DIMACS Technical Report 97-35
September 1997

On the Limit Values of Probabilities for the First Order Properties of Graphs

by

Joel Spencer
Courant Institute
New York University
251 Mercer St.
New York NY 10012

Luboš Thoma[1]
DIMACS, P.O. Box 1179
Rutgers University
Piscataway, NJ 08855

[1]The second author gratefully acknowledges support as a DIMACS Postdoctoral Fellow.

DIMACS is a cooperative project of Rutgers University, Princeton University, AT&T Bell Laboratories and Bellcore.

DIMACS is an NSF Science and Technology Center, funded under contract STC–91–19999; and also receives support from the New Jersey Commission on Science and Technology.

ABSTRACT

Consider the random graph $\mathcal{G}(n,p)$, where $p = p(n)$ is any threshold function satisfying $p(n) = \Theta(\ln n/n)$. We give a full characterization of the limit values of probabilities of $\mathcal{G}(n,p)$ having a property ψ, where ψ is any sentence of the first order theory of graphs.

Contents

1 Introduction

Let $n \in \mathbb{N}$ and $0 \leq p = p(n) \leq 1$. The probability space $\mathcal{G}(n,p)$ is the usual probability space over the set of graphs on the vertex set $[n] = \{1, \ldots, n\}$ with the product measure $Prob(G) = p^{e_G}(1-p)^{\binom{n}{2}-e_G}$.

Let us define the first order language L of graphs. The language L consists of an alphabet containing a countable number of variable symbols $(x, y, z, ...)$ interpreted as vertices; logical symbols as Boolean connectives $\neg$, $\wedge$, $\vee$, quantifiers $\forall$, $\exists$, and brackets; the predicate of equality '=', and the adjacency '$\sim$' predicate, where $x \sim y$ is interpreted as $\{x, y\}$ being an edge. For $\psi \in L$, we write $Prob(\mathcal{G}(n,p) \models \psi)$ for the probability that a graph G in $\mathcal{G}(n,p)$ has the property ψ.

Our paper deals with the behaviour of $\mathcal{G}(n,p)$ for $p = p(n)$ being a threshold satisfying $p = \Theta(\ln n/n)$. Set

$$p_{k,l} = p_{k,l}(n; c) := \frac{\ln n}{kn} + \frac{l \ln \ln n}{kn} + \frac{c}{kn}\ , \tag{1}$$

where $k \geq 1, l \geq k-1$ are integers and $c \in \mathbb{R}$. In their 1991 paper, T. Łuczak and the first author proved that the zero - one law for $\mathcal{G}(n,p), p = \Theta(\ln n/n)$, and the language L (cf. Theorem 5 in [ŁS 91] for precise statement) holds except for p being "close" to $p_{k,l}$, in other words the probabilities $p_{k,l}$ are the only thresholds in the range $\Theta(\ln n/n)$.

As the general case is somewhat complicated we here describe our results when $k = 1$ and l is an arbitrary nonnegative integer. For a nonnegative integer s, let $\sigma_{l,s}$ be the following sentence of L : "there are *precisely* s vertices of degree *precisely* l". The number of vertices of degree l in $\mathcal{G}(n,p)$ possesses asymptotically Poisson distribution with mean $e^{-c}/l!$. Set $q_{l,s} := e^{-e^{-c}/l!}(e^{-c}/l!)^s/s!$ which is the limiting probability of $\mathcal{G}(n,p)$ satisfying $\sigma_{l,s}$.

Theorem 1 *Let $k = 1$, l be an arbitrary but fixed nonnegative integer, and $p = p_{k,l}(n; c)$. Let ψ be any sentence from the first order language of graphs L. Then $\mathcal{G}(n,p)$ satisfying ψ*

has a limiting probability (as $n \to \infty$) and the limit equals either a finite sum of the values $q_{l,s}$ defined above or one minus such a finite sum. [This includes zero and one with the null sum.] Furthermore, all such values are limiting probabilities of some first order sentence ψ. Further, the indices s of the finite sum are at most the quantifier rank of ψ.

The general characterization of limit probabilities is given in Section 6. While we give our results for $p(n) = p_{k,l}(n; c)$ we note that an additive $o(n^{-1})$ term in (1) will not affect any of the limit probabilities. This can be shown by checking that it is so for each of the axioms of the almost sure theories as well as their extensions we develop. We omit the calculations.

We are inspired by the results of Lynch [L 92] on the limit probabilities for first order sentences over $\mathcal{G}(n,p)$ where $p = c/n$ or $p = cn^{-(u+1)/u}, u \in \mathbb{N}$. Lynch showed that all such limit probabilities could be expressed as functions of c that were generated from addition, subtraction, multiplication, division, and base e exponentiation. In one important sense our range $p = \Theta(\frac{\ln n}{n})$ is considerably simpler than the case $p = c/n$. Our limit probabilities can and will include hyperexponential terms, e.g. $e^{-e^{-c}}$, but they are never a tower of exponentials of any higher length. In this manner our range is similar to the other range studied by Lynch: $p = cn^{-(u+1)/u}, u \in \mathbb{N}$.

The paper is organized as follows. The next section contains a necessary background on Ehrenfeucht games and complete theories. Section 3 deals with a general procedure for computing the limit values of probabilities.

Sections 4 and 5 contain different cases of the proof of Theorem 1. In Section 4 we treat the classical $[k = 1, l = 0]$ Erdős - Rényi case. We give there the main ideas used later in the complete proof. The case $k = 1$ and $l \geq 0$ is dealt with in Section 5. We give all the details of the probabilistic as well as logical parts of the proof. The statement and proof of a generalization of Theorem 1 for $p_{k,l}(n; c)$, which is somewhat technical, is given in Section 6.

Our asymptotics are always for k, l fixed and n approaching infinity. Notation: For two functions $f(n)$ and $g(n)$, $f(n) = \Theta(g(n))$ means that there are two constants c_1 and c_2 such that $c_1 g(n) \leq f(n) \leq c_2 g(n)$ for n large enough.

2 Ehrenfeucht Games, Complete Theories, and Countable Models

In this section we introduce the tool used later in the proof - Ehrenfeucht games. Throughout this section, let L' denote a first order language with a finite number of predicates.

We start with the (standard) definition of *quantifier rank*, $qr(\psi)$, of a formula $\psi \in L'$. For $\psi \in L'$, the quantifier rank of ψ is defined recursively as follows

$$\begin{aligned} qr(\psi) &= 0, \quad \text{for } \psi \text{ atomic} \\ qr(\neg\psi) &= qr(\psi) \\ qr(\psi \vee \varphi) &= \max\{qr(\psi), qr(\varphi)\} \\ qr((\exists\, x)\psi) &= qr(\psi) + 1\ . \end{aligned}$$

For an L'-model M we denote by $Th_t(L'; M)$ the set of all sentences of L' of quantifier rank at most t which hold for M, i.e.

$$Th_t(L'; M) = \{\psi \in L' \; ; \; M \models \psi \wedge qr(\psi) \leq t\} \, .$$

Ehrenfeucht games were introduced in [Eh 61]. The Ehrenfeucht game, $Ehr_t(M_1, M_2)$, is a t-round perfect information game between two players, played on two L'-models M_1 and M_2. In each round the first player (Spoiler) picks one of the two models M_1, M_2 and chooses one of its elements. The second player (Duplicator) must answer by selecting an element from the other model. For $i, 1 \leq i \leq t$, let a_i^1, a_i^2 denote the elements chosen by the players in the ith round of the game from M_1 and M_2, respectively. In the case the structures induced by $a_1^1, \ldots, a_t^1$ in M_1 and by $a_1^2, \ldots, a_t^2$ in M_2 are naturally isomorphic, the second player wins. We are going to use the following fact about Ehrenfeucht games (cf. [EF 95]).

Fact 2 *Let L' be a first order language with a finite number of predicates and let M_1, M_2 be two L'-models. The second player has a winning strategy for the game $Ehr_t(M_1, M_2)$ if and only if $Th_t(L'; M_1) = Th_t(L'; M_2)$.*

Two L'-models M_1, M_2 are called *elementarily equivalent* if and only if for every positive integer t : $Th_t(L'; M_1) = Th_t(L'; M_2)$. Let Σ be a theory. If Σ is not complete it has two extensions Σ_1, Σ_2 that have different truth value on some sentence ψ and by the Gödel Completeness Theorem there would be countable or finite models M_1, M_2 of those theories. Thus we obtain (cf. [BM 77]):

Fact 3 *Let Σ be a consistent first order theory which has no finite models. Then the following are equivalent.*

ι. *Σ is complete.*

ιι. *Any pair of countable models of Σ are elementarily equivalent.*

ιιι. *For any two countable models M_1, M_2 of Σ and any positive integer t, the second player has a winning strategy for the game $Ehr_t(M_1, M_2)$.*

Later we will be using the third condition, i.e. Ehrenfeucht games, to show a given theory is complete. We will need the following two facts about Ehrenfeucht games played on graphs. All graphs involved are considered as models of the first order language of graphs. The first fact describes a local sufficient condition for Duplicator to win the Ehrenfeucht game played on graphs.

Fact 4 *Let t be a positive integer. Let G_1 and G_2 be graphs satisfying the following property: for every $i = 1, \ldots, t$, and for every choice of $w \in V(G_1)$ and $v_1, \ldots, v_{i-1} \in V(G_2)$ there exists a vertex $v_i \in V(G_2)$ such that*

1. *the distance between v_i and v_α is at least 3^{t-i+1}, for every $\alpha = 1, \ldots, i-1$, and*
2. *the 3^{t-i} neighborhood of w is isomorphic to the 3^{t-i} neighborhood of v_i, with an isomorphism mapping w into v_i.*

Further, assume the above condition holds with roles of G_1 and G_2 reversed. Then Duplicator has a winning strategy for $Ehr_t(G_1, G_2)$.

Proof. Let r be a nonnegative integer and v a vertex of a graph. We define the rth (distance) neighborhood of v, denoted by $N(v;r)$, as the set of vertices of distance at most r from the vertex v.

We will denote by $a_1^1, \ldots, a_t^1$ and $a_1^2, \ldots, a_t^2$ elements chosen in G_1 and G_2, respectively, by the players during the game $Ehr_t(G_1, G_2)$.

We are going to show by induction that Duplicator has a winning strategy during which he can maintain that for every $i = 1, \ldots, t$, the neighborhoods $N(a_i^1; 3^{t-i})$ and $N(a_i^2; 3^{t-i})$ are isomorphic with the isomorphism mapping all a_j^1 to a_j^2 where $a_j^1 \in N(a_i^1; 3^{t-i})$. This implies that for every $i = 1, \ldots, t$, the structures induced on the sets $\{a_1^1, \ldots, a_i^1\}$ and $\{a_1^2, \ldots, a_i^2\}$ are isomorphic.

For $i = 1$, if Spoiler chose a vertex a_1^1 in G_1, say, using the assumptions of the fact Duplicator can find a vertex a_1^2 in G_2 such that the neighborhoods $N(a_1^1; 3^t)$ and $N(a_1^2; 3^t)$ are isomorphic with the isomorphism mapping a_1^1 to a_1^2.

Assume the induction hypothesis to hold for all $j = 1, \ldots, i-1$. Without loss of generality, suppose Spoiler chose its next move to be a vertex a_i^1 in G_1. In the case $a_i^1 \in \cup_{j=1}^{i-1} N(a_j^1; 2 \cdot 3^{t-i})$, Duplicator chooses a_i^2 according to the isomorphism between the neighborhoods $N(a_l^1; 3^{t-l})$ and $N(a_l^2; 3^{t-l})$ where the index l is such that $a_i^1 \in N(a_l^1; 2 \cdot 3^{t-i})$. As $N(a_i^1; 3^{t-i}) \subseteq N(a_l^1; 3^{t-l})$, we get that the neighborhoods $N(a_i^1; 3^{t-i})$ and $N(a_i^2; 3^{t-i})$ are isomorphic. Moreover, the isomorphisms of neighborhoods map a_j^1 to $a_j^2, 1 \leq j \leq i$. If $a_i^1 \notin \cup_{j=1}^{i-1} N(a_j^1; 2 \cdot 3^{t-i})$, Duplicator can choose a_i^2 according to the assumptions of the fact by setting $w = a_i^1$ and $v_1 = a_1^2, \ldots, v_{i-1} = a_{i-1}^2$. Thus, the neighborhoods $N(a_i^1; 3^{t-i})$ and $N(a_i^2; 3^{t-i})$ are again isomorphic with the isomorphism mapping a_i^1 to a_i^2. Further, note that $N(a_i^1; 3^{t-i}) \bigcap (\cup_{j=1}^{i-1} N(a_j^1; 3^{t-i})) = \emptyset$ and $N(a_i^2; 3^{t-i}) \bigcap (\cup_{j=1}^{i-1} N(a_j^2; 3^{t-i})) = \emptyset$. This finishes the induction step. $\blacksquare_4$

Fact 5 (cf. Proposition 1.3.9 in [EF 95]) *Let t be a positive integer and let G_1, G_2, and H be graphs. If Duplicator has a winning strategy for the game $Ehr_t(G_1, G_2)$, then he has also a winning strategy for $Ehr_t(G_1 + H, G_2 + H)$ where '+' is the disjoint union of graphs.*

Proof. We need to show Duplicator has a winning stategy for the game $Ehr_t(G_1 + H, G_2 + H)$. If Spoiler chooses a vertex in the graph G_1 or G_2 then Duplicator uses his winning strategy from $Ehr_t(G_1, G_2)$. If Spoiler chooses a vertex from either copy of H, Duplicator selects the identical vertex in the other copy. $\blacksquare_5$

3 Completions of Almost Sure Theories and Limit Probabilities

In this section we describe a general procedure for characterizing the limit values of probabilities. The main theorem of this section is Theorem 8. We start with assumptions for this section.

Let L' be a (first or second order) language. For each $n \in \mathbb{N}$ we are given a probability space (Ω_n, μ_n), where Ω_n is a finite set of models of L' and μ_n is a probabilistic measure on Ω_n. [In the following sections we set (Ω_n, μ_n) to be $\mathcal{G}(n,p)$.] We assume all subsets of Ω_n are μ_n-measurable. For $\psi \in L'$ we write $\mu_n(\psi)$ for $\mu_n(\{M \in \Omega_n : M \models \psi\})$, the probability ψ holds in a "random" model.

A theory T is called the *almost sure theory* for the sequence (Ω_n, μ_n) if it consists of those sentences $\psi \in L'$ for which $\lim_{n\to\infty} \mu_n(\psi) = 1$.

Remark 6 Observe that any almost sure theory T is consistent. This holds due to the finiteness of proof, i.e. "compactness" . If ψ is a theorem of T, i.e. $T \vdash \psi$, then it follows from a finite number of axioms and therefore $\lim_{n\to\infty} \mu_n(\psi) = 1$. Hence, all theorems of T hold almost surely and therefore the sentence "false" is not in T.

Let $\sigma \in L'$ be a sentence with positive limit probability, i.e. $\lim_{n\to\infty} \mu_n(\sigma) > 0$, and let T' be the extension of T by the new sentence σ. We claim that the theory T' is consistent. Similarly as above, if ψ is a theorem of T', we get $\liminf_{n\to\infty} \mu_n(\psi) \geq \lim_{n\to\infty} \mu_n(\sigma) > 0$. Therefore, the sentence "false" cannot be in T'.

Definition 7 *Let T be a subtheory for the almost sure theory. Let I be a finite or countable index set. A family σ_s, $s \in I$, of sentences of L is called a* complete family of completions *if*

ι. *For all $s \in I$ the extension of T by adding the new sentence σ_s is a complete theory. [Henceforth we write $T(s)$ for this extended theory.]*

$\iota\iota$. *For all $s \in I$*

$$a_s := \lim_{n\to\infty} \mu_n(\sigma_s)$$

exists and $0 < a_s < 1$.

$\iota\iota\iota$. *The σ_s are mutually exclusive in the sense that $T \vdash (\sigma_s \to \neg\sigma_{s'})$ for all distinct $s, s' \in I$.*

$\iota\iota\iota\iota$. $\sum_{s\in I} a_s = 1$.

For $\psi \in L'$, let $\mathfrak{M}[\psi] = \{s \in I; T(s) \vdash \psi\}$.

Theorem 8 *Given a complete family of completions $\sigma_s, s \in I$,*

$$\lim_{n\to\infty} \mu_n(\psi) = \sum_{s\in\mathfrak{M}[\psi]} a_s$$

for all $\psi \in L'$. In particular, the limiting probability always exists.

Remark 9 In general a complete family of completions implies convergence, i.e. that the limiting probabilities always exist. When combined with a characterization of the sets $\mathfrak{M}[\psi]$ a description of the limiting probabilities is achieved.

The assumptions of Definition 7 imply that the theory $\cap_{s\in I} T(s)$ is the almost sure theory for the sequence (Ω_n, μ_n).

Proof of Theorem 8. Let $\epsilon > 0$ be arbitrarily small. Pick $I_0 \subseteq I$ finite so that

$$\sum_{s \in I \setminus I_0} a_s < \epsilon$$

and set

$$MAIN := \vee_{s \in I_0} \sigma_s \ .$$

By $\iota\iota$, $\mu_n(MAIN) = \sum_{s \in I_0} \mu_n(\sigma_s)$ so, as I_0 is finite, $\lim_{n\to\infty} \mu_n(MAIN) = \sum_{s \in I_0} a_s$. Hence by $\iota\iota\iota$ we have $\lim_{n\to\infty} \mu_n(\neg MAIN) = \sum_{s \in I \setminus I_0} a_s$ and therefore for n sufficiently large $\mu_n(\neg MAIN) < \epsilon$. Set

$$\psi^* := \vee_{s \in I_0 \cap \mathfrak{M}[\psi]} \sigma_s$$

Then ψ, ψ^* are close in the sense that

$$(\psi \equiv \neg\psi^*) \Rightarrow \neg MAIN \bigvee \vee_{s \in I_0 \cap \mathfrak{M}[\psi]} (\sigma_s \wedge \neg\psi) \bigvee \vee_{s \in I_0 \setminus \mathfrak{M}[\psi]} (\sigma_s \wedge \psi) \ .$$

For $s \in I_0 \cap \mathfrak{M}[\psi]$, $\psi \in T(s)$ so $\mu_n(\sigma_s \wedge \neg\psi) \to 0$. For $s \in I_0, s \notin \mathfrak{M}[\psi]$, $\psi \notin T(s)$ so by completeness $\neg\psi \in T(s)$ and hence $\mu_n(\sigma_s \wedge \psi) \to 0$. As I_0 is finite for n sufficiently large

$$\mu_n(\psi \equiv \neg\psi^*) < \epsilon + \epsilon + \epsilon = 3\epsilon$$

so that

$$|\mu_n(\psi) - \mu_n(\psi^*)| < 3\epsilon \ .$$

But for n sufficiently large we also have

$$|\mu_n(\psi^*) - \sum_{s \in I_0 \cap \mathfrak{M}[\psi]} a_s| < \epsilon$$

so that

$$|\mu_n(\psi^*) - \sum_{s \in \mathfrak{M}[\psi]} a_s| < 2\epsilon$$

and thus

$$|\mu_n(\psi) - \sum_{s \in \mathfrak{M}[\psi]} a_s| < 5\epsilon$$

which, as ϵ was arbitrary, completes the proof. $\blacksquare_8$

4 The Erdős-Rényi Case

The classic work of Paul Erdős and Alfred Rényi [ER 60] contains many beautiful results. Surely one of the most striking is their description of the threshold for that most fundamental of graph concepts - connectivity. They showed that if

$$p = \frac{\ln n}{n} + \frac{c}{n} + o(\frac{1}{n}) \tag{3}$$

then

$$Prob(\mathcal{G}(n,p) \text{ connected}) = e^{-e^{-c}} + o(1) \ . \tag{4}$$

The precision of the result and the novelty of the double exponential have well stood the test of time.

In one sense this result is not directly pertinent to our current investigations as connectivity is not a first order property. But in their proof Erdős and Rényi actually examined the property of containing no isolated vertices, which we shall denote by σ_0 and can be written $(\forall x)(\exists y)(x \sim y)$ in the First Order world. They showed that σ_0 has the limit probability (4) and that connectivity and σ_0 have the same limit probability in this range of p.

We show that the double exponential is essentially the only kind of limit probability that can appear in the First Order world. Our results are given more formally and more generally in Sections 5 and 6, here we sketch the situation in this historically interesting case.

The first order sentences that hold almost surely for these p can be split into positive and negative statements. On the positive side:

- For every $m \geq 3$ and every $r \geq 1$: There are (at least) r cycles of length (precisely) m.
- For every $r \geq 1$ and every $d \geq 1$: There are (at least) r vertices of degree (precisely) d.

We think of small cycles and small degree vertices as special phenomenon. The above two schema say that they do occur, the final three schema say that they hold in isolation.

- For every $t, d \geq 1$: There are no vertices x_1, x_2 with degrees both at most d which are less than t apart.
- For every $t, d \geq 1$ and every $m \geq 3$: There is no vertex x of degree at most d which lies less than distance t from a cycle of length at most m.
- For every $t \geq 1$ and $m \geq 3$: There are no two cycles of length at most m which lie less than distance t apart.

Roughly speaking, the special phenomenon hold only polylogarithmically often and they are spread out in the graph so that locally one never sees two such phenomenon. Let T denote the theory given by these schema, which turns out to be the almost sure theory for p in this range. We prove this in the next section.

The countable models M of T are particularly nice. Because no two special phenomenon can be any finite distance apart any component of T can have only one of them. If a component C contains a vertex of finite degree d then it *must* be the infinite tree TR_d with all other vertices having infinite degree. Note TR_d is uniquely determined and note that M must have countably many copies of TR_d. Similarly if C contains a cycle of size m then it *must* be the unicyclic C_m generated from the cycle by generating a disjoint infinite rooted tree, all vertices of infinite degree, from each of the vertices of the cycle. Again C_m

is uniquely determined and M must have countably many copies of C_m. Let TR_ω denote the infinite tree with all vertices of infinite degree. The model M may have countably many copies of TR_ω, it may have any finite number of copies of TR_ω, in particular it may have no copies of TR_ω.

What about isolated vertices? Their number cannot be specified in the almost sure theory T since this range of p is precisely the threshold function for having an isolated vertex. We let σ_s denote the sentence: " There exist (precisely) s isolated vertices", defined for all nonnegative integers s including (especially) zero.

We claim that for any s the theory $T(s) = T + \sigma_s$ is complete. The countable models M_s of $T + \sigma_s$ are those models M of T that have precisely s isolated vertices. So $T(s)$ is not $\aleph_0$-categorical, as the models may or may not have copies of TR_ω. But it suffices to show that any two models M_s, M_s' of $T(s)$ are elementarily equivalent (i.e., have the same first order properties) for which it suffices to show that for any fixed t Duplicator can win the t-move Ehrenfeucht game on M_s, M_s'. To show that we are going to use Facts 4 and 5. First, we observe that due to Fact 5 it is enough to show that Duplicator can win the game $Ehr_t(\widehat{M}_s, \widehat{M}_s')$, where $\widehat{M}_s, \widehat{M}_s'$ denote the models M_s, M_s', respectively, after deleting the s isolated vertices. Let $i \in [t]$ and suppose vertices $w \in \widehat{M}_s$ and $v_1, \ldots, v_{i-1} \in \widehat{M}_s'$ are given. We show that there is a vertex $v_i \in \widehat{M}_s'$ satisfying the assumptions of Fact 4. Let U be the component of $\widehat{M}_s$ containing w. If U is not TR_ω, then there is an isomorphic copy of U in $\widehat{M}_s'$ such that this copy does not contain any of the vertices $v_1, \ldots, v_{i-1}$. This holds due to the fact that there are countably many of these components in M_s' and, thus, in $\widehat{M}_s{}'$. Hence, there is a vertex $v_i \in \widehat{M}_s'$ satisfying the assumptions of Fact 4. In the case U is a copy of TR_ω, we choose a copy U' of TR_1 in $\widehat{M}_s'$ such that U' does not contain any of the vertices $v_1, \ldots, v_{i-1}$. This is again possible due to the fact that M_s', and thus $\widehat{M}_s'$, contains a countable number of components isomorphic to TR_1. To satisfy the assumptions of Fact 4 we choose for v_i any vertex of U' which is at distance more than 3^t from the vertex of degree 1 in U'. The argument given above works also with the roles of $\widehat{M}_s$ and $\widehat{M}_s'$ reversed. Combining Facts 4 and 5 we get that Duplicator has a winning strategy for the game $Ehr_t(M_s, M_s')$.

Let $\overline{M}_s$ denote that model of $T + \sigma_s$ that has no copies of TR_ω, in some sense a minimal model. Then a sentence ϕ is a theorem of $T + \sigma_s$ if and only if ϕ holds in $\overline{M}_s$. Consider the t move Ehrenfeucht game on $\overline{M}_s, \overline{M}_{s'}$ with $s, s' \geq t$. Duplicator has a clear win since the only "nonisomorphism" that Spoiler can hope to exploit is in the number of isolated vertices but he has only t moves and every time he selects an isolated vertex Duplicator can likewise select an isolated vertex in the other model. Any given ϕ has some finite quantifier rank t and therefore ϕ holds in either all or none of the theories $T + \sigma_s$, $s \geq t$. Letting $\mathfrak{M}[\phi]$ denote the set of s for which $T + \sigma_s \models \phi$ we can rephrase this by saying $\mathfrak{M}[\phi]$ is either finite or cofinite.

What does all this say about the possible limit probabilities for a first order sentence ϕ? Let X be the number of isolated vertices in $\mathcal{G}(n,p)$. Then $\mathrm{Exp}[X] = n(1-p)^{n-1} \sim e^{-c}$ for p given by (3). Further it can be shown that X approaches the Poisson in distribution so that

$$Prob(\mathcal{G}(n,p) \models \sigma_s) = Prob(X = s) \to q_s := e^{-e^{-c}} \frac{(e^{-c})^s}{s!} \tag{5}$$

as $n \to \infty$. Note that $\sum_{s\geq 0} q_s = 1$, meaning roughly that the number of isolated vertices is almost surely bounded. This, and the careful epsilondeltamanship of Section 3, implies that any first order ϕ has limiting probability $\sum_{s\in\mathfrak{M}[\phi]} q_s$. Thus the possible limiting probabilities are finite sums of the q_s or one minus such a sum. Further, the σ_s are themselves first order sentences and for any finite subset S of nonnegatice integers the sentence $\phi = \vee_{s\in S}\sigma_s$ does have limiting probability $\sum_{s\in S} q_s$ while $\neg\phi$ has limiting probability one minus that – therefore we have a *complete characterization* of the possible limiting probabilities of first order sentences ϕ.

5 The Case of $k = 1$ and $l \geq 0$.

In this section we prove Theorem 1. We introduce the almost sure theories $T_{1,l}$ and their completions $T_{1,l}(s)$. At the end of the section, the proof of Theorem 1 is given. Solely for the purpose of this section we set $p = p_{1,l} = (\ln n + l\ \ln\ln n + c)/n$.

5.1 Axioms of the almost sure theories $T_{1,l}$

We start with defining axioms of the almost sure theories $T_{1,l}$, where $l \geq 0$ is an integer. For two vertices v, w in a graph, we use $dist(v, w)$ to denote the distance of v and w.

For a graph H, let $\varphi(H)$ be the statement that there is a subgraph isomorphic to the graph H, i.e. $(\exists\, x_1, \ldots, x_{v_H})\ (\wedge_{\{i,j\}\in E(H)}(x_i \sim x_j))$. The theories $T_{1,l}$ have the following seven schema, $\Lambda^1(.) - \Lambda^7(.)$:

- For every graph H with $v_H < e_H$: $\Lambda^1(H) \equiv \neg\varphi(H)$.
- For every graph H which is a tree or unicyclic connected: $\Lambda^2(H) \equiv \varphi(H)$.
- For all integers $t \geq 1$ and $m \geq 3$: $\Lambda^3(t, m) \equiv (\exists\ t$ distinct cycles $C_m^1, \ldots, C_m^t$ each on m vertices).
- For all integers $r, s, t \geq 0, s \geq r$: $\Lambda^4(r, s, t) \equiv \neg(\exists$ a pair of vertices $x, y)(deg(x) = r \wedge deg(y) = s \wedge dist(x, y) = t)$.
- For all integers $r, t \geq 0$ and $m \geq 3$: $\Lambda^5(t, r, m) \equiv \neg(\exists x, deg(x) = r)(\exists$ cycle C_m on m vertices) $(dist(x, C_m) = t)$.

and schema dealing with vertices of degree $r, r \geq 0$. The case $r = l$ is not included here as $p = p_{1,l}$ is the threshold function for the existence of vertices with degree equal to l.

- For all integers $r, 0 \leq r < l$: $\Lambda^6(r) \equiv \neg(\exists x)(deg(x) = r)$.
- For all integers $t \geq 1$ and $r > l$: $\Lambda^7(t, r) \equiv (\exists$ distinct vertices $x_1, \ldots, x_t)(deg(x_1) = \cdots = deg(x_t) = r)$.

Note that the group of axioms $\Lambda^6(.)$ is void for $l = 0$, i.e. for the theory $T_{1,0}$. In the following lemma, we show that all the schema of $T_{1,l}$ have the limiting probability equal to 1. In fact Lemma 12, proved later, and the results of Section 3 imply that the theories $T_{1,l}$ are the almost sure theories for $p = p_{1,l}$.

Lemma 10 *For every $l \geq 0$, all the axioms of $T_{1,l}$ are first order sentences and all of them hold for $\mathcal{G}(n,p)$ with probability $1 - o(1)$.*

Proof. Clearly, all the axioms $\Lambda^1(.) - \Lambda^7(.)$ are expressible in the first order language of graphs.

(1) The fact that $\Lambda^1(H)$ holds almost surely was originally proved by P. Erdős and A. Rényi in [ER 60]. For a graph H with $v_H < e_H$, let Y_H be a random variable counting subgraphs of $\mathcal{G}(n,p)$ isomorphic to H. Using Markov inequality, we get

$$Prob(Y_H > 0) \leq \text{ Exp}[Y_H] = \binom{n}{v_H} p^{e_H} = O(n^{v_H} p^{e_H}) = O\Big(\frac{(\ln n)^{e_H}}{n^{e_H - v_H}}\Big) = o(1) \ .$$

Thus, $\lim_{n\to\infty} Prob(\mathcal{G}(n,p) \models \Lambda^1(H)) = \lim_{n\to\infty} Prob(Y_H = 0) = 1$.

(2) For a graph H, let $m(H) = \max_{F\subseteq H} e_F/v_F$. It was proved by Bollobás that the function $n^{-1/m(H)}$ is a threshold for $\mathcal{G}(n,p) \models \varphi(H)$. Recall that $m(TR_t) = 1 - \frac{1}{t}$ for a tree TR_t on t vertices and $m(U) = 1$ for a unicyclic connected graph U. Thus, $\lim_{n\to\infty} Prob(\mathcal{G}(n,p) \models \Lambda^2(H)) = 1$.

(3) Let Y_m be a random variable counting the cycles C_m in $\mathcal{G}(n,p)$. We get for the moments of the random variable Y_m

$$\text{Exp}[Y_m] = \binom{n}{m} p^m = O((np)^m) = O(\ln^m n)$$

and

$$\text{Var}[Y_m] = \binom{n}{m}\binom{n-m}{m} p^{2m} + O\Big(\sum_{i=1}^{m} \binom{n}{i}\binom{n-i}{m-i}\binom{n-m}{m-i} p^{2m-i}\Big) - (\text{ Exp}[Y_m])^2 \ .$$

Thus, we have

$$\begin{aligned}
\frac{\text{Var}[Y_m]}{(\text{ Exp}[Y_m] - t)^2} &= (1+o(1)) \cdot \frac{\text{Var}[Y_m]}{(\text{ Exp}[Y_m])^2} \\
&= (1+o(1)) \cdot \Big[O\Big(\sum_{i=1}^{m} \frac{\binom{n}{i}\binom{n-i}{m-i}\binom{n-m}{m-i}}{\binom{n}{m}^2} \cdot p^{-i}\Big) + \Big(\frac{\binom{n}{m}\binom{n-m}{m}p^{2m}}{\binom{n}{m}^2 p^{2m}} - 1\Big)\Big] \\
&= O\Big(\sum_{i=1}^{m} \frac{1}{(np)^i}\Big) + o(1) \\
&= o(1) \ .
\end{aligned}$$

We get via the Chebyshev inequality $Prob(Y_m < t) \leq \text{ Var}[Y_m]/(\text{ Exp}[Y_m] - t)^2 = o(1)$ and, hence, $Prob(\mathcal{G}(n,p) \models \Lambda^3(t,m)) = 1 - Prob(Y_m < t) = 1 - o(1)$.

(4) Let Y be a random variable counting pairs of vertices x, y satisfying $deg(x) = r \wedge deg(y) = s \wedge dist(x,y) = t$. Computing the expectation of Y, one gets

$$\begin{aligned}
\mathrm{Exp}[Y] &= O\Big(\binom{n}{2}\binom{n-2}{t-1}\binom{n-t-1}{r-1}\binom{n-t-r}{s-1}p^{s-1+r-1+t}(1-p)^{2(n-s+1-r+1-t-1)}\Big) \\
&= O\Big(n(np)^{s+r+t-2}e^{-2np}\Big) \\
&= o(1)\ .
\end{aligned}$$

By Markov inequality, $Prob(Y > 0) \leq \mathrm{Exp}[Y] = o(1)$. Hence, $Prob(\mathcal{G}(n,p) \models \Lambda^4(r,s,t)) = 1 - o(1)$.

(5) Let Y be a random variable counting pairs (x, C_m) where x is a vertex of degree r, C_m is a cycle on m vertices, and $dist(x, C_m) = t$. Computing the expectation of Y, we get

$$\begin{aligned}
\mathrm{Exp}[Y] &= O\Big(n\binom{n-1}{r-1}p^{r-1}(1-p)^{n-t+1-m-r+1}\binom{n-r}{m}m!\cdot m\cdot p^m\binom{n-r-m}{t-1}p^t\Big) \\
&= O\Big((np)^{r+m+t-1}e^{-np}\Big) \\
&= o(1)\ .
\end{aligned}$$

By Markov inequality, $Prob(Y > 0) \leq \mathrm{Exp}[Y] = o(1)$. Therefore, $\lim_{n\to\infty} Prob(\mathcal{G}(n,p) \models \Lambda^5(t,r,m)) = 1$.

(6) Let Y_r be a random variable counting the vertices of $\mathcal{G}(n,p)$ with degree equal to r. We are going to show $Prob(Y_r > 0) = o(1)$, for $r < l$, and $Prob(Y_r \geq t) = 1 - o(1)$, for $r > l$. We start with computing the expectation of Y_r

$$\mathrm{Exp}[Y_r] = n\binom{n-1}{r}p^r(1-p)^{n-r-1} = O\Big(n^{r+1}\frac{\ln^r n}{n^r}e^{-\ln n - l\ln\ln n}\Big) = O\Big(\ln^{r-l} n\Big)\ . \tag{6}$$

In the case $r < l$, we get by Markov inequality

$$Prob(Y_r > 0) \leq \mathrm{Exp}[Y_r] = O(\ln^{r-l} n) = o(1)$$

and, thus, $Prob(\mathcal{G}(n,p) \models \Lambda^6(r)) = 1 - o(1)$.

For $r > l$, we are going to use Chebyschev inequality. We have $Prob(Y_r < t) \leq Prob(|Y_r - \mathrm{Exp}[Y_r]| \geq \mathrm{Exp}[Y_r] - t) \leq \mathrm{Var}[Y_r]/(\mathrm{Exp}[Y_r] - t)^2$. Thus, we need to bound $\mathrm{Var}[Y_r]$ from above. One gets for the second moment of the random variable Y_r

$$\begin{aligned}
\mathrm{Var}[Y_r] &= 2\ \mathrm{Exp}[\binom{Y_r}{2}] + \ \mathrm{Exp}[Y_r] - (\ \mathrm{Exp}[Y_r])^2 \\
&= 2\binom{n}{2}\Big[\sum_{i=0}^{r}\binom{n-2}{i}\binom{n-2-i}{r-i}\binom{n-r-2}{r-i}p^{2r}(1-p)^{2n-2r-2} \\
&+ \sum_{i=0}^{r-1}\binom{n-2}{i}\binom{n-2-i}{r-i-1}\binom{n-r-1}{r-i-1}p^{2r-1}(1-p)^{2n-2r-2}\Big] \\
&+ \ \mathrm{Exp}[Y_r] - (\ \mathrm{Exp}[Y_r])^2 \\
&= \ \mathrm{Exp}[Y_r] + \Big[2\binom{n}{2}\binom{n-2}{r}\binom{n-r-2}{r}p^{2r}(1-p)^{2n-2r-2} - (\ \mathrm{Exp}[Y_r])^2\Big] \\
&+ 2\binom{n}{2}p^{2r}(1-p)^{2n-2r-2}\Big[\sum_{i=1}^{r}\binom{n-2}{i}\binom{n-2-i}{r-i}\binom{n-r-2}{r-i} \\
&+ \sum_{i=0}^{r-1}\binom{n-2}{i}\binom{n-2-i}{r-i-1}\binom{n-r-1}{r-i-1}p^{-1}\Big] .
\end{aligned}$$

Hence,

$$\begin{aligned}
\frac{\mathrm{Var}[Y_r]}{(\ \mathrm{Exp}[Y_r]-t)^2} &= (1+o(1))\cdot\frac{\mathrm{Var}[Y_r]}{(\ \mathrm{Exp}[Y_r])^2} \\
&= (1+o(1))\Big[\frac{1}{\mathrm{Exp}[Y_r]} + \Big(\frac{2\binom{n}{2}\binom{n-2}{r}\binom{n-r-2}{r}}{n^2\binom{n-1}{r}^2} - 1\Big) \\
&+ \frac{2\binom{n}{2}}{n^2\binom{n-1}{r}^2}\cdot\Big(\sum_{i=1}^{r}\binom{n-2}{i}\binom{n-2-i}{r-i}\binom{n-r-2}{r-i} \\
&+ \sum_{i=0}^{r-1}\binom{n-2}{i}\binom{n-2-i}{r-i-1}\binom{n-r-1}{r-i-1}p^{-1}\Big)\Big] \\
&= (1+o(1))\Big[\frac{1}{\mathrm{Exp}[Y_r]} + o(1) + \frac{1}{n^{2r}}\Big(O(n^{2r-1}) + O(\frac{n^{2r-1}}{\ln n})\Big)\Big] \\
&= o(1)\ .
\end{aligned}$$

Therefore, we get $Prob(\mathcal{G}(n,p) \models \Lambda^7(l,r)) = Prob(Y_r \geq t) = 1 - o(1)$. $\blacksquare_{10}$

5.2 Complete theories $T_{1,l}(s)$

For fixed integers $l, s \geq 0$, let $T_{1,l}(s)$ be the extension of the theory $T_{1,l}$ by the axiom $\sigma_{l,s}$. In this section, we are going to show that these theories are complete. To be able to employ the technique of Ehrenfeucht games, we need the following description of countable models for the theory $T_{1,l}(s)$. Note that due to the schema $\Lambda^7(.)$, the theories $T_{1,l}(s)$ do not possess finite models.

Lemma 11 *Let $l, s \geq 0$ be integers and let M be a countable model for the theory $T_{1,l}(s)$. Then M contains*

I. in the case $l = 0$, precisely s finite components each of them being an isolated vertex,

in the case $l > 0$, precisely s infinite trees, each with precisely one vertex of finite degree equal to l and with all other degrees infinite,

and every other component of M is one of the following three types

II. an infinite tree with precisely one vertex of finite degree bigger than l and with all other degrees infinite,

III. an infinite tree with all vertices of infinite degree,

IV. an infinite unicyclic connected component with all vertices of infinite degree, i.e. an infinite component consisting of a finite cycle and infinite rooted trees from each vertex of the cycle having all degrees infinite. These infinite rooted trees are pairwise vertex disjoint.

Moreover, for every integer $r > l$, M must contain an infinite number of trees of type II with a vertex of degree r and, for every (finite) cycle C of length m, the model M must contain an infinite number of components of type IV containing C. Furthermore, the number of component of type III can be either infinite or any nonnegative integer, including 0.

Proof. The structure of a countable model M of $T_{1,l}(s)$ is determined by the axioms $\Lambda^1(.), \ldots, \Lambda^7(.)$, and $\sigma_{l,s}$ of the theory.

The tree-like or unicyclic structure of components of M is implied by axioms $\Lambda^1(.)$ and $\Lambda^2(.)$. Moreover, axioms $\Lambda^4(.)$ and $\Lambda^5(.)$ imply that no two special phenomena - a vertex of a finite degree or a finite cycle - are in the same component of M.

The existence and number of components of type I is determined by the axiom $\sigma_{l,s}$. Finally, the existence and number of components of types II and IV is implied by axioms $\Lambda^3(.)$, $\Lambda^6(.)$, and $\Lambda^7(.)$. Note that the axioms of the theory $T_{1,l}(s)$ do not determine neither the number nor the existence of components of type III. $\blacksquare_{11}$

Lemma 12 *For all integers $l, s \geq 0$, the theories $T_{1,l}(s)$ are complete.*

Proof. Fix $l, s \geq 0$. In the proof we use T to denote the theory $T_{1,l}(s)$. Note that T is a consistent theory as it is an extension of the theory $T_{1,l}$ by the axiom $\sigma_{l,s}$ with nonzero limiting probability (cf. Remark 6).

To show T is complete, we are going to use the condition $\iota\iota$ of Fact 3. Thus the following claim already implies the completeness of T.

Claim 13 *For every two countable models M_1, M_2 of the theory T and for every positive integer t, the second player has a winning strategy for the Ehrenfeucht game $Ehr_t(M_1, M_2)$.*

Proof. The proof is based on Facts 4 and 5. When using Fact 4 we maintain the notation introduced there. Due to Fact 5 it is enough to show that Duplicator can win the game $Ehr_t(\widehat{M}_1, \widehat{M}_2)$, where $\widehat{M}_1, \widehat{M}_2$ denote the models M_1, M_2, respectively, after deleting all the components of type I.

Let $i \in [t]$ and suppose, without loss of generality, vertices $w \in \widehat{M}_1$ and $v_1, \ldots, v_{i-1} \in \widehat{M}_2$ are given. We show that there is a vertex $v_i \in \widehat{M}_2$ satisfying the assumptions of Fact 4. Let U be the component of $\widehat{M}_1$ containing the vertex w. In the case the component U is of the type II or IV, we can always find a component U' of the same type, i.e. isomorphic, in $\widehat{M}_2$ such that none of the vertices $v_1, \ldots, v_{i-1}$ belongs to U'. Hence, we can choose the vertex v_i to satisfy the conditions of Fact 4. If U is a component of type III, let U' be a component of $\widehat{M}_2$ isomorphic to the tree TR_{l+1}, i.e. the tree with precisely one vertex of finite degree $l+1$ and all other degrees infinite, and such that U' does not contain any of $v_1, \ldots, v_{i-1}$. We choose for the vertex v_i any vertex of U' of distance more than 3^t from the vertex of finite degree $l+1$. The argument given above works also with the roles of $\widehat{M}_1$ and $\widehat{M}_2$ reversed. Thus Fact 4 implies that Duplicator has a winning strategy for $Ehr_t(\widehat{M}_1, \widehat{M}_2)$ and our claim follows. $\blacksquare_{13}$

This also completes the proof of Lemma 12. $\blacksquare_{12}$

Remark 14 Fix $l, s \geq 0$. Note that $T_{1,l}(s)$ is not $\aleph_0$-categorical. However, the proof of Lemma 12 implies that there is a particular model $\overline{M}$ of $T_{1,l}(s)$ not containing any trees of type III (cf. Lemma 11) such that for every t and every model M of $T_{1,l}(s)$ one has

$$Th_t(L; \overline{M}) = Th_t(L; M) \ .$$

This $\overline{M}$ is in some sense a minimal model of $T_{1,l}(s)$. Later, we are going to refer to this "bar" notation.

To justify that $\sigma_{l,s}$ form a complete family of completions and, also, to determine the possible limit values, we need the following fact giving the limit probabilities $q_{l,s}$. The proof of it is standard and can be found for example in [B 85].

Fact 15 *For every $l \geq 0$, let $X_l(n)$ be a random variable counting the vertices of degree precisely l in $\mathcal{G}(n, p_{1,l})$. Then, for every integer $l \geq 0$, the variables $X_l(n)$ converge in distribution, as $n \to \infty$, to a random variable with the Poisson distribution and mean equal to $e^{-c}/l!$. Thus,*

$$q_{l,s} = \lim_{n\to\infty} Prob(\mathcal{G}(n, p_{1,l}) \models \sigma_{l,s}) = \lim_{n\to\infty} Prob(X_l(n) = s) = \frac{e^{-sc}}{(l!)^s s!} \cdot exp\Big(-\frac{e^{-c}}{l!}\Big) \ .$$

5.3 Proof of Theorem 1

We can now quickly prove Theorem 1. The results of Section 3 give the statement of Theorem 1 except that we still need to characterize the sets $\mathfrak{M}[\psi]$. The following shows that all $\mathfrak{M}[\psi]$ are finite or cofinite.

Claim 16 *Let $t \geq 0$ be an integer. For all $s, s' \geq t$, let M_s be a model for $T_{1,l}(s)$ and $M_{s'}$ be a model for $T_{1,l}(s')$. Then, $Th_t(L; M_s) = Th_t(L; M_{s'})$.*

Proof. In view of Remark 14 after the proof of Lemma 12, instead of the models M_s and $M_{s'}$ we can equivalently use models $\overline{M}_s$ and $\overline{M}_{s'}$. Both of the models $\overline{M}_s$ and $\overline{M}_{s'}$ do not contain trees of type III. (cf. Lemma 11). Thus, after deleting all the components of type I (i.e. all isolated vertices in the case $l = 0$ or all infinite trees containing a vertex of degree l in the case $l > 0$) the remaining parts of the models $\overline{M}_s$ and $\overline{M}_{s'}$ are *isomorphic*. Due to the assumption that the number of the components of type I in both models is at least t, the second player can win $Ehr_t(\overline{M}_s, \overline{M}_{s'})$. Due to the structure of the models $\overline{M}_s$ and $\overline{M}_{s'}$ described above, Duplicator can win by using a rather straightforward strategy - he just precisely copies the moves of the Spoiler. Hence, the claim follows. $\blacksquare_{16}$

The claim implies the finite/co-finite part of the statement in Theorem 1. Moreover, for a given $\psi \in L$ we can give a bound on the size of $\mathfrak{M}[\psi]$ or of its complement, $\mathfrak{M}^C[\psi] = (\mathbb{N} \cup \{0\}) \setminus \mathfrak{M}[\psi]$, depending on which of these two sets is finite. To see that, let $qr(\psi) = h$. Recall that the theories $T_{1,l}(s)$ are complete, i.e. that either $T_{1,l}(s) \vdash \psi$ or $T_{1,l}(s) \vdash \neg\psi$. We need to distinguish the following two cases.

In the case $T_{1,l}(h) \vdash \neg\psi$, Claim 16 implies that for every $s \geq h$ one has $T_{1,l}(s) \vdash \neg\psi$. Thus, $\mathfrak{M}[\psi] \subseteq \{0, 1, \ldots, h-1\}$. In the other case, $T_{1,l}(h) \vdash \psi$, we define $\varphi \equiv \neg\psi$. Hence, $\mathfrak{M}^C[\psi] = \mathfrak{M}[\varphi] \subseteq \{0, 1, \ldots, h-1\}$.

To finish the proof of Theorem 1, i.e. to get the full characterization, we need to show witnessing sentences in the language L. For $S \subset \mathbb{N} \cup \{0\}$ finite, let $\psi_S \equiv \vee_{s \in S} \sigma_{l,s}$. In the case $S \subseteq \mathbb{N} \cup \{0\}$ is co-finite, we set $\psi_S \equiv \neg \vee_{s \in (\mathbb{N} \cup \{0\}) \setminus S} \sigma_{l,s}$.

6 The General Case

In this section we give a generalization of Theorem 1 for $p = p_{k,l} = (\ln n + l \, \ln\ln n + c)/(kn)$, though omitting some of the more detailed calculations. The results are summarized in Theorem 22.

The section is organized similar to Section 5. We start with describing schema for almost sure theories $T_{k,l}$; then we describe the completions of these - $T_{k,l}(\mathbf{r})$. Finally, we give a description of sets $\mathfrak{M}[\psi]$, ψ being a first order sentence of graphs.

We begin with a special notation. Let $\mathbf{d} = (d_1, \ldots, d_{k'})$ with all $d_i \geq 0$ integers. Let TR be a tree on $\{1, \ldots, h\}$ with $h \geq k'$ and with all leaves of TR in $\{1, \ldots, k'\}$. We say $(TR, \mathbf{d})$ has type (k', l') where

$$l' = h - 1 + \sum_{i=1}^{k'} d_i \, .$$

We call $(TR, \mathbf{d})$ a *low-degree tree* and, for $r \geq 1$, define the first order sentence expressing that there exist (at least) r such low degree trees

> $LDT(TR, \mathbf{d}, r)$: There exist distinct $x_1^{(u)}, \ldots, x_h^{(u)}$, $1 \leq u \leq r$, with $x_i^{(u)} \sim x_j^{(u)}$ if and only if $\{i, j\} \in E(TR)$ [i.e., the x_i's form r copies of TR] and $x_i^{(u)}, 1 \leq i \leq k'$, is adjacent to precisely d_i vertices other than the $x_1^{(u)}, \ldots, x_h^{(u)}$.

In the case $r = 1$, we write $LDT(TR, \mathbf{d})$ for $LDT(TR, \mathbf{d}, 1)$.

Remark 17 $LDT(TR, \mathbf{d})$ gives the existence of k' low degree vertices in close proximity. When $l' = k' - 1$, the minimal possible value, we must have $h = k'$ and all $d_i = 0$ and then $LDT(TR, \mathbf{d})$ means there exists a component TR. When $h = k' = 1$ and $\mathbf{d} = (d_1)$ the sentence $LDT(TR, \mathbf{d})$ means there exists a vertex of degree precisely d_1. When, for example, $k' = 2, h = 3$, $TR = \{\{1,3\},\{2,3\}\}$ and $\mathbf{d} = (5, 7)$ the sentence $LDT(TR, \mathbf{d})$ means there exist two vertices of (total) degrees 6 and 8 that are at distance two. Note that no condition is placed on the degree of the vertex between them.

The theories $T_{k,l}$ have the following common schema:

- There are no bicyclic subgraphs.
- For every $r \geq 1$ there are (at least) r copies of every tree and unicyclic graph.
- For every $d \geq 0$, $t \geq 0$, $m \geq 3$ there is no vertex of degree d at distance t from a vertex on a cycle of size m.

With k given the following common schema hold in the theories $T_{k,l}$ for all $l \geq k - 1$:

- For $(TR, \mathbf{d})$ of type $(k + 1, l')$, any l': $\neg LDT(TR, \mathbf{d})$.
- For $(TR, \mathbf{d})$ of type (k', l') with $0 < k' < k$, any l', and any r : $LDT(TR, \mathbf{d}, r)$.

With k, l fixed we add:

- For $(TR, \mathbf{d})$ of type (k, l') with $l' < l$: $\neg LDT(TR, \mathbf{d})$.
- For $(TR, \mathbf{d})$ of type (k, l') with $l' > l$, any r: $LDT(TR, \mathbf{d}, r)$.

We are going to show that all the axioms of the theories $T_{k,l}$ hold with probability $1 - o(1)$ which combined with Lemma 19 and the results of Section 3 implies that $T_{k,l}$ are almost sure theories. Identically as in Section 5, we get that every sentence in the first three schema holds with probability $1 - o(1)$.

Let $(TR, \mathbf{d})$ have type (k', l') and set $X = X_{(TR,\mathbf{d})}$ equal the number of choices of $x_1, \ldots, x_h$ having the $LDT(TR, \mathbf{d})$ property. Neglecting constants

$$\mathrm{Exp}[X] = \Theta\left(n^h p^{h-1} \prod_{i=1}^{k'} \binom{n}{d_i} p^{d_i} (1-p)^n\right) = \Theta\left(n (\ln n)^{l'} e^{-pnk'}\right) .$$

Neglecting polylogarithmic terms we get $\mathrm{Exp}[X] = n^{1-(k'/k)+o(1)}$. If $k' > k$ then $\mathrm{Exp}[X] \rightarrow 0$ while if $k' < k$ then $\mathrm{Exp}[X] \rightarrow \infty$. In the latter case a variance calculation, which is omitted, gives that almost surely the random variable X is close to $\mathrm{Exp}[X]$ and, thus, $LDT(TR, \mathbf{d}, r)$ almost surely for any r. In the case $k' = k$ the powers of n cancel and $\mathrm{Exp}[X] = \Theta((\ln n)^{l'-l})$. If $l' < l$ then $\mathrm{Exp}[X] \rightarrow 0$ while if $l' > l$ then $\mathrm{Exp}[X] \rightarrow \infty$. When $\mathrm{Exp}[X] \rightarrow 0$ almost surely $\neg LDT(TR, \mathbf{d})$ while when $\mathrm{Exp}[X] \rightarrow \infty$ a variance calculation, which we omit, shows $X \sim \text{Exp}[X]$ almost surely so that $LDT(TR, \mathbf{d}, r)$ almost surely for any r.

There are, critically, only a finite number of $(TR, \mathbf{d})$ of a given type (k, l). We call them *equivalent* if the sentences $LDT(TR, \mathbf{d})$ are tautologically equivalent. [For example, $k = h = 2$, TR an edge on $\{1, 2\}$, $\mathbf{d} = (5, 7)$ and $\mathbf{d} = (7, 5)$ are equivalent.] Let $\beta = \beta(k, l)$ be the number of equivalence classes and let $(TR_\gamma, \mathbf{d}_\gamma)$, $1 \le \gamma \le \beta$, be a list containing precisely one $(TR, \mathbf{d})$ from each equivalence class. Let $\mathbf{r} = (r_1, \dots, r_\beta)$ with $r_\gamma \ge 0$ arbitrary integers. Define

$$\sigma_{\mathbf{r}} := \wedge_\gamma \left[LDT(TR_\gamma, \mathbf{d}_\gamma, r_\gamma) \wedge \neg LDT(TR_\gamma, \mathbf{d}_\gamma, r_\gamma + 1) \right] .$$

(When $r_\gamma = 0$ the bracketed expression is to be replaced by simply $\neg LDT(TR_\gamma, \mathbf{d}_\gamma)$.)

We claim that the family of $\sigma_{\mathbf{r}}$ forms a complete family of completions in the sense of Section 3. Our index set will be

$$I = \{(r_1, \dots, r_\beta) : r_\gamma \ge 0 \text{ for all } 1 \le \gamma \le \beta\} .$$

Note that, for example, due to the second schema of $T_{k,l}$ the theories $T_{k,l}$ as well as $T_{k,l}(\mathbf{r})$ do not possess finite models.

Lemma 18 *Let k and $l, l \ge k - 1 \ge 0$, be integers and $\mathbf{r} \in I$. Every countable model M of the theory $T_{k,l}(\mathbf{r})$ contains*

A. for every low degree tree $(TR_\gamma, \mathbf{d}_\gamma), \gamma = 1, \dots, \beta$, precisely r_γ components containing this tree. Moreover, if $l = k - 1$, all these componets are finite and consist of a single copy of the tree $(TR_\gamma, \mathbf{d}_\gamma)$. In the case $l > k - 1$, all these components are infinite with all vertices of infinite degree except vertices $x_1, \dots, x_k$ in the copy of the low degree tree $(TR_\gamma, \mathbf{d}_\gamma)$.

Every other component of the model M is one of the following four types:

B. an infinite tree containing a copy of a tree $(TR, \mathbf{d})$ of type $(k, l'), l' > l$, with all vertices of infinite degree except the vertices $x_1, \dots, x_k$ in the copy of $(TR, \mathbf{d})$,

C. a finite tree on k' vertices where $0 < k' < k$,

D. an infinite tree with all but at most $k - 1$ vertices of infinite degree,

E. an infinite unicyclic connected component with all vertices of infinite degree, i.e. an infinite component consisting of a finite cycle and infinite rooted trees from each vertex of the cycle having all degrees infinite. These infinite rooted trees are pairwise vertex disjoint.

The model M must contain an infinite number of components of type B for every low degree tree $(TR, \mathbf{d})$ of type $(k, l'), l' > l$. Moreover, it must contain an infinite number of the finite components of type C for every tree $(TR, \mathbf{d})$ of type $(k', k' - 1), 0 < k' < k$. Also, for every integer $m \ge 3$ and a cycle on m vertices the model M must contain an infinite number of components of type E containing that cycle. Finally, the number of components of type D can be either infinite or any nonnegative number, including 0.

Proof. The axioms of the theory $T_{k,l}$ and the sentence $\sigma_{\mathbf{r}}$ determine the structure of the model M. The first schema of $T_{k,l}$ implies that every component of M has a tree-like or unicyclic structure. Moreover, using the third and fourth schema of $T_{k,l}$ we infere that there are at most k vertices of finite degree in a given component and that there is no component with both a finite cycle and a vertex of finite degree.

The existence and number of components of type A is given by the axiom $\sigma_{\mathbf{r}}$. Because there are at most k vertices of finite degree in each component of M, every copy of $(TR_\gamma, \mathbf{d}_\gamma), 1 \le \gamma \le \beta$, must lie in a separate component. If $l = k-1$, the low degree tree $(TR_\gamma, \mathbf{d}_\gamma)$ contains necessary k vertices and all $d_i, i = 1, \ldots, k$, equal zero. Thus, every component containing a copy of $(TR_\gamma, \mathbf{d}_\gamma)$ consists of a single copy of the tree $(TR_\gamma, \mathbf{d}_\gamma)$. In the case $l > k-1$, there is either more than k vertices in $(TR_\gamma, \mathbf{d}_\gamma)$ and/or a vertex joined to some of the vertices $x_1, \ldots, x_k$ of $(TR_\gamma, \mathbf{d}_\gamma)$. Thus, there must be a vertex, say v, which degree is not determined by the d_i's. As there cannot be more than k vertices of finite degree in a component, the vertex v must have an infinite degree. Hence, the component is infinite.

The last schema of $T_{k,l}$ implies that there is an infinite number of components of type B. Moreover, as $l' > l \ge k-1$, such a component must contain a vertex which does not have a degree determined by any of $d_i, i = 1, \ldots, k$. Similarly as above, due to the fact that there cannot be $k+1$ vertices of finite degree in a component, we conclude that such a vertex must have an infinite degree. This implies that the component itself is infinite.

If a component of M contains $k', k' < k$ vertices of finite degree which form a low degree tree $(TR, \mathbf{d})$ of type $(k', k'-1)$, then TR has necessary k' vertices and all $d_i, i = 1, \ldots, k'$, equal zero. Thus, such a component consists only of a single copy of the tree $(TR, \mathbf{d})$. Moreover, the fifth schema of $T_{k,l}$ implies that there is an infinite number of such components.

The existence and number of components containing none or $k', 0 < k' < k$, vertices of finite degree which in the latter case form a low degree tree of type $(k', l'), l' > k'-1$, is not determined by the theory $T_{k,l}(\mathbf{r})$.

The second schema of $T_{k,l}$ implies the infinite number of unicyclic components.

Finally, the axioms $\neg LDT(TR, \mathbf{d})$ for any low degree tree $(TR, \mathbf{d})$ of type $(k, l'), l' < l$, or $(k+1, l')$, any $l' \ge k$, and the nonexistence of bicyclic subgraphs imply that all the components of the model M must be one of the types A - E. $\blacksquare_{18}$

Lemma 19 *Let k and $l, l \ge k-1 \ge 0$, be integers. For every $\mathbf{r} \in I$, the theory $T_{k,l}(\mathbf{r}) := T_{k,l} + \sigma_{\mathbf{r}}$ is complete.*

Proof. The theory $T_{k,l}(\mathbf{r})$ is consistent by Remark 6. To show it is complete we are going to use the condition $\iota\iota\iota$ of Fact 3. We use a notation for components of countable models of the theory $T_{k,l}(\mathbf{r})$ introduced in Lemma 18.

Let M_1, M_2 be two countable models for $T_{k,l}(\mathbf{r})$ and consider the t-move Ehrenfeucht game $Ehr_t(M_1, M_2)$. To show Duplicator has a winning strategy for this game we are going to use Facts 4 and 5. First, we observe that due to Fact 5 it is enough to show that Duplicator can win the game $Ehr_t(\widehat{M}_1, \widehat{M}_2)$, where $\widehat{M}_1, \widehat{M}_2$ denote the models M_1, M_2, respectively, after deleting all components of type A. Let $i \in [t]$ and suppose vertices $w \in \widehat{M}_1$ and $v_1, \ldots, v_{i-1} \in \widehat{M}_2$ are given. We show that there is a vertex $v_i \in \widehat{M}_2$ satisfying the assumptions of Fact 4. Let U be the component of $\widehat{M}_1$ containing w. If U is a component of type B, C, or E, there must be a component U' in $\widehat{M}_2$ isomorphic to U such that none of the vertices $v_1, \ldots, v_{i-1}$ belongs to U'. This holds due to the fact that M_2, and thus $\widehat{M}_2$, contains a countable number of such components. Hence, there is a vertex $v_i \in \widehat{M}_2$ satisfying the assumptions of Fact 4. Thus, suppose U is a component of type D, i.e. U is an infinite tree containing $k', 0 \leq k' < k$, vertices of finite degree. Set $a = 3^{t-i}$ in Fact 4. The a-neighborhood of the vertex w, $N(w; a)$, contains at most k' vertices of finite degree, including the possibility of no such vertex. Consider a low degree tree $(TR', \mathbf{d}')$ which consists of the low degree tree generated by the vertices of finite degree in the neighborhood $N(w; a)$ plus additional vertices of finite degree to bring the total to k which are all at distance $a + 1$ from w and all are given degree $l + 1$. Giving them this high degree gives this low degree tree $(TR', \mathbf{d}')$ type (k, l'), where $l' \geq l + 1$, just from a single one of the new vertices. Therefore, there must be a component U' (of type B) in $\widehat{M}_2$ containing the low degree tree $(TR', \mathbf{d}')$ and such that it does not contain any of the vertices $v_1, \ldots, v_{i-1}$. Further, there is a vertex $v_i \in U'$ corresponding to w such that the a-neighborhoods of v_i and w are isomorphic as required in Fact 4. The argument given above works also with the roles of $\widehat{M}_1$ and $\widehat{M}_2$ reversed. Thus, combining Facts 4 and 5 we get that Duplicator has a winning strategy for the game $Ehr_t(M_1, M_2)$.

By Fact 3 the theory $T_{k,l}(\mathbf{r})$ is complete. ▮19

Let M be a countable model of the theory $T_{k,l}(\mathbf{r})$. Based on the proof of Lemma 19, we can define the *minimal model* $\overline{M}$ as that model which we obtain by deleting all components of type D from M, i.e. the model which contains all the components of M of types A, B, C, and E. The proof of Lemma 19 implies that $Th_t(L; M) = Th_t(L; \overline{M})$ for every t and every countable model M of $T_{k,l}(\mathbf{r})$.

For $1 \leq \gamma \leq \beta$, let X_γ be the number of low degree trees of type $(TR_\gamma, \mathbf{d}_\gamma)$ in $\mathcal{G}(n, p_{k,l})$. We calculate $\mathrm{Exp}[X_\gamma]$ as before but watch the constants. The e^{-npk} term yields an additional e^{-c} factor so that

$$\mu_\gamma := \lim_{n\to\infty} \mathrm{Exp}[X_\gamma] = \lambda_\gamma e^{-c} ,$$

where λ_γ is a calculatable positive rational number that involves the $d_i!$ and the number of automorphisms of $(TR_\gamma, \mathbf{d}_\gamma)$ The values of λ_γ are described in full detail in Lemma 21.

Remark 20 Describing the λ_γ can be somewhat complex. For example, let $l = k - 1$ so all $d_i = 0$. Then β is the number of isomorphism types of trees on k vertices. X_γ is the number of isolated trees of type TR_γ, and λ_γ involves the number of automorphisms of TR_γ. Note, however, that for k, l fixed we have a finite list of $(TR, \mathbf{d})$ to consider and therefore a finite and computable list of the λ_γ.

By the following lemma, the X_γ have a joint limit distribution as independent random variables having Poisson distribution with means μ_γ so that

$$a_{\mathbf{r}} := \lim_{n\to\infty} \mu_n(\sigma_{\mathbf{r}}) = \prod_{\gamma=1}^{\beta} e^{-\mu_\gamma} \frac{\mu_\gamma^{r_\gamma}}{r_\gamma!}$$

and so $\sum_{\mathbf{r}\in I} a_{\mathbf{r}} = 1$ and the $\sigma_{\mathbf{r}}, \mathbf{r} \in I$, form a complete set of completions.

Lemma 21 *Let $X_\gamma, 1 \le \gamma \le \beta$, be random variables as above counting the trees $(TR_\gamma, \mathbf{d}_\gamma)$ in $\mathcal{G}(n, p_{k,l})$. Then the multidimensional random variable $(X_1, \ldots, X_\beta)$ converges in probability to the multidimensional random variable consisting of β independent random variables with Poisson distribution and means, $1 \le \gamma \le \beta$,*

$$\mu_\gamma = \frac{1}{k^l \cdot aut(TR_\gamma) \cdot \prod_{i=1}^k d_i!} \cdot e^{-c} ,$$

where $aut(TR_\gamma)$ is the size of the automorphism group of the tree TR_γ and $\mathbf{d}_\gamma = (d_1, \ldots, d_k)$. In particular, for $\mathbf{r} \in I$

$$\lim_{n\to\infty} Prob((X_1, \ldots, X_\beta) = \mathbf{r}) = \prod_{\gamma=1}^{\beta} e^{-\mu_\gamma} \frac{\mu_\gamma^{r_\gamma}}{r_\gamma!} .$$

Proof. Recall that all the trees $(TR_\gamma, \mathbf{d}_\gamma)$ are of the type (k, l). To be precise, let us write $X_\gamma(n) = X_\gamma$ for the random variable counting the trees $(TR_\gamma, \mathbf{d}_\gamma)$ in $\mathcal{G}(n, p_{k,l})$. To show convergence to Poisson distribution it is enough to show convergence of moments.

First, let us compute the expectation of the random variable $X_\gamma(n)$. Let h_γ denote the number of vertices of TR_γ. We get

$$\begin{aligned} \mathrm{Exp}[X_\gamma(n)] &= (1+o(1)) \binom{n}{h_\gamma} \frac{h_\gamma!}{aut(TR_\gamma)} p^{h_\gamma - 1} \prod_{i=1}^{k} \binom{n}{d_i} p^{d_i} (1-p)^n \\ &= (1+o(1)) \frac{1}{aut(TR_\gamma) \cdot \prod_{i=1}^k d_i!} \cdot n(np)^l e^{-knp} \qquad (7) \\ &= (1+o(1)) \frac{1}{k^l \cdot aut(TR_\gamma) \cdot \prod_{i=1}^k d_i!} \cdot e^{-c} . \end{aligned}$$

Using (7), we get for the t-factorial moment $\mathrm{Exp}_t[X_\gamma(n)]$ of $X_\gamma(n)$

$$\mathrm{Exp}_t[X_\gamma(n)] = \mathrm{Exp} \prod_{\alpha=0}^{t-1} (X_\gamma(n) - \alpha) = \prod_{\alpha=0}^{t-1} \mathrm{Exp}[X_\gamma(n - \alpha h_\gamma)] = (1+o(1)) \left(\frac{1}{k^l \cdot aut(TR_\gamma) \cdot \prod_{i=1}^k d_i!} \cdot e^{-c} \right)^t .$$

Thus, the random variable $X_\gamma(n)$ converges in probability $(n \to \infty)$ to a random variable with Poisson distribution $Po(\mu_\gamma)$.

Further, we get for the joint factorial moments

$$\mathrm{Exp} \prod_{t=1}^{u} \prod_{\alpha_t=0}^{w_t - 1} (X_{\gamma_t}(n) - \alpha_t) = (1+o(1)) \prod_{t=1}^{u} \left(\frac{1}{k^l \cdot aut(TR_\gamma) \cdot \prod_{i=1}^k d_i!} \cdot e^{-c} \right)^{w_t} ,$$

$1 \le u \le \beta$. Thus, the random variable $(X_{\gamma_1}, \ldots, X_{\gamma_u})$ converges in probability $(n \to \infty)$ to a multidimensional random variable $(Y_{\gamma_1}, \ldots, Y_{\gamma_u})$ with independent random variables Y_{γ_i} possessing Poisson distributions $Po(\mu_{\gamma_i}), 1 \le \gamma \le u$. ∎21

Now we determine the sets $\mathfrak{M}[\psi]$. When ψ has quantifier rank s and $\mathbf{r}, \mathbf{r}'$ differ only on coordinates on which both are at least s then we claim $\mathbf{r} \in \mathfrak{M}[\psi]$ if and only if $\mathbf{r}' \in \mathfrak{M}[\psi]$. This is because when both minimal countable models $\overline{M}_\mathbf{r}$ and $\overline{M}_{\mathbf{r}'}$ have s or more components of the same type Duplicator can play isomorphically on them throughout an s-move Ehrenfeucht game.

By an s-box on I let us mean the set of $\mathbf{r} \in I$ where for each coordinate r_γ we require either that that coordinate have a specific value, with that value less than s, or we require that the coordinate be at least s. Then $\mathfrak{M}[\psi]$ is the disjoint union of a finite number of s-boxes. But conversely, given a finite union of disjoint s-boxes $\cup_{j=1}^{m} B_j$, let $\mathbf{r}(j) = (r_1(j), \ldots, r_\beta(j)) \in B_j$ and define

$$\varphi_j := \wedge_{\{\gamma: r_\gamma(j) < s\}} \left[LDT(TR_\gamma, \mathbf{d}_\gamma, r_\gamma(j)) \wedge \neg LDT(TR_\gamma, \mathbf{d}_\gamma, r_\gamma(j)+1) \right] \bigwedge \wedge_{\{\gamma: r_\gamma(j) \geq s\}} LDT(TR_\gamma, \mathbf{d}_\gamma, s) \ ,$$

$j = 1, \ldots, m$. (When $r_\gamma(j) = 0$ then the expression in the brackets above should be replaced by $\neg LDT(TR_\gamma, \mathbf{d}_\gamma)$.) Then φ_j and, thus, $\varphi := \vee_{j=1}^{m} \varphi_j$ are themselves first order sentences. Further, $\mathfrak{M}[\varphi] = \cup_{j=1}^{m} \mathfrak{M}[\varphi_j] = \cup_{j=1}^{m} B_j$, i.e. any disjoint union of s-boxes is $\mathfrak{M}[\varphi]$ for some first order sentence φ. (We note that φ may have higher quantifier depth than s.) By Remark 9 this completes the complete characterization of the limit probabilities.

Let $f_\psi(c)$ denote the limit probability of ψ as a function of the real number c. The form of $f_\psi(c)$ is simple to describe if we relax the condition of a complete characterization. The basic sentence $LDT(TR_\gamma, \mathbf{d}_\gamma, i) \wedge \neg LDT(TR_\gamma, \mathbf{d}_\gamma, i+1)$ has a limiting probability f of the form $qe^{-\lambda e^{-c}} e^{-ci}$ where λ, q are positive rational numbers. The sentence $LDT(TR_\gamma, \mathbf{d}_\gamma, s)$ will have an f of the form one minus a finite sum of such terms. To conclude this section, we summarize the results in the following theorem.

Theorem 22 *Let k, l be integers such that $l \geq k-1 \geq 0$. Let ψ be any sentence of the first order language of graphs. Then, in general, the limiting probability $f_\psi(c)$ has the form of a finite sum and difference of rational numbers times finite products of terms of the form above, i.e. a linear combination with rational coefficients of terms of the form $e^{-c\lambda_1} e^{-\lambda_2 e^{-c}}$ where the λ_1, λ_2 are themselves rational numbers. [We include 1 as the null product here.]*

In particular, $f_\psi(c)$ always exists and is always an infinitely differentiable function of c.

References

[B 85] B. Bollobás, Random graphs, Academic Press, 1985.

[BM 77] J.L. Bell, M. Machover, A Course in Mathematical Logic, North-Holland Publishing Company, 1977.

[CK 90] C.C. Chang, H.J. Keisler, Model theory, Elsevier Science Publishers B.V., 1990.

[EF 95] H.-D. Ebbinghaus, J. Flum, Finite model theory, Springer-Verlag, 1995.

[Eh 61] A. Ehrenfeucht, An application of games to the completeness problem for formalized theories, Fundamentæ Math. 49 (1961) 129-141.

[ER 60] P. Erdős, A. Rényi, On the evolution of random graphs, Magyar Tud. Akad. Mat. Kutató Int. Közl 5 (1960), 17-61.

[L 92] J. F. Lynch, Probabilities of Sentences about Very Sparse Random Graphs, Random Struct. Alg. 3 (1992), 33 - 53.

[ŁS 91] T. Łuczak, J. Spencer, When does the zero-one law hold ?, J. Amer. Math. Soc. 4 (1991), 451 - 468.