\left(1-\frac{1}{r_3}-\frac{1}{r_4}-\frac{\alpha}{6}+\frac{\alpha}{3r_3}\right)k_{34}$$
$$+\left(1-\frac{2}{r_4}\right)k_{40}+\left(1-\frac{1}{r_4}-\frac{1}{4}\right)k_{41}+\left(1-\frac{2}{r_4}-\frac{\alpha}{4}\right)k_{44}.$$

In $k_{22}, k_{23}, k_{24}, k_{33}, k_{34}, k_{44}$ there are summands involving α. In all six cases, the sum of these summands is of the form αx where $x \leq 0$. This is because $r_2, r_3, r_4 \geq 3$ (unless the respective set S_i is empty, but in that case the corresponding k-factor equals 0), which follows from Property (P1) of the model (and our assumption that (k_{ij}) is σ-valid). Also note that $\alpha \leq 1$ by definition. Therefore, replacing every appearance of α with 1 can only decrease the sum. By doing so (and simplifying) we obtain

$$-\frac{\ln A}{\ln n} \geq -\frac{s}{\ln n}+\left(1-\frac{2}{r_2}\right)k_{20}+\left(\frac{1}{2}-\frac{1}{r_2}\right)k_{21}$$
$$+\left(\frac{1}{2}-\frac{1}{r_2}\right)k_{22}+\left(\frac{5}{6}-\frac{1}{2r_2}-\frac{1}{r_3}\right)k_{23}$$
$$+\left(\frac{3}{4}-\frac{1}{2r_2}-\frac{1}{r_4}\right)k_{24}+\left(\frac{2}{3}-\frac{1}{r_3}\right)k_{31}$$
$$+\left(\frac{2}{3}-\frac{4}{3r_3}\right)k_{33}+\left(\frac{5}{6}-\frac{2}{3r_3}-\frac{1}{r_4}\right)k_{34}$$
$$+\left(1-\frac{2}{r_4}\right)k_{40}+\left(\frac{3}{4}-\frac{1}{r_4}\right)k_{41}+\left(\frac{3}{4}-\frac{2}{r_4}\right)k_{44}.$$

The right-hand side can be equivalently rewritten as

$$-\frac{\ln A}{\ln n} \geq -\frac{s}{\ln n}+\left(\frac{r_2-2}{r_2}\right)k_{20}+\left(\frac{r_2-2}{2r_2}\right)k_{21}$$
$$+\left(\frac{r_2-2}{2r_2}\right)k_{22}+\left(\frac{2r_2-3}{6r_2}+\frac{r_3-2}{2r_3}\right)k_{23}$$
$$+\left(\frac{r_2-2}{4r_2}+\frac{r_4-2}{2r_4}\right)k_{24}+\left(\frac{2r_3-3}{3r_3}\right)k_{31}$$
$$+\left(\frac{2r_3-4}{3r_3}\right)k_{33}+\left(\frac{r_3-2}{3r_3}+\frac{r_4-2}{2r_4}\right)k_{34}$$
$$+\left(\frac{r_4-2}{r_4}\right)k_{40}+\left(\frac{3r_4-4}{4r_4}\right)k_{41}+\left(\frac{3r_4-8}{4r_4}\right)k_{44}.$$

Since $r_2, r_3, r_4 \geq 3$ (unless the respective k_i equals 0), all the numerators in all the fractions above are ≥ 1, so replacing them with 1 only decreases the sum. By doing so, and rearranging summands, we obtain

$$-\frac{\ln A}{\ln n} \geq -\frac{s}{\ln n}+\frac{1}{r_2}\left[k_{20}+\frac{k_{21}}{2}+\frac{k_{22}}{2}+\frac{k_{23}}{6}+\frac{k_{24}}{4}\right]$$
$$+\frac{1}{r_3}\left[\frac{k_{31}}{3}+\frac{k_{32}}{2}+\frac{k_{33}}{3}+\frac{k_{34}}{3}\right]$$

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

$$
\begin{aligned}
&+ \frac{1}{r_4}\left[k_{40} + \frac{k_{41}}{4} + \frac{k_{42}}{2} + \frac{k_{43}}{2} + \frac{k_{44}}{4}\right] \\
&\overset{(16)}{\geq} -\frac{s}{\ln n} + \frac{s_2}{6} + \frac{s_3}{6} + \frac{s_4}{8} \\
&\geq -\frac{s}{\ln n} + \frac{s}{8} = \left(\frac{1}{8} - o(1)\right)s \\
&\geq \frac{2}{100}s.
\end{aligned}
$$

In conclusion, we have proved Inequality (A.3), and as explained this completes the proof. ⊣

COROLLARY A.17.

$$P_\sigma \leq (\Delta s + 1)^{11} \frac{1}{n^{(1+\frac{1}{100})s}} s_2^{s_2/2} s_3^{s_3/3}.$$

PROOF. We have $P_\sigma = \sum_{(k_{ij})} P_{\sigma,(k_{ij})}$ where the sum is over all σ-valid lists (k_{ij}). Due to Lemma A.16, it only remains to explain why there at most $(\Delta s + 1)^{11}\sigma$-valid lists (k_{ij}).

Every σ-valid list is a list of 11 non-negative integers, and their sum k must satisfy $k \leq \Delta s$ (due to Property (P2) of the model). Therefore, the number of σ-valid lists is bounded by the number of 11-tuples of integers, each between 0 and Δs, which is $(\Delta s + 1)^{11}$. ⊣

We are now ready to prove the asymmetry of $\mathbf{G}_n$.

PROOF OF THEOREM 2.16. As previously explained, it suffices to prove Equation (A.1). Let $M(s, s_2, s_3)$ be the set of permutations σ with s non-fixed vertices, s_2 vertices in 2-cycles and s_3 vertices in 3-cycles. Then we shall use the following fairly coarse bound:

$$|M(s, s_2, s_3)| \leq (n)_s \cdot \frac{1}{\left(\frac{s_2}{2}\right)!} \cdot \frac{1}{\left(\frac{s_3}{3}\right)!} \leq \frac{n^s}{\left(\frac{s_2}{2}\right)!\left(\frac{s_3}{3}\right)!}.$$

From Corollary A.17 we deduce

$$
\begin{aligned}
\sum_{\sigma \in M(s,s_2,s_3)} P_\sigma &\leq \frac{n^s}{\left(\frac{s_2}{2}\right)!\left(\frac{s_3}{3}\right)!} \cdot (\Delta s + 1)^{11} \frac{1}{n^{(1+\frac{1}{100})s}} s_2^{s_2/2} s_3^{s_3/3}, \\
&\leq c_3^s n^{-\frac{s}{100}}
\end{aligned}
$$

where c_3 is some positive constant, since $s > 1$ provided by $\sigma \neq \mathrm{id}$. Every $\sigma \neq \mathrm{id}$ belongs to some set $M(s, s_2, s_3)$ with $2 \leq s \leq n$, and for every $2 \leq s \leq n$ there at most s^2 choices for s_2, s_3 such that $M(s, s_2, s_3)$ is non-empty. Overall,

$$\sum_{\sigma \in \Sigma_n \setminus \{\mathrm{id}\}} P_\sigma \leq \sum_{s=2}^{n} s^2 c_3^s n^{-\frac{s}{100}} = O\left(\sum_{s=2}^{n} n^{-\frac{s}{200}}\right) = o(1).$$

Thus the proof is completed. ⊣

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

B. Properties of extensions in random graphs. The following properties of safe and rigid extensions will be useful. For their proofs, see (22, Chapter 4).

PROPOSITION B.1. *Let (R,H) be a rooted graph.*

1. *If (R,H) is not safe then it has a rigid subextension.*
2. *Assume that (R,H) is rigid, H is a subgraph of some larger graph G, and $X \subseteq V(G)$. If $R \cup X \neq V(H) \cup X$ then $(R \cup X, G[V(H) \cup X])$ is rigid.*

Recall that $\mathbf{G} \sim G(n, n^{-\alpha})$, $M = M(n) = (\ln n)^{1/(10d)}$, and $r, v, e = O(M)$.

Given an r-tuple of vertices $\vec{x}$, we let $\mathbf{N}_{\vec{x}}^{(R,H)}$ denote the random variable which counts (R,H)-extensions of $\vec{x}$ in $\mathbf{G}$. By definition,

$$\mathbb{E}\left(\mathbf{N}_{\vec{x}}^{(R,H)}\right) = \frac{1}{a}(n-r)_v p^e \sim \frac{1}{a} n^v p^e = \frac{1}{a} n^{v-\alpha e},$$

where a is the number of automorphisms of the rooted graph (R,H) which preserve the root vertices.

THEOREM B.2. *W.h.p. in $\mathbf{G}$, for every safe (R,H) with $r \leq M$ roots and $v \leq M$ non-roots, and for every* r*-tuple of vertices $\vec{x}$, we have $\mathbf{N}_{\vec{x}}^{(R,H)} \sim \mathbb{E}\mathbf{N}_{\vec{x}}^{(R,H)}$ uniformly. More explicitly, there exists $\delta = o(1)$ (which is uniform w.r.t. (R,H) and $\vec{x}$) such that w.h.p.*

$$\frac{\mathbf{N}_{\vec{x}}^{(R,H)}}{\mathbb{E}\mathbf{N}_{\vec{x}}^{(R,H)}} \in [1-\delta, 1+\delta]. \tag{B.1}$$

PROOF. To prove the theorem, we shall prove that there exists $\delta = o(1)$ such that for any (R,H) with $r \leq M$ roots and $v \leq M$ non-roots and any r-tuple $\vec{x}$, (B.1) holds with probability $1 - \exp(-\lambda)$, where $\lambda = \exp\left(\Omega\left(\frac{\ln n}{M^d}\right)\right)$. Then it would only remain to apply the union bound over (R,H) and $\vec{x}$. Indeed, the number of possible safe rooted graphs with $r \leq M$ roots and $v \leq M$ non-roots is trivially bounded by $\exp\left(O(M^2)\right)$. The number of r-tuples of vertices $\vec{x}$ is $O(n^r) = \exp\left(O(M)\ln n\right)$. So the probability of having (R,H) and $\vec{x}$ such that $N_{\vec{x}}^{(R,H)}$ does not satisfy (B.1) is

$$\exp\left(O(M^2) + O(M)\ln n - \lambda\right) = o(1).$$

So, from now on let us fix a safe rooted graph (R,H) of type (v,e) with $r \leq M$ roots and $v \leq M$, and also fix an r-tuple of vertices $\vec{x}$. Denote $\mathbf{N} = \mathbf{N}_{\vec{x}}^{(R,H)}$. We prove concentration of $\mathbf{N}$ via Kim and Vu's powerful concentration result (see 14). Note that for the rest of the proof, e denotes the number of edges in the rooted graph (not the Napier's constant).

Let $\mathcal{Y}$ be the set of all $\frac{1}{a}(n-r)_v v$-tuples $\vec{y}$ which are potential (R,H)-extensions of $\vec{x}$. Let $\mathcal{E}$ be the set of all $\binom{n}{2} - \binom{r}{2}$ pairs of vertices which do not connect two vertices of $\vec{x}$. For every $\vec{y} \in \mathcal{Y}$, let $P_{\vec{y}}$ be the set of all edges required by the (R,H)-extension $(\vec{x}, \vec{y})$. This is a subset of $\mathcal{E}$ of size e. Finally, for every $i \in \mathcal{E}$ let $\mathbf{I}_i$ be the indicator of the event $i \in E(\mathbf{G})$. Then we can write $\mathbf{N}$ as the value of a multivariate

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

polynomial on $E(\mathbf{G})$ as follows:

$$\mathbf{N} = \sum_{\vec{y}\in\mathcal{Y}} \left[\prod_{i\in P_{\vec{y}}} \mathbf{I}_i \right].$$

For a set of edges $A \subseteq \mathcal{E}$ with $|A| = e_0 \leq e$, we define

$$\mathcal{Y}_A = \left\{ \vec{y} \in \mathcal{Y} : A \subseteq P_{\vec{y}} \right\}, \quad \mathbf{N}_A = \sum_{\vec{y}\in\mathcal{Y}_A} \left[\prod_{i\in P_{\vec{y}}\setminus A} \mathbf{I}_i \right].$$

For every $0 \leq e_0 \leq e$ let

$$E_{e_0} = \max_{A\subseteq\mathcal{E},|A|=e_0} \mathbb{E}\left(\mathbf{N}_A\right).$$

Also define

$$E = \max_{e_0\geq 0} E_{e_0}, \quad E' = \max_{e_0\geq 1} E_{e_0}.$$

Notice that $E_0 = \mathbb{E}\mathbf{N}$, thus $E = \max\{E', \mathbb{E}\mathbf{N}\}$. Kim–Vu's concentration result asserts that for every $\lambda > 1$,

$$\mathbb{P}\left(|\mathbf{N} - \mathbb{E}\mathbf{N}| > \frac{8^e}{\sqrt{e!}} \sqrt{E \cdot E'} \cdot \lambda^e \right) = O\left(\exp\left(-\lambda + (e-1)\ln|\mathcal{E}|\right)\right). \tag{B.2}$$

To make use of this result we first bound $\mathbb{E}(\mathbf{N}_A)$ for every set A of $e_0 \leq e$ edges. It is easy to see that $\mathbb{E}(\mathbf{N}_A) = |\mathcal{Y}_A|\, p^{e-e_0}$; let us bound $|\mathcal{Y}_A|$. Consider the set of vertices incident to the edges of A, excluding the vertices of $\vec{x}$; denote it V_0 and its size v_0. We can bound $|\mathcal{Y}_A|$ by the number of v-tuples $\vec{y}$ which contain all the vertices of V_0 (in any positions), which is

$$(v)_{v_0} \cdot (n - r - v_0)_{v-v_0} \leq v! \cdot n^{v-v_0}.$$

If the vertices of V_0 and the edges of A cannot be completed into an (R, H)-extension of $\vec{x}$, then actually $\mathcal{Y}_A$ is empty and $\mathbb{E}(\mathbf{N}_A) = 0$ trivially. Otherwise, from safeness of (R, H) we deduce $v_0 - \alpha e_0 > 0$. Overall $\mathbb{E}(\mathbf{N}_A) \leq v! \cdot \frac{n^{v-\alpha e}}{n^{v_0-\alpha e_0}}$.

Comparing $\mathbb{E}(\mathbf{N}_A)$ to $\mathbb{E}(\mathbf{N})$ yields

$$\frac{\mathbb{E}(\mathbf{N}_A)}{\mathbb{E}(\mathbf{N})} \leq \frac{av!}{n^{v_0-\alpha e_0}} \leq \frac{(v!)^2}{n^{v_0-\alpha e_0}}.$$

Maximizing over all non-empty sets A of at most e edges, we obtain the bound

$$E' \leq \frac{(v!)^2}{n^\beta} \cdot \mathbb{E}(\mathbf{N}),$$

where $\beta = \min_{v_0,e_0}(v_0 - \alpha e_0)$, with the minimum taken over $1 \leq v_0 \leq v$ and $e_0 < \frac{v_0}{\alpha}$.

It is at this point that we encounter the main difference from the original argument: since v may now grow with n, we get that β is no longer constant, but approaches 0 in a way that reflects the behavior of rational approximations of α. To provide a lower bound on β we invoke the assumption that $\left|\alpha - \frac{p}{q}\right| \geq \frac{1}{q^d}$ for all but finitely many

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

rationals $\frac{p}{q}$. Taking any $1 \leq v_0 \leq v$ and $e_0 < \frac{v_0}{\alpha}$ (with a finite number of possible exceptions) we have

$$v_0 - \alpha e_0 = e_0 \left(\frac{v_0}{e_0} - \alpha \right) = e_0 \left| \alpha - \frac{v_0}{e_0} \right| \geq \frac{1}{e_0^{d-1}}.$$

Overall, this proves the lower bound $\beta \geq \frac{c}{e^{d-1}}$ where c is a constant (which comes from the finite number of exceptions). For convenience, we rewrite this lower bound in terms of v: since $e < \frac{v}{\alpha}$, we have $\beta \geq \frac{c}{v^{d-1}}$ (where now c might be a different constant).

We now return to Equation (B.2). Let us first take care of the left-hand side. From what we have seen so far,

$$\begin{aligned} \frac{E'}{\mathbb{E}\mathbf{N}} \leq \frac{(v!)^2}{n^\beta} \leq \frac{v^{2v}}{n^\beta} &= \exp\left(2v \ln v - \beta \ln n\right) \\ &\leq \exp\left(2v \ln v - \frac{c}{v^{d-1}} \ln n\right). \end{aligned}$$

Since $v = O(M) = O\left((\ln n)^{\frac{1}{10d}}\right)$, we have

$$E' \leq \exp\left(-(c - o(1)) \frac{\ln n}{v^{d-1}}\right) \cdot \mathbb{E}\mathbf{N}.$$

We can also trivially bound $\frac{8^e}{\sqrt{e!}}$ from above by a constant, and for every $\lambda > 1$, we have $\lambda^e \leq \lambda^{\frac{1}{\alpha} v}$. Let us take $\lambda = \exp\left(\frac{\alpha c}{4} \cdot \frac{\ln n}{v^d}\right)$. This choice guarantees

$$\frac{8^e}{\sqrt{e!}} \sqrt{E \cdot E'} \cdot \lambda^e \leq \exp\left(-\left(\frac{c}{4} - o(1)\right) \frac{\ln n}{v^{d-1}}\right) \cdot \mathbb{E}\mathbf{N}.$$

In particular $\frac{8^e}{\sqrt{e!}} \sqrt{E \cdot E'} \cdot \lambda^e \leq \delta \mathbb{E}\mathbf{N}$ for $\delta = o(1)$ which does not depend on the specific (R, H) or $\vec{x}$.

Finally, from Equation (B.2) with λ, δ as above, we have

$$\begin{aligned} \mathbb{P}\left(|N - \mathbb{E}N| > \delta \mathbb{E}N\right) &= O\left(\exp\left(-\lambda + (e-1)\ln|\mathcal{E}|\right)\right) \\ &= \exp\left[-\lambda + O(v \ln n)\right] \\ &= \exp\left(-(1 + o(1))\lambda\right) \end{aligned}$$

with $\lambda = \exp\left(\Omega\left(\frac{\ln n}{M^d}\right)\right)$. That finishes the proof. ⊣

Remark B.3. For technical reasons, we also need the following straightforward generalization of Theorem B.2. *Following the same notation, w.h.p. for every tuple $\vec{x}$ and every set $U_{\vec{x}}$ of "forbidden" vertices of size $O(M^{d+1})$ the conclusion of Theorem B.2 holds for the random variable $\mathbf{P}^{(R,H)}_{\vec{x}}$ counting only (R,H)-extensions $\vec{y}$ of $\vec{x}$ which do not use any vertices from $U_{\vec{x}}$.* This generalization can be proved as a direct corollary of Theorem B.2, by adding the forbidden vertices $U_{\vec{x}}$ to the rooted graph as roots, without any edges. Technically, this generalization requires the replacement of $r \leq M$ with $r = O(M^d)$, but it has negligible effect on computations since d is a constant.

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

B.1. Proof of bounded closure lemma. The proof is similar to the proof of the Finite Closure Theorem from 22. For convenience, we present the full argument.

We start by fixing $0 \leq r, t \leq M$. As in the original argument, set

$$\beta = \min_{v_0, e_0} \left(\frac{\alpha e_0 - v_0}{v_0} \right),$$

where the minimum is over all $1 \leq v_0 \leq t$ and $e_0 > \frac{1}{\alpha} v_0$. Again, this is no longer a positive constant; we now have $\beta \geq \frac{c}{t^d}$ where $c = c_\alpha$ is a positive constant.

Define $K = \left\lceil \frac{r}{\beta} \right\rceil + 1$ and let us show that with probability $1 - o(M^{-2})$, the t-closure of any r-tuple contains at most $r + K$ vertices. Then, taking a union bound over all possible $0 \leq r, t \leq M$ will finish the proof.

Assume that a set R of r vertices has $|\mathrm{cl}_t(R)| > r + K$. This means there exists a sequence $R = S_0 \subsetneq S_1 \subsetneq S_2 \subsetneq \cdots \subsetneq S_u \subseteq \mathrm{cl}_t(R)$ where each (S_{i-1}, S_i) is a dense rooted graph of type (v_i, e_i) with $v_i \leq t$ and also the number of non-roots of (R, S_u) satisfies $\sum_{i=1}^{u} v_i \in [K, K + t]$.

The induced subgraph on S_u has $v_* = r + \sum_{i=1}^{u} v_i$ vertices and at least $e_* = \sum_{i=1}^{u} e_i$ edges. Notice that

$$v_* - \alpha e_* = r + \sum_{i=1}^{u} (v_i - \alpha e_i) \leq r - \sum_{i=1}^{u} \beta v_i \leq r - \beta K \leq -\beta.$$

It is therefore sufficient to prove the following statement. W.h.p., there exists no subgraph H of $\mathbf{G}$ such that $v(H) \in [r + K, r + K + t]$ and $v(H) - \alpha e(H) \leq -\beta$. Fix a graph H which satisfies the above conditions. Its expected number of copies is

$$\begin{aligned} \mathbb{E}(\mathbf{X}_H) \leq n^{v(H) - \alpha e(H)} \leq n^{-\beta} &= \exp(-\beta \ln n) \\ &= \exp\left(-\Omega\left(\frac{\ln n}{M^d} \right) \right) \end{aligned}$$

since $\beta \geq \frac{c}{t^d}$ and $t = O(M)$. Let us recall that $r = O(M)$ and $K = O\left(rt^d\right) = O\left(M^{d+1}\right)$. It is sufficient to take the union bound only over those H with minimal number of edges; that is, we may assume $e(H) = \frac{1}{\alpha} v(H) + O(1)$. Therefore, the probability of existence of a subgraph H of $\mathbf{G}$ as above is at most

$$\begin{aligned} &\exp\left[O\left((r + K + t) \ln(r + K + t) \right) - \Omega\left(\frac{\ln n}{M^d} \right) \right] \\ &= \exp\left(O\left(M^{d+1} \ln M \right) - \Omega\left(\frac{\ln n}{M^d} \right) \right). \end{aligned}$$

This is $o(M^{-2})$ since $M = (\ln n)^{\frac{1}{10d}}$.

B.2. Proof of Lemma 5.8. The idea is to use Theorem B.2 to claim that there are many (R, H)-extensions, and also that the majority of them must be generic.

From Theorem B.2, Remark B.3, and the Bounded Closure Lemma (with M replaced by $2M$), w.h.p. the following properties hold true in $\mathbf{G}$:

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

1. For every safe rooted graph (R', H') with $r' \leq 2M$ roots and $v' \leq 2M$ non-roots, and for every r'-tuple $\vec{x}'$,
$$\mathbf{N}_{\vec{x}'}^{(R',H')} \sim \mathbb{E}\left(\mathbf{N}_{\vec{x}'}^{(R',H')}\right). \tag{B.3}$$
2. For every r-tuple $\vec{x}$, $r \leq M$, letting $U_{\vec{x}} = \mathrm{cl}_{2M}(\vec{x}) \setminus \vec{x}$, we have $|U_{\vec{x}}| = O\left(M^{d+1}\right)$.
3. For every safe rooted graph (R, H) with $r \leq M$ roots and $v \leq M$ non-roots, and for every r-tuple $\vec{x}$ of vertices,
$$\mathbf{P}_{\vec{x}}^{(R,H)} \sim \mathbb{E}\left(\mathbf{P}_{\vec{x}}^{(R,H)}\right),$$
where $\mathbf{P}_{\vec{x}}^{(R,H)}$ counts (R, H)-extensions $\vec{y}$ which do not use any vertex from $U_{\vec{x}}$ (we call such (R, H)-extensions *good*).

Let $\mathcal{A}$ be the event that the three conditions above hold; then $\mathbb{P}(\mathcal{A}) = 1 - o(1)$. We now fix a safe rooted graph (R, H) with $r \leq M$ roots and $v \leq M$ non-roots, a non-negative integer $t \leq M$ and an r-tuple of vertices $\vec{x}$. To complete the proof, it suffices to show that given the event $\mathcal{A}$, $\vec{x}$ has a t-generic (R, H)-extension $\vec{y}$.

Let (v, e) denote the type of (R, H). Since $\mathcal{A}$ holds, we know that
$$\mathbf{P}_{\vec{x}}^{(R,H)} \sim \mathbb{E}\left(\mathbf{P}_{\vec{x}}^{(R,H)}\right) \sim \frac{n^{v-\alpha e}}{a},$$
where a is the number of automorphisms of (R, H) which preserve the roots. We prove that, given that $\mathcal{A}$ holds, the number of good (R, H)-extensions $\vec{y}$ of $\vec{x}$ which fail to be t-generic is $o\left(\frac{n^{v-\alpha e}}{a}\right)$. First, let us bound the number of good (R, H)-extensions $\vec{y}$ which violate the first part of Definition 5.7, meaning that they contain additional edges except those specified by H. Assume that $\vec{y}$ is such an extension; then it forms an (R, H^+)-extension of $\vec{x}$ of type (v, e^+) with $e^+ > e$. This extension cannot have any rigid subextensions, because $\vec{y}$ shares no vertices with $U_{\vec{x}}$. From part 1 of Proposition B.1 it is safe. The number of (R, H^+)-extensions of $\vec{x}$ is therefore
$$O\left(n^v p^{e^+}\right) = O\left(an^{-\alpha} \cdot \frac{n^{v-\alpha e}}{a}\right) = n^{-\alpha+o(1)} \cdot \frac{n^{v-\alpha e}}{a}.$$
The last equality follows from the bound $a \leq v! = n^{o(1)}$. Summing over $\exp\left(O(M^2)\right)$ possibilities for the rooted graph (R, H^+), we obtain a bound of $o\left(\frac{n^{v-\alpha e}}{a}\right)$ on the number of (R, H)-extensions of $\vec{x}$ which violate the first part of Definition 5.7.

Now we bound the number of good (R, H)-extensions $\vec{y}$ which violate the second part of Definition 5.7. Assume that $\vec{y}$ is such an extension, and let $\vec{z}$ be a rigid extension of $\vec{x} \cup \vec{y}$ with at most t non-roots and at least one edge between $\vec{z}$ and $\vec{y}$. Let $(V(H), H_1)$ denote the rooted graph corresponding to this extension, and let (v_1, e_1) denote its type. Without loss of generality, we may assume that $\vec{z}$ is minimal rigid extension of $\vec{x} \cup \vec{y}$, in the sense that $(V(H), H_1)$ has no non-trivial rigid subextensions. Note that $\vec{z} \cup \vec{y}$ is an (R, H_1)-extension of $\vec{x}$, which is of type $(v + v_1, e + e_1)$, with $v + v_1 \leq v + t \leq 2M$. We then consider cases.

Case 1. (R, H_1) is safe. The number of (R, H_1)-extensions of $\vec{x}$ is
$$O\left(n^{v+v_1} p^{e+e_1}\right) = O(n^{v_1-\alpha e_1} n^{v-\alpha e}).$$

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

$(V(H), H_1)$ is rigid, so $v_1 - \alpha e_1 < 0$. Actually, we have $v_1 - \alpha e_1 = -\Omega\left(\frac{1}{M^{d-1}}\right)$. Therefore,

$$O(n^{v_1-\alpha e_1} n^{v-\alpha e}) = \exp\left(-\Omega\left(\frac{\ln n}{M^{d-1}}\right)\right) n^{v-\alpha e}.$$

Summing over $\exp\left(O(M^2)\right)$ possibilities for $(V(H), H_1)$, we obtain a bound of $o\left(\frac{n^{v-\alpha e}}{a}\right)$ on the number of $\vec{y}$ which violate the second part of Definition 5.7.

Case 2. (R, H_1) is not safe. We first show that in this case, $\vec{z}$ is a rigid extension of $\vec{x}$. Since (R, H_1) is not safe, from part 1 of Proposition B.1 it has a rigid subextension (R, H_2). The vertices of $\vec{y} \cup \vec{z}$ which form this (R, H_2)-extension of $\vec{x}$ are actually all in $\vec{z}$; we denote them by $\vec{z}'$. Indeed, by definition these vertices are contained in $U_{\vec{x}}$ (as they form a rigid extension of $\vec{x}$ with at most $v + t \leq 2M$ vertices) but $\vec{y}$ contains no vertices from $U_{\vec{x}}$. From part 2 of Proposition B.1 we deduce that $\vec{x} \cup \vec{y} \cup \vec{z}'$ is a rigid extension of $\vec{x} \cup \vec{y}$. It corresponds to a rigid subextension of $(V(H), H_1)$, but from minimality it must be $(V(H), H_1)$ itself, so $\vec{z}' = \vec{z}$. Therefore $\vec{z}$ is a rigid extension of $\vec{x}$.

Now, since $\vec{z}$ is a rigid extension of $\vec{x}$, its vertices are all in $U_{\vec{x}}$. Since $\vec{z}$ has at most t vertices and $|U_{\vec{x}}| = O(M^{d+1})$, there are $O\left(M^{(d+1)t}\right) = \exp\left(O(M \ln M)\right)$ possibilities for $\vec{z}$.

Fix a possible $\vec{z}$. Consider $\vec{y}$ as an extension of $\vec{x} \cup \vec{z}$. It does not have a rigid subextension, since $\vec{y}$ contains no vertex from $U_{\vec{x}}$, and from part 1 of Proposition B.1 it is safe. Let (R', H_1) denote the rooted graph corresponding to this extension. Its type is (v, e^+) with $e^+ > e$, because there is at least one edge between $\vec{y}$ and $\vec{z}$. Its number of roots is at most $r + t \leq 2M$. Therefore the number of (R', H')-extensions of $\vec{x} \cup \vec{z}$ is

$$O\left(n^v p^{e^+}\right) = O(n^{-\alpha}) n^{v-\alpha e}.$$

Summing over $\exp\left(O(M^2)\right)$ possibilities for the rooted graph (R', H_1) and $\exp\left(O(M \ln M)\right)$ possibilities for $\vec{z}$, we get a bound of $o\left(\frac{n^{v-\alpha e}}{a}\right)$ on the number of $\vec{y}$ which violate the second part of Definition 5.7. That completes the proof.

Acknowledgments. This is the full version of the extended abstract that appears in the proceedings of the Annual ACM/IEEE Symposium on Logic in Computer Science (LICS) 2024.

REFERENCES

[1] L. Babai, P. Erdős, and S. M. Selkow, *Random graph isomorphism*. ***SIAM Journal on Computing***, vol. 9 (1980), pp. 628–635.

[2] L. Babai and L. Kučera, *Canonical labeling of graphs in linear average time*, ***20th Annual IEEE Symposium on Foundations of Computer Science***, IEEE Computer Society Press, Silver Spring, Maryland, USA, 1979, pp. 39–46.

[3] A. Baker, ***Transcendental Number Theory***, Cambridge University Press, Cambridge, United Kingdom, 1975.

[4] I. Benjamini and M. Zhukovskii, *A very sharp threshold for first order logic distinguishability of random graphs*, ***Discrete Analysis***, 2025, Article 6, 17 pp. https://doi.org/10.19086/da.138190.

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press

[5] B. Bollobás, *The asymptotic number of unlabelled regular graphs*. ***Journal of the London Mathematical Society***, vols. 2–26 (1982), pp. 201–206.

[6] A. Ehrenfeucht, *An application of games to the completeness problem for formalized theories*. ***Fundamenta Mathematicae***, vol. 49 (1961), no. 2, pp. 129–141.

[7] R. Fagin, *Probabilities on finite models*. ***The Journal of Symbolic Logic***, vol. 41 (1976), no. 1, pp. 50–58.

[8] Y. V. Glebskii, D. I. Kogan, M. I. Liogon'kii, and V. A. Talanov, *Range and degree of realizability of formulas in the restricted predicate calculus*. ***Cybernetics***, vol. 5 (1969), no. 2, pp. 142–154.

[9] I. Hoshen and W. Samotij, *Simonovits's theorem in random graphs*. ***Proceedings of the London Mathematical Society*** (2026). Published online 13 October 2026.

[10] N. Immerman and E. Lander, *Describing graphs: A first-order approach to graph canonization*, In A. L. Selman (Ed.), ***Complexity Theory Retrospective*** (1990), pp. 59–81. Springer, New York, NY. https://doi.org/10.1007/978-1-4612-4478-3_5.

[11] S. Janson, T. Łuczak, and A. Ruciński, ***Random Graphs***, Wiley, New York, NY, USA, 2000.

[12] A. Khinchin, *Einige sätze über kettenbrüche, mit anwendungen auf die theorie der diophantischen approximationen*. ***Mathematische Annalen***, vol. 92 (1924), pp. 115–125.

[13] J. H. Kim, O. Pikhurko, J. Spencer, and O. Verbitsky, *How complex are random graphs in first order logic?* ***Random Structures & Algorithms***, vol. 26 (2005), nos. 1–2, pp. 119–145.

[14] J. H. Kim and H. V. Van, *Concentration of multivariate polynomials and its applications*. ***Combinatorica***, vol. 20 (2000), pp. 417–434.

[15] L. Leonid, ***Elements of Finite Model Theory***, Texts in Theoretical Computer Science. An EATCS Series, Springer, Berlin, Germany, 2004.

[16] B. D. McKay, *Subgraphs of dense random graphs with specified degrees*. ***Combinatorics, Probability and Computing***, vol. 20 (2011), no. 3, pp. 413–433.

[17] O. Pikhurko, H. Veith, and O. Verbitsky, *The first order definability of graphs: Upper bounds for quantifier depth*. ***Discrete Applied Mathematics***, vol. 154 (2006), no. 17, pp. 2511–2529.

[18] K. F. Roth, *Rational approximations to algebraic numbers*. ***Mathematika***, vol. 168 (1955), no. 2, pp. 1–20.

[19] A. Ruciński, *Matching and covering the vertices of a random graph by copies of a given graph*. ***Discrete Mathematics***, vol. 105 (1992), pp. 185–197.

[20] A. Ruciński and A. Vince, *Strongly balanced graphs and random graphs*. ***Journal of Graph Theory***, vol. 10 (1986), pp. 251–264.

[21] S. Shelah and J. Spencer, *Zero-one laws for sparse random graphs*. ***Journal of the American Mathematical Society***, vol. 1 (1988), pp. 97–115.

[22] J. Spencer, ***The Strange Logic of Random Graphs***, Springer, Berlin, Germany, 2001.

[23] J. Spencer and K. St. John, *The tenacity of zero-one laws*. ***Electronic Journal of Combinatorics***, vol. 8 (2001), no. 2, p. 14.

DEPARTMENT OF MATHEMATICS
CALIFORNIA INSTITUTE OF TECHNOLOGY
USA
E-mail: thershko@caltech.edu

SCHOOL OF COMPUTER SCIENCE
UNIVERSITY OF SHEFFIELD
UK
E-mail: m.zhukovskii@sheffield.ac.uk

https://doi.org/10.1017/jsl.2026.10271 Published online by Cambridge University Press