

# 1 Introduction

1. Already Fagin considers a stronger language and shows that in second order there is no convergence.
2. Many others asked the following question: Find a language which is both stronger than the first order language of graph and at the same time still obeys the 0-1 law.
3. From a graph theorist point of view, the question looks like the following: find a language obeying a 0-1 law which is strong enough to express  $P_1, P_2, \dots$ , where  $P_i$  is a property of graphs that is of interest to graph theorists.

## 1.1 Lindström quantifiers

The definition in Flum's book is the weakest definition, but probably the standard one. Consider proving in that sense and only mentioning that our technique will work in the most general sense as well.

### 1.1.1 notation

Let  $K$  be a class of graphs and let  $\varphi(x, y)$  be a predicate. We denote the Lindström quantifier for that class by  $Q_K$  and use  $Q_K xy. \varphi(x, y)$  to say "the graph with edges defined by  $\varphi$  is in  $K$ ". Formally, let  $K$  and  $\varphi$  be as above, then

$$G \models Q_K xy. \varphi(x, y) \iff (V[G], \{(x, y) \mid G \models \varphi(x, y)\}) \in K.$$

In this paper  $K$  will be one of the following sets:

1. CONN, the set of all connected graphs,
2. HAM, the set of all Hamiltonian graphs, that is, the set of graphs having a Hamilton cycle, and
3.  $CH_k$ , the set of all graphs having chromatic number  $k$ .

## 1.2 results

Our results are of mixed nature. We show that there are languages able to express any first order sentence, connectivity and  $k$ -colorability for any fixed  $k$ , and still obey the zero-one law for  $G(n, p)$  for any constant  $0 < p < 1$ . On the other hand, for the same model of random graphs we show that in any language stronger than the first order language of graphs that is also able to express Hamiltonicity, there is a sentence with no limiting probability.

**Theorem 1.** *For every constant  $0 < p < 1$  and for every sentence  $\varphi$  in  $\mathcal{L}(Q_{\text{CONN}}, Q_{\text{CH}_2}, Q_{\text{CH}_3}, \dots)$ ,*

$$\lim \Pr[G(n, p) \models \varphi] \in \{0, 1\}.$$

**Theorem 2.** Fix a constant  $0 < p < 1$ . There exists a sentence  $\varphi \in \mathcal{L}(Q_{\text{HAM}})$  such that

$$\lim_{n \rightarrow \infty} \Pr[G(n, p) \models \varphi]$$

does not exist.

## 2 Connectivity and chromatic number

In this section we prove Theorem 1.

*Proof of Theorem 1.* Let  $p$  be a constant,  $0 < p < 1$  and let  $\varphi$  be a sentence in the language  $\mathcal{L} = \mathcal{L}(Q_{\text{CONN}}, Q_{\text{CH}_2}, Q_{\text{CH}_3}, \dots)$ , which is the first-order closure of the quantifiers  $Q_{\text{CONN}}, Q_{\text{CH}_2}, Q_{\text{CH}_3}, \dots$  (using the weakest sense of a Lindström quantifier). We wish to show that the limiting probability of  $\varphi$  in  $G(n, p)$  is either zero or one. We use induction on the structure of  $\varphi$ .

claim: Let  $\varphi(\bar{a})$  be a predicate in  $\mathcal{L}$ , where  $\bar{a}$  are free variables (seen as parameters of  $\varphi$ ). Then the limiting probability of  $\varphi(\bar{a})$  in  $G(n, p)$  with  $0 < p < 1$  is either zero or one, depending only on the relations among  $\bar{a}$ .

claim proof: By induction on the structure of the predicate. For atomic predicate it is clear (that is, the truth value of  $a_1 = a_2$  and  $a_1 \sim a_2$  depend on the relations among  $a_1$  and  $a_2 \dots$ ).

If  $\varphi$  is of the form  $\neg\psi$  or  $\psi_1 \wedge \psi_2$  then the claim is immediate.

Assume that  $\varphi(\bar{a})$  is of the form  $\varphi(\bar{a}) = Q\bar{b}.\psi(\bar{a}, \bar{b})$  where  $Q$  is one of  $\exists, Q_{\text{CONN}}, Q_{\text{CH}_k}$  and  $\bar{b}$  is of length one or two. Consider the equivalence relation  $E_{\bar{a}}$  over vertices defined by

$$xE_{\bar{a}}y \iff \forall a \in \bar{a}.(x = a \leftrightarrow y = a) \wedge (x \sim a \leftrightarrow y \sim a),$$

and let  $B_i, 1 \leq i \leq k_* = |\bar{a}| + 2^{|\bar{a}|}$  be the equivalence classes in  $[n]/E_{\bar{a}}$ . By the induction hypothesis, the limiting probability of  $\psi(\bar{a}, \bar{b})$  depends only on the relations among  $\bar{a}$  and  $\bar{b}$ . In other words, if  $Q = \exists$  then the aforementioned limiting probability is determined by the relations among  $\bar{a}$  and the  $i$  for which  $b \in B_i$ . If  $Q = Q_{\text{CONN}}$  or  $Q = Q_{\text{CH}_k}$  then the limiting probability depends on relations among  $\bar{a}$ , the equivalence classes containing the two variables of  $\bar{b}$  and the relations between these variables.

Now, for the first time, probability enters the proof. By Lemma ?? we know that a.a.s. in  $G(n, p)$  all of the  $B_i$ 's are either singletons of the form  $\{a_i\}$  for some  $a_i \in \bar{a}$  or have linear size.

Assume first that  $Q$  is the existential quantifier. If for any of the (finitely many)  $B_i$ 's the limiting probability of  $\psi(\bar{a}, b)$  is one for  $b \in B_i$ , then  $\varphi$  also has limiting probability one. Otherwise  $\varphi$ 's limiting probability is zero.

Next assume that  $\varphi(\bar{a})$  is of the form  $\varphi(\bar{a}) = Q_{\text{CONN}}b_1, b_2.\psi(\bar{a}, b_1, b_2)$ . We want to show that the limiting probability for the connectivity of the graph  $F = ([n], \{\{b_1, b_2\} | \psi(\bar{a}, b_1, b_2)\})$  depends solely on the interrelations among the members of  $\bar{a}$ . Notice however, that by the induction hypothesis, once  $\bar{a}$  is given the limiting probability of  $\psi(\bar{a}, b_1, b_2)$  depends upon the classes of  $[n]/E_{\bar{a}}$

containing  $b_1$  and  $b_2$ , whether  $b_1 \sim_{G(n,p)} b_2$  and whether  $b_1 = b_2$ . In particular, if both  $B_1$  and  $B_2$  are not singleton, then we are in one of three situations. It may be that for every  $b_1 \in B_1, b_2 \in B_2$  one has  $\psi(\bar{a}, b_1, b_2)$ , or that for every such pair one has  $\neg\psi(\bar{a}, b_1, b_2)$  or that the  $\psi(\bar{a}, b_1, b_2)$  depends *solely* on the truth value of  $b_1 \sim_{G(n,p)} b_2$ .

Consider the graph  $H$  whose vertex set is the quotient set  $[n]/E_{\bar{a}}$  and its edge set is determined by

$$B_1 \sim_H B_2 \iff \exists b_1 \in B_1, b_2 \in B_2. \psi(\bar{a}, b_1, b_2).$$

We argue that if  $H$  is disconnected then so is  $F$ , and if  $H$  is connected that  $F$  is connected with probability tending to one with  $n$ . Let  $u$  and  $w$  be two vertices of  $F$  and assume that there is a path  $v_0 = u, v_1, \dots, v_l = w$  in  $F$  connecting  $u$  and  $w$ . Then  $[v_0], [v_1], \dots, [v_l]$  is a walk connecting  $u$  and  $w$  in  $H$  (formally this may not be a walk since it may be that  $[v_i] = [v_j]$  without a loop. Still it contains a path connecting  $[v]$  and  $[w]$ ). Hence, if  $H$  is disconnected then  $F$  is disconnected as well.

Assume now that  $H$  is connected and let  $u, w$  be two vertices in  $F$ . Denote the vertices of the shortest path connecting  $[u]$  and  $[w]$  in  $H$  by  $[v_0 = u], [v_1], \dots, [v_l = w]$ . We find a path connecting  $u$  and  $w$  in  $F$  as follows. Start by denoting  $u$  as  $v'_0$  and assume we have a path in  $F$  connecting  $u$  to  $v'_i$  where  $v'_j \in [v_j]$  for every  $0 \leq j \leq i$ . Since  $[v_i] \sim_H [v_{i+1}]$ , we know that there are two vertices  $b_1 \in [v_i]$  and  $b_2 \in [v_{i+1}]$  such that  $\psi(\bar{a}, b_1, b_2)$  holds. If either  $[v_i]$  or  $[v_{i+1}]$  is a singleton (or both), then, by definition of  $H$ 's vertices, all the vertices of  $[v_i]$  relate in the same manner to all the vertices of  $[v_{i+1}]$ . Hence we can take  $v'_{i+1} = v_{i+1}$ . If both  $[v_i]$  and  $[v_{i+1}]$  are not singletons, then by Lemma ?? both are of size linear in  $n$ . If the value of  $\psi(\bar{a}, b_1, b_2)$  does not depend on  $b_1 \sim_{G(n,p)} b_2$ , we can simply pick  $v'_{i+1} = v_{i+1}$ . Otherwise we use the fact that probability of  $v'_i$  not having a neighbor or a non-neighbor in a vertex set of size  $cn$  is exactly  $p^{cn} + (1-p)^{cn} \leq c'^n$  where  $0 < c' < 1$  is some constant. Hence, with probability tending to one exponentially fast we can find a vertex  $v'_{i+1} \in [v_{i+1}]$  satisfying  $\psi(\bar{a}, v'_i, v'_{i+1})$ . When picking  $v'_{i-1}$  we need to take extra care and make sure that it is a neighbor of  $w$  in  $F$  as well. This is easily done using a similar computation.

We have shown that with probability tending exponentially fast to one, the connectivity of  $F$  depends only on the connectivity of  $H$ . By the definition of  $J$ , it is connected depending only  $\bar{a}$ .

Finally let  $Q = \text{CH}_k$  for some constant  $k \in \mathbb{N}$ . We look again on the graphs  $F$  and  $H$  defined above. By arguments similar to the above, there are  $k_*$  vertices  $v_1, \dots, v_{k_*} \in F$  such that  $v_i \in B_i$  and  $v_i \sim_F v_j \iff B_i \sim_H B_j$ . Hence  $F$  has a copy of  $H$  as a subgraph and therefore  $\chi(F) \geq \chi(H)$ . We argue that with probability tending to one there two possibilities: either  $\chi(F) = \chi(H)$  or  $\chi(F) = \Omega(n/\ln n)$ . If for every  $i$  the vertices of  $B_i$  form an independent set in  $F$ , then we can color  $F$  by assigning the color of  $B_i$  to all its vertices. Hence in this case  $\chi(F) = \chi(H)$ . Assume now that there is a set  $B_i$  that is not an independent set and let  $n' = |B_i|$ . The graph spanned by  $F$  on  $B_i$  is either a

clique, a spanned subgraph of  $G(n, p)$  or the complement of a spanned subgraph of  $G(n, p)$ . In the first case its chromatic number is  $n'$ . In the second and third the chromatic number of  $F[B_i]$  is in fact the chromatic number of a random graph with  $n'$  vertices and edge probability  $p'$  being equal to  $p$  or to  $1 - p$ . In an exciting paper [?] Bollobás showed that with probability tending to one this number is  $(1/2 + o(1))n'/\log_b n'$  where  $b = 1/(1 - p')$ . Thus in this case the chromatic number of  $F$  is of order  $n/\ln n$ .

Summarizing the above argument we get that the chromatic number of  $([n], \{\{b_1, b_2\} | \psi(\bar{a}, b_1, b_2)\})$  depends only on the relations among  $\bar{a}$  and it is either a specific constant or growing to infinity with  $n$ . Therefore the limiting probability of  $Q_{\text{CH}_k} xy.\psi(\bar{a}, x, y)$  is either zero or one, depending on the relations among  $\bar{a}$ .

This ends the proof of the claim. The theorem follows immediately as  $\varphi$  is a sentence in  $\mathcal{L}$  with no free variables.  $\square$

### 3 Hamiltonicity

In this section we prove Theorem 2 by showing that using  $Q_{\text{HAM}}$  we can express any monadic second order sentence on a set with size tending to infinity as  $n$  grows.

*Proof of Theorem 2.* We start by enriching the language. Let  $S$  and  $T$  be two sets of vertices. We define

$$S > T \iff Q_{\text{HAM}} x, y. (x \in T \setminus S \wedge y \in S \setminus T) \vee \\ (x \in S \setminus T \wedge y \notin S \Delta T) \vee (x \notin S \Delta T \wedge y \notin S \Delta T),$$

where  $S \Delta T$  is the symmetric difference of  $S$  and  $T$ , that is,  $S \Delta T = (S \cup T) \setminus (S \cap T)$ . Notice that the sentence holds if and only if  $|S| > |T|$  and  $|\overline{S \Delta T}| > |S \setminus T|$ . This is always true as there is no probability here — we do not look at the edges of the underlying graph.

We will consider sets of two types, the neighborhood of a vertex  $v$ , denoted  $N(v)$ , and the set of vertices having the same degree as  $v$  denoted here by  $D(v)$ . The degree of a vertex in  $G(n, p)$  is a random variable following a binomial distribution with parameters  $n - 1$  and  $p$ . In our case  $p$  is constant and all the degrees lie in the range  $(np - c\sqrt{n \ln n}, np + c\sqrt{n \ln n})$  for some constant  $c$  (see [?] or [1, Chapter 3]). We also mention the similar fact that a.s. for any two vertices  $u, v$ , the *codegree* of  $u$  and  $v$ , which is  $|N(u) \cap N(v)|$ , lies in the range  $(np^2 - c\sqrt{n \ln n}, np^2 + c\sqrt{n \ln n})$ . Therefore for neighborhoods of vertices we can express “larger”, that is, asymptotically almost surely for every two vertices  $u, v$  one has  $N(u) > N(v)$  if and only if  $|N(u)| > |N(v)|$ . The distribution of  $D(v)$  is also quite well known. In particular, if  $v$  is a vertex of minimal degree then a.s.  $|D(v)| = 1$ , if  $d(v) = \lceil pn \rceil$  then  $|D(v)| = O(\sqrt{n})$  and the transition is rather smooth (there are jumps in the beginning, but it does not affect this study). Given a vertex  $v$  we can express  $u \in D(v)$  by writing  $\neg(N(v) > N(u)) \wedge \neg(N(u) > N(v))$ . By the above remark a.s.  $u \in D(v)$  if

and only of  $d(u) = d(v)$ . By the same remark and by the fact that a.a.s. for every  $v$  one has  $|D(v)| = O(\sqrt{n})$  we get that we can compare the sizes of  $D(u)$  and  $D(v)$ .

We want to be able to express monadic second order, that is, to be able to quantify over sets. We do this by looking at intersections of neighborhoods with the set under discussion. Lemma 3 states that for small enough sets we can express any monadic second order sentence in this manner. That is, if  $S$  is a set of vertices of size at most  $\ln \ln n$  then given a monadic existential second order sentence of the form  $A = \exists P_1, P_2, \dots, P_k. \varphi(P_1, P_2, \dots, P_k)$  we write instead

$$A' = \exists v_1, v_2, \dots, v_k \notin S. \varphi'(P_1, P_2, \dots, P_k),$$

where  $\varphi'$  is obtained from  $\varphi$  by replacing each term of the form  $P_i(u)$  with  $v_i \sim u$ .

In [2] Kaufmann and Shelah demonstrated that using monadic second order language one can interpret a segment of arithmetic. In particular, letting  $n$  be the size of the model, one may say “ $0 \leq \log_*(n) \bmod 100 \leq 50$ ” which obviously has no limiting probability. We will use same sentence encoded using the method aforementioned. Let  $\text{BIGGAP}'(S)$  be this encoded sentence.

Now we can write the desired sentence:

$$\exists v. \text{BIGGAP}'(D(v)) \wedge \forall u. \text{BIGGAP}'(D(u)) \rightarrow \neg D(u) > D(v). \quad (1)$$

By Lemma 3 we know that the maximal  $|D(v)|$  (recall, there is no probability involved in  $<$ ) is a.a.s. of size between  $\ln \ln n$  and  $n$ , and therefore the sentence in Equation (1) has no limiting probability as required.  $\square$

**Lemma 3.** *Let  $0 < p < 1$  be constant and consider  $G(n, p)$ . Then a.a.s for every set of vertices  $S$  of size  $|S| \leq \ln \ln n$  the following holds: For every subset  $A \subset S$  there is a vertex  $v \notin S$  such that*

$$N(v) \cap S = A.$$

*Proof.* Let  $A \subset S$  and let  $v \notin S$  be a vertex. The probability that  $N(v) \cap S = A$  is exactly  $p^{|A|}(1-p)^{\binom{|S|}{2}-|A|} \leq 2^{-(\ln \ln n)^2}$ . Therefore the probability that there is no witness for the set is bounded by  $(1 - 2^{-(\ln \ln n)^2})^{n - \ln \ln n} \leq e^{-\sqrt{n}}$ . Take a union bound over all the  $2^{\ln \ln n} \leq \ln n$  possible subsets of  $S$  and then another union bound over all the  $\sum_{k=1}^{\ln \ln n} \binom{n}{k} \leq \binom{n}{\ln \ln n + 1} \leq e^{\ln n \ln \ln n}$  sets of size at most  $k \leq \ln \ln n$ . Hence we have that a.a.s. for every set  $S$  of size at most  $\ln \ln n$  there is a witness for every subset, and the proof is complete.  $\square$

## 4 Concluding remarks

1. This study suggests a notion of simplicity for properties of graphs. Properties for which the first order closure obeys a 0-1 law are simpler than properties  $P$  for which  $\mathcal{L}(Q_P)$  has nonconverging sentences. There is an intermediate level of properties for which  $\mathcal{L}(Q_P)$  does not obey the 0-1 law but every sentence has a limiting probability.

2. adding density to the language
3. other properties
4. general sorting scheme
5. languages able to define more sets and still obey the 0-1 law

## References

- [1] B. Bollobás. *Random Graphs*. Number 73 in Cambridge studies in advanced mathematics. Cambridge University Press, Cambridge, UK, Second edition, Jan. 2001.
- [2] M. Kaufmann and S. Shelah. On random models of finite power and monadic logic. *Discrete Mathematics*, 54(3):285–293, 1985.