Random Databases and Threshold for Monotone Non-Recursive Datalog

Konstantin Korovin and Andrei Voronkov *

School of Computer Science, The University of Manchester
{korovin|voronkov}@cs.man.ac.uk

Abstract. In this paper we define a model of randomly generated databases and show how one can compute the threshold functions for queries expressible in monotone non-recursive datalog$^{\neq}$. We also show that monotone non-recursive datalog$^{\neq}$ cannot express any property with a sharp threshold. Finally, we show that non-recursive datalog$^{\neq}$ has a $0-1$ law for a large class of probability functions, defined in the paper.

1 Introduction

In this paper we consider random databases, in which relations are generated based on some standard probability distributions. Namely, we assume that every tuple belongs to a relation with an equal probability p, called the *tuple probability*. Similar probabilistic models have been intensively studied in the theory of random graphs (see e.g. [2, 11]), but also in the database theory (see e.g. [15, 10, 4]). We are interested in investigating the behavior of queries expressible in database query languages on the random databases. There are several interesting questions that can be asked about such queries, for example, calculating the probability of a query to be true as a function of p and the domain size. One of the most important characteristics of probabilistic behavior of monotone properties of structures is their *threshold functions*. The threshold functions can be used for characterizing asymptotic probabilistic behavior of queries, that is, probabilistic behavior when the database is growing. Intuitively, if a database grows faster than the threshold function, then the probability of the query to be true converges to 1; likewise, if a database grows slower than the threshold function, then the probability of the query to be true converges to 0. Although it is known that every monotone property has a threshold function, the problem of determining threshold functions for particular properties of random structures is an active research area in combinatorics. In this paper we show how one can compute the threshold functions for queries expressible in monotone non-recursive datalog$^{\neq}$. We also introduce a notion of *density* of queries expressible in this language and show the exact relationship between densities and threshold functions of the queries.

Another natural question is the power of databases query languages to express some phenomena related to randomly generated structures. For example, it is known that

* The authors are partially supported by grants from EPSRC and the Faculty of Science and Technology.

some monotone properties actually have a *sharp threshold* (see, e. g.,[8]). Finding interesting properties with a sharp threshold or identifying whether some natural properties have a sharp threshold are questions intensively studied in combinatorics, and in particular in the theory of random graphs. We are interested in studying the expressive power of query languages on random databases with respect to the threshold behavior. In particular, we show that monotone non-recursive datalog$^{\neq}$ cannot express any property with a sharp threshold. It is not hard to give an example of a first-order property having a sharp threshold, which shows that relatively simple extensions of monotone non-recursive datalog$^{\neq}$ can express such properties. However, the exact relation between query languages and expressibility of properties with a sharp threshold remains an open question.

The main contributions of this paper are the following.

1. We define a probabilistic model of randomly generated databases.
2. We show how one can compute the threshold functions for queries expressible in monotone non-recursive datalog$^{\neq}$.
3. We introduce the notion of density of queries expressible in monotone non-recursive datalog$^{\neq}$, and show the exact relationship between densities and threshold functions of the queries.
4. We show that monotone non-recursive datalog$^{\neq}$ has a $0-1$ law w.r.t. every probability function $p(n)$ satisfying the following condition: for every rational $q > 0$ either $p \gg n^{-q}$ or $p \ll n^{-q}$.
5. We show that monotone non-recursive datalog$^{\neq}$ cannot express any property with a sharp threshold.

For the future research, it would be interesting to consider different probabilistic models for some big existing and evolving databases such as WEB, and study behavior of properties expressible by database query languages in such models (see e.g. [13]).

2 Preliminaries

In this paper we study asymptotic properties of finite relational structures with constants. In particular, we are interested in properties expressible by existential formulas in the language with equality in which negation can only be applied to equalities. In this section we introduce definitions of structures and the query language datalog$^{\neq}$.

Generally speaking, we are dealing with randomly generated finite structures. For simplicity, we assume that the structure contains only one relation symbol. We only consider boolean queries, that is, queries with a yes-no answer. In order to deal with some standard queries as boolean queries, we introduce constants in the language. For example, for an input binary relation R, reachability can be formulated as follows: given two elements (x, y), is there a path from x to y in which arcs are pairs belonging to R? Normally, such a query would be formulated using a binary output relation T that is the transitive closure of R. However, since we are restricted to boolean queries, we cannot use a binary output relation. Instead, we can introduce two constants c_1 and c_2 and use $T(c_1, c_2)$ to represent reachability of c_2 from c_1.

Structures. We consider signatures with one relation symbol R and constants. Consider a finite signature $\Sigma = \{R, c_1, \ldots, c_l\}$, where R is a relation symbol of some arity $r > 0$ and each c_i is a constant. Denote by $\mathbb{N}$ the set of natural numbers and by $[n]$ the finite set $\{1, \ldots, n\}$. For every $n \geq l$ denote by $\mathbb{M}_n$ the set of all structures of the signature Σ with the domain $[n]$ in which the constants $c_1, \ldots c_l$ are interpreted as the numbers $1, \ldots, l$ respectively. For $n < l$ we let $\mathbb{M}_n = \emptyset$. Without loss of generality we can restrict ourselves to structures in $\cup_{n>0}\mathbb{M}_n$. For a structure M we denote the interpretation of R by R_M.

For simplicity of the presentation we require all constants to be interpreted by distinct elements of the domain. However, our main results remain valid if the interpretations of constants can coincide.

Properties of structures. A *property of structures* is a parametrized family $\{A_n\}_{n=0}^{\infty}$ of structures such that for each $n \in \mathbb{N}$ the set A_n consists of structures in $\mathbb{M}_n$. A property $\{A_n\}_{n=0}^{\infty}$ is called *monotone* if for each $n \in \mathbb{N}$ and $M, M' \in \mathbb{M}_n$ such that $R_M \subseteq R_{M'}$ and $M \in A_n$ we have $M' \in A_n$. In other words, if $M \in A_n$, then by adding a tuple to R_M we obtain a structure in A_n.

For example, if R is binary and we have no constants then each structure is a directed graph; it is easy to see that the following properties are monotone: to contain a given graph, to have no isolated vertices, to be a non-planar graph, and connectivity.

Language. Evidently, every formula φ of the signature Σ defines a property: we let $M \in A_n$ if $M \in \mathbb{M}_n$ and $M \models \varphi$. We consider the subset of all formulas of Σ built from formulas $R(\bar{s})$, $s = t$ and $s \neq t$ using only $\vee$, $\wedge$, and $\exists$. Denote this set of formulas by $\Sigma_1^{\neq}$. For example, if R is ternary and $c_1 \in \Sigma$ then $\exists x, y(R(x, c_1, y) \wedge x \neq y \wedge R(x, x, y))$ is a formula in $\Sigma_1^{\neq}$. It is easy to see that all properties definable by formulas in $\Sigma_1^{\neq}$ are monotone. It is also not hard to argue that $\Sigma_1^{\neq}$ has the same expressive power as non-recursive datalog$^{\neq}$.

Theorem 1. *Every query defined by a formula in $\Sigma_1^{\neq}$ is definable by a non-recursive datalog$^{\neq}$-program and vice versa.*

The proof is standard.

Monotone properties of undirected random graphs have been intensively studied (see e.g. [2, 11]). One of the most important characteristics of a monotone property of a random graph is to have the *threshold function*. To define this notion for structures, we first introduce a probabilistic model of structures, which is similar to the binomial model of the random graphs.

Random structures. Let us introduce the probability space $(\mathbb{M}_n, \mathcal{F}, p, \mu)$ on structures, where $\mathcal{F}$ is the set of all subsets of $\mathbb{M}_n$, $0 \leq p \leq 1$ and the probability function μ is defined as follows:

$$\mu(M) = p^m (1-p)^{n^r - m},$$

where $M \in \mathbb{M}_n$, m is the number of tuples satisfying the relation R in M (remember that r is the arity of R). This can be viewed as a result of n^r independent coin flippings with the probability p of success: every tuple is included in R_M with the probability p. We will refer to p as the *tuple probability*. We denote by $M_{n,p}$ the random structure corresponding to this probability distribution, that is a random element of this probability space. When we consider asymptotic behavior of properties of structures we assume that p is a function of n. Let us note that the presence of constants in the language introduces certain peculiarities for the probabilistic analysis. For example, it is well-known that the first-order logic on random graphs (without constants) has a $0-1$ law for the constant distribution, e.g. $p(n) = 1/2$, (see [7, 9]). However, when we consider the first-order logic with constants then the logic fails to have a $0-1$ law (see Section 6 for more on $0-1$ laws for datalog$^{\neq}$).

Notation. Asymptotics. For functions f, g, we write $f(n) \asymp g(n)$ if $f(n) = O(g(n))$ and $g(n) = O(f(n))$, also we write $f(n) = \Theta(g(n))$ if $f(n) \asymp g(n)$, and write $f(n) \ll g(n)$ if $f(n) = o(g(n))$.

Probability. The expected value of a random variable X is denoted by $\mathbb{E}(X)$. The indicator variable of a property A will be denoted as I_A.

Thresholds. Let $\{A_n\}_{n=0}^{\infty}$ be a monotone property. A function $p'(n)$ is called the *threshold function*, or simply *threshold*, for this property if

$$\lim_{n\to\infty} \mu(M_{n,p} \in A_n) = \begin{cases} 0 \textit{ if } p \ll p'; \\ 1 \textit{ if } p \gg p'. \end{cases}$$

Bollobás and Thomason [3] prove that every monotone property has a threshold. *One of our main results is finding threshold functions for all properties of structures expressible by formulas in $\Sigma_1^{\neq}$. Namely, we show that for every formula $\varphi \in \Sigma_1^{\neq}$ the threshold function is either constant or of the form n^{-q_φ}, where q_φ is a non-negative rational number, and give an algorithm for computing q_φ from a given formula φ.*

The problem of determining threshold functions for properties of random structures is an active research area in combinatorics. In particular, the problem of finding thresholds for graph containment has a long history. Starting from the paper [6] where this problem was solved for a special case of balanced undirected graphs, and culminating 21 years later in [1] where this problem is solved for arbitrary undirected graphs. Let us note that the corresponding *structure containment problem*, defined in Section 3, is expressible in $\Sigma_1^{\neq}$.

Sharp thresholds. Some monotone properties of structures can possess a *sharp threshold*. We call the threshold p' for a property $\{A_n\}_{n=0}^{\infty}$ *sharp* if for every $\epsilon > 0$

$$\lim_{n\to\infty} \mu(M_{n,p} \in A_n) = \begin{cases} 0 \textit{ if } p \leq (1-\epsilon)p'; \\ 1 \textit{ if } p \geq (1+\epsilon)p'. \end{cases}$$

If the threshold for a monotone property is not sharp, then we say that the property has a *coarse threshold*. *We will show that for every property definable by a formula in $\Sigma_1^{\neq}$ the threshold is coarse.* We will use the following reformulation of the notion of sharp threshold (see [11]). Consider a property $\{A_n\}_{n=0}^{\infty}$. For each ϵ such that $0 < \epsilon < 1$ define $p_\epsilon(n)$ be the tuple probability such that $\mu(M_{n,p_\epsilon} \in A_n) = \epsilon$. The property $\{A_n\}_{n=0}^{\infty}$ has a sharp threshold if and only if $\lim_{n\to\infty} p_\epsilon(n)/p_{1/2}(n) = 1$ for every ϵ such that $0 < \epsilon < 1$.

Outline. The rest of this paper is structured as follows. In Section 3 we introduce and study the structure containment and the weak structure containment problems. We prove that every sentence in $\Sigma_1^{\neq}$ is equivalent to a weak structure containment problem. In Section 4 we show how one can calculate the threshold for any structure containment problem and show that the threshold is coarse. In Section 5 we prove similar results for the weak structure containment problems. Finally, in Section 6 we present the main results of this paper.

3 Structure Containment

In this section we introduce the structure containment and the weak structure containment properties. We prove that every formula in $\Sigma_1^{\neq}$ is equivalent to a weak structure containment property.

Structure containment. Consider an arbitrary but fixed structure $\mathcal{M}$. We say that a structure M *contains* $\mathcal{M}$, denoted by $M \sqsupseteq \mathcal{M}$, if there is an injective homomorphism from $\mathcal{M}$ into M. Note that structure containment is *language-dependent*, since every homomorphism maps an integer m into itself if $c_m \in \Sigma$. We define the *structure containment property*, (also referred to as $\mathcal{M}$*-containment property* when we want to emphasise the structure $\mathcal{M}$), to be the set of all structures that contain $\mathcal{M}$. Evidently, for every structure $\mathcal{M}$ the structure containment property is monotone.

Weak structure containment. Let S be a finite family of structures. We say that a structure M *weakly contains* S, denoted by $M \sqsupseteq_w S$, if it contains at least one structure from S. We define the *weak structure containment property*, (also referred to as *weak* S*-containment property* when we want to emphasise the family S), to be the set of all structures that weakly contain S. Evidently, for every finite family S of structures the weak structure containment property is monotone.

Theorem 2. *Given a sentence A of $\Sigma_1^{\neq}$, one can effectively find a finite family of structures S such that for every structure M we have $M \models A$ if and only if $M \sqsupseteq_w S$.*

Proof. We assume that $c_1, \ldots, c_l$ are all constants of the language. Let B be any formula with free variables $x_1, \ldots, x_n$. Denote by $\exists B$ the *existential closure of* B*,* that is, the formula $\exists x_1 \ldots \exists x_n\ B$. Also, denote by $\exists^n B$ the following formula:

$$\exists x_1 \ldots \exists x_n \Big(\bigwedge_{1 \le i \le n,\ 1 \le j \le l} x_i \neq c_j \wedge \bigwedge_{1 \le i < j \le n} x_i \neq x_j \wedge B \Big).$$

We say that a sentence B is *simple* if it has the form $\top$, $\bot$, or $\exists^n B'$ such that $n \ge 0$ and B' is a conjunction of atomic formulas $R(t_1, \ldots, t_r)$. (By $\top$, $\bot$ we denote the logical constants true and false respectively).

Let us prove the following statement: *for every sentence D of $\Sigma_1^{\neq}$ one can effectively find a finite set of simple formulas $D_1, \ldots, D_u$, where $u \ge 0$, such that D is equivalent to $D_1 \vee \ldots \vee D_u$.* The proof is as follows. First, we can convert D to a formula $\exists C_1 \vee \ldots \vee \exists C_v$, such that C_i is a conjunction of atomic formulas, using prenexing, CNF transformation, and anti-prenexing. Hence, it is enough to prove this statement when D has the form $\exists D'$, where D' is a conjunction $D_1 \wedge \ldots \wedge D_v$ and each D_i is an atomic formula. Such a formula D may still be not simple, since some of the conditions on simple formulas may not be satisfied. When D is not simple, do the following.

1. If D' contains an equality $t = t$, replace this equality by $\top$ in D'.
2. If D' contains an equality $c_p = c_q$, where $p \neq q$, replace D by $\perp$.
3. If D' contains an inequality $t \neq t$, replace D by $\perp$.
4. If D' contains an inequality $c_p \neq c_q$, where $p \neq q$, replace this inequality by $\top$ in D'.
5. If D' contains an equality $x = t$, such that t is different from x, replace x by t in D'.

If none of the following transformations is applicable to D, then D' contains no equalities and no formulas $c_p \neq c_q$. If D is not simple, then apply also the following transformations to D.

1. If, for some variables x_p, x_q occurring in D', D' does not contain $x_p \neq x_q$, replace D by $\exists(D' \wedge x_p = x_q) \vee \exists(D' \wedge x_p \neq x_q)$.
2. If, for some variable x_p occurring in D' and constant c_q, D' does not contain $x_p \neq c_q$, replace D by $\exists(D' \wedge x_p = c_q) \vee \exists(D' \wedge x_p \neq c_q)$.

Then proceed as before. It is not hard to argue that after some finite number of steps we will terminate and obtain a disjunction of simple formulas equivalent to D.

Now let us prove the following statement: *for every simple formula $D = \exists^n D'$ one can effectively find a structure $\mathcal{M}$ such that for all structures M, $M \models A$ if and only if $M \sqsupseteq \mathcal{M}$*. Without loss of generality one can assume that the free variables of D' are $x_1, \ldots, x_n$. Define the structure $\mathcal{M}$ as follows. The domain of $\mathcal{M}$ is $[l+n]$. To define $R_{\mathcal{M}}$, let us introduce an *interpretation function* f mapping all variables occurring in D and all constants to numbers in the domain of $\mathcal{M}$ as follows: $f(c_i) = i$ and $f(x_j) = j + l$. Then define the relation $R_{\mathcal{M}}$ as follows:

$$R_{\mathcal{M}} = \{(f(t_1), \ldots, f(t_r)) \mid R(t_1, \ldots, t_r) \text{ occurs in } D'\}.$$

Let us show that $\mathcal{M}$ satisfies the claim. Indeed, it is not hard to argue that $\mathcal{M} \models D$. Let M be any structure such that $M \models \exists^n D'$. Then there exist pairwise distinct numbers $k_1, \ldots, k_n$ such that (i) $\{k_1, \ldots, k_n\} \cap [l] = \emptyset$; (ii) the domain of M contains $k_1, \ldots, k_n$; (iii) for the function g defined by $g(x_i) = k_i$ we have $M, g \models D'$. It is not hard to argue that the function $h : \mathbb{N} \to \mathbb{N}$ defined by

$$h(s) = \begin{cases} s, & \text{if } 1 \leq s \leq l; \\ k_{s-l}, & \text{if } l+1 \leq s \leq l+n \end{cases}$$

is an injective homomorphism from $\mathcal{M}$ into M, so $M \sqsupseteq \mathcal{M}$.

From what we proved it follows that the sentence A (from the statement of the theorem) is equivalent to a disjunction $D_1 \vee \ldots \vee D_u$ of simple formulas and that there are structures $\mathcal{M}_1, \ldots, \mathcal{M}_u$ such that the following holds. For every structure M and $i = 1, \ldots, u$, $M \models D_i$ if and only if $M \sqsupseteq \mathcal{M}_i$. Therefore, for every structure M we have $M \models A$ if and only if $M \sqsupseteq_w \{\mathcal{M}_1, \ldots, \mathcal{M}_u\}$, hence we can define S to be $\{\mathcal{M}_1, \ldots, \mathcal{M}_u\}$.

Let us note that by a simple modification of this proof we could have proved a similar result for the case when different constants may also be interpreted by the same element.

4 Threshold for the Structure Containment

In this section we show how to compute the threshold function for any structure containment property and show that every such property has a coarse threshold.

Let us consider $\mathcal{M}$-containment property for an arbitrary but fixed structure $\mathcal{M}$. For a structure $M \in \mathbb{M}_n$, we write $M \sqsupset \mathcal{M}$ if $M \sqsupseteq \mathcal{M}$ and $\mathcal{M}$ is not isomorphic to M. Since we are studying the asymptotic behavior of the properties we can always assume that n is greater than the number of elements of $\mathcal{M}$ and therefore for every structure M from $\mathbb{M}_n$ we have $M \sqsupseteq \mathcal{M}$ if and only if $M \sqsupset \mathcal{M}$. We denote by $\sqsubseteq$ and $\sqsubset$ the relations inverse to $\sqsupseteq$ and $\sqsupset$, respectively.

Notation. For a structure M denote by d_M the number of elements of the domain of M and by r_M the number of tuples in R_M. We say that a tuple of elements of M is a *constant tuple* if all its elements are constants (from Σ), and a *non-constant tuple* otherwise. Let $\bar{d}_M$ denote the number of non-constant elements of M, $\bar{r}_M$ denote the number of non-constant tuples in R_M, and $\hat{r}_M$ denote the number of constant tuples in R_M.

First we show exponential bounds for the probability of the random structure $M_{n,p}$ to contain $\mathcal{M}$. In proofs we will use some techniques developed in the theory of random graphs (see [11]).

Exponential bounds. Let us note that if $R_{\mathcal{M}}$ is the empty relation then all structures with a sufficiently large domain contain $\mathcal{M}$, so the containment problem is trivial. Furthermore if all tuples in $R_{\mathcal{M}}$ are constant (i.e. $\hat{r}_{\mathcal{M}} = r_{\mathcal{M}}$) then $\mu(M_{n,p} \sqsupset \mathcal{M}) = p^{r_{\mathcal{M}}}$. In this case, trivially, the threshold function for the containment property is just a constant function and the threshold is coarse. We consider the case when $R_{\mathcal{M}}$ has at least one constant tuple later in Corollary 3.

Now assume that $R_{\mathcal{M}}$ is non-empty and contains no constant tuples. Define

$$\Phi_{\mathcal{M}} = \Phi_{\mathcal{M}}(n,p) = \min_{Q \sqsubseteq \mathcal{M}, r_Q > 0} n^{\bar{d}_Q} p^{r_Q}. \tag{1}$$

Then the following theorem holds.

Theorem 3. *Let $\mathcal{M}$ be a structure in the signature Σ such that $R_{\mathcal{M}}$ is non-empty and without constant tuples. Then, for every sufficiently large n and for every sequence $p = p(n) < 1$ the following holds:*

$$1 - \exp\left\{-\Theta(\Phi_{\mathcal{M}})\right\} \leq \mu(M_{n,p} \sqsupset \mathcal{M}) \leq 1 - \exp\left\{-\frac{\Theta(\Phi_{\mathcal{M}})}{1-p}\right\}. \tag{2}$$

Proof. Let us rewrite (2) in a more convenient for us form:

$$\exp\left\{-\frac{\Theta(\Phi_{\mathcal{M}})}{1-p}\right\} \leq \mu(M_{n,p} \not\sqsupset \mathcal{M}) \leq \exp\left\{-\Theta(\Phi_{\mathcal{M}})\right\}. \tag{3}$$

We will use the following inequality (4) (see e.g. [11] where this inequality is proved in a more general setting). We say that a structure M *extends* a structure M' if they have the same domain and $R_{M'} \subseteq R_M$. Consider a family S of structures on the domain $[n]$ and such that for each structure $M \in S$ the relation R_M is non-empty. For each structure $M \in S$, let I_M be the indicator variable denoting that the random structure $M_{n,p}$ extends M. Let $X_S = \sum_{M \in S} I_M$, so X_S denotes the number of structures from S which are extended by $M_{n,p}$. Then the following holds:

$$\exp\left\{-\frac{\mathbb{E}(X_S)}{1-p}\right\} \leq \mu(X_S = 0) \leq \exp\left\{-\frac{(\mathbb{E}(X_S))^2}{\sum_{M',M'' \in S \wedge R_{M'} \cap R_{M''} \neq \emptyset} \mathbb{E}(I_{M'} I_{M''})}\right\}. \quad (4)$$

Now let us consider a structure M such that $d_M \leq n$. We call an *n-copy* of M any structure M' with the domain $[n]$ which contains M and has a minimal $R_{M'}$, i.e., after removing any tuple from $R_{M'}$, M' will not contain M. When n is clear from the context we say a copy of M instead of an n-copy. It is clear that a structure with the domain $[n]$ contains M if and only if it contains an n-copy of M. Let X_M be the random variable denoting the number of different n-copies of M in the random structure $M_{n,p}$. Let us calculate $\mathbb{E}(X_M)$. There are exactly $f(n, M) = \binom{n-l}{\bar{d}_M} \bar{d}_M!/aut(M) = \Theta(n^{\bar{d}_M})$ different n-copies of M, where $aut(M)$ is the number of automorphisms of M, (remember that l is the number of constants in the signature). Using the linearity of the expectation we have

$$\mathbb{E}(X_M) = f(n, M)p^{r_M} = \Theta(n^{\bar{d}_M} p^{r_M}). \quad (5)$$

From this it follows that

$$\Phi_{\mathcal{M}} \asymp \min_{Q \sqsubseteq \mathcal{M}, r_Q > 0} \mathbb{E}(X_Q). \quad (6)$$

We will prove the left-hand side of (3) using the left-hand side of (4). To do so we take a structure $H \sqsubseteq \mathcal{M}$ such that $\mathbb{E}(X_H) = \min_{Q \sqsubseteq \mathcal{M}, r_Q > 0} \mathbb{E}(X_Q)$ and consider the family S_H of all n-copies of H. Then, it is easy to see that $X_{S_H} = X_H$. From (4) using (6) we have

$$\exp\left\{-\frac{\Theta(\Phi_{\mathcal{M}})}{1-p}\right\} \leq \exp\left\{-\frac{\mathbb{E}(X_H)}{1-p}\right\} \leq \mu(X_H = 0) = \mu(M_{n,p} \not\supseteq H). \quad (7)$$

It is obvious that if a structure does not contain a copy of H then it does not contain a copy of $\mathcal{M}$ and hence $\mu(M_{n,p} \not\supseteq H) \leq \mu(M_{n,p} \not\supseteq \mathcal{M})$. Therefore the left-hand side of (3) follows from (7).

Let us prove the right-hand side of (3). To this end, we use the right-hand side of (4) for the family $S_{\mathcal{M}}$ of all n-copies of $\mathcal{M}$. Again we have $X_{S_{\mathcal{M}}} = X_{\mathcal{M}}$. Now we estimate the sum in the denominator of the exponent in (4). Let M', M'' be structures with the domain $[n]$, then we can define a new structure $M' \cap M''$ to be a structure with the domain $[n]$ and the relation $R_{M' \cap M''} = R_{M'} \cap R_{M''}$. For each structure $Q \sqsubseteq \mathcal{M}$ there are $\Theta(n^{\bar{d}_Q} n^{2(\bar{d}_{\mathcal{M}} - \bar{d}_Q)}) = \Theta(n^{2\bar{d}_{\mathcal{M}} - \bar{d}_Q})$ pairs (M', M'') such that $M', M'' \in S_{\mathcal{M}}$ and $M' \cap M''$ is isomorphic to an n-copy of Q. So using (6) we have

$$\begin{aligned} \sum_{M',M'' \in S_{\mathcal{M}} \wedge R_{M'} \cap R_{M''} \neq \emptyset} \mathbb{E}(I_{M'} I_{M''}) &\asymp \sum_{Q \sqsubseteq \mathcal{M}, r_Q > 0} n^{2\bar{d}_{\mathcal{M}} - \bar{d}_Q} p^{2r_{\mathcal{M}} - r_Q} \\ &\asymp \max_{Q \sqsubseteq \mathcal{M}, r_Q > 0} \frac{(\mathbb{E}(X_{\mathcal{M}}))^2}{\mathbb{E}(X_Q)} \asymp \frac{(\mathbb{E}(X_{\mathcal{M}}))^2}{\Phi_{\mathcal{M}}}. \end{aligned} \quad (8)$$

Direct substitution of (8) into (4) gives us the right-hand side of (3).

Let $\mathcal{M}$ be a structure such that $R_{\mathcal{M}}$ is non-empty and without constant tuples. Then, we define the *density* of the structure $\mathcal{M}$ to be

$$m(\mathcal{M}) = \max_{Q \sqsubseteq \mathcal{M}, r_Q > 0} r_Q / \bar{d}_Q.$$

Corollary 1. *Let $\mathcal{M}$ be a structure such that $R_{\mathcal{M}}$ is non-empty and without constant tuples. Then the threshold function for the $\mathcal{M}$-containment property is $n^{-1/m(\mathcal{M})}$, i.e., the following holds:*

$$\lim_{n\to\infty} \mu(M_{n,p} \sqsupset \mathcal{M}) = \begin{cases} 0, \text{ if } \ p \ll n^{-1/m(\mathcal{M})}, \\ 1, \text{ if } \ p \gg n^{-1/m(\mathcal{M})}. \end{cases}$$

Proof. To prove this corollary it is sufficient to prove that (i) if $pn^{1/m(\mathcal{M})} \to \infty$ then $\Phi_{\mathcal{M}} \to \infty$ and (ii) if $pn^{1/m(\mathcal{M})} \to 0$ then $\Phi_{\mathcal{M}} \to 0$. Assume that $pn^{1/m(\mathcal{M})} \to \infty$. Then for every structure $Q \sqsubseteq \mathcal{M}$ with $r_Q > 0$ we have $1/m(\mathcal{M}) \leq \bar{d}_Q / r_Q$ and hence

$$n^{\bar{d}_Q} p^{r_Q} = (pn^{\bar{d}_Q / r_Q})^{r_Q} \to \infty.$$

Therefore $\Phi_{\mathcal{M}} = \min_{Q \sqsubseteq \mathcal{M}, r_Q > 0} n^{\bar{d}_Q} p^{r_Q} \to \infty$, which proves (i).

Now let H be a structure contained in $\mathcal{M}$ such that $r_H / \bar{d}_H = m(\mathcal{M})$. To prove (ii) notice that

$$\Phi_{\mathcal{M}} = \min_{Q \sqsubseteq \mathcal{M}, r_Q > 0} n^{\bar{d}_Q} p^{r_Q} \leq n^{\bar{d}_H} p^{r_H} = (pn^{1/m(\mathcal{M})})^{r_H} \to 0.$$

Corollary 2. *Let $\mathcal{M}$ be a structure such that $R_{\mathcal{M}}$ is non-empty and without constant tuples. Then the threshold for the $\mathcal{M}$-containment property is coarse.*

Proof. If $\mathcal{M}$-containment would have a sharp threshold then $\lim_{n\to\infty} p_\epsilon(n)/p_{1/2}(n) = 1$ for every ϵ such that $0 < \epsilon < 1$. Let us show that this is not the case. Take an arbitrary ϵ such that $0 < \epsilon < 1$. Let $p_\epsilon(n)$ be the tuple probability such that $\mu(M_{n,p_\epsilon} \sqsupset \mathcal{M}) = \epsilon$. Corollary 1 implies that $p_\epsilon(n) \to 0$ and therefore $p_\epsilon(n)$ is bounded away form 1. Using this and (2) we obtain that for some constants $A > 0$ and $B > 0$,

$$1 - \exp\{-A\Phi_{\mathcal{M}}\} \leq \mu(M_{n,p_\epsilon} \sqsupset \mathcal{M}) = \epsilon \leq 1 - \exp\{-B\Phi_{\mathcal{M}}\} \tag{9}$$

for all sufficiently large n.

Now consider an ϵ such that

$$0 < \epsilon < 1 - 2^{-2^{-r_{\mathcal{M}}} A/B}. \tag{10}$$

It is straightforward to check that $0 < \epsilon < 1$. Let us show that for this ϵ we have $p_\epsilon(n)/p_{1/2}(n) \not\to 1$. To avoid technicalities with $\Phi_{\mathcal{M}}(p_\epsilon, n) = \min_{Q \sqsubseteq \mathcal{M}, r_Q > 0} n^{\bar{d}_Q} p_\epsilon^{r_Q}$ we consider a structure $W \sqsubseteq \mathcal{M}$ on which the minimum is reached infinitely often. So let $\{n_i | i \in \mathbb{N}\}$ be an infinite subset of $\mathbb{N}$ such that $\Phi_{\mathcal{M}}(p_\epsilon, n_i) = n_i^{\bar{d}_W} p_\epsilon^{r_W}$. It is clear that for our goal it is enough to prove that $p_\epsilon(n_i)/p_{1/2}(n_i)$ is bounded away from 1. In order to prove it we show lower and upper bounds for $p_\epsilon(n_i)$.

From the right-hand side of (9) we have that

$$\epsilon = \mu(M_{n_i,p_\epsilon} \sqsupset \mathcal{M}) \leq 1 - \exp\{-B(\Phi_{\mathcal{M}})\}.$$

Straightforward calculations yield

$$\ln(1-\epsilon)^{-1/B} \leq n_i^{\bar{d}_W} p_\epsilon^{r_W}.$$

Finally, the lower bound is

$$\frac{\left(\ln(1-\epsilon)^{-1/B}\right)^{1/r_W}}{n_i^{\bar{d}_W/r_W}} \leq p_\epsilon(n_i). \tag{11}$$

Obtaining an upper bound for $p_\epsilon(n_i)$ is similar.

The left-hand side of (9) yields

$$1 - \exp\{-A\Phi_{\mathcal{M}}\} \leq \mu(M_{n_i,p_\epsilon} \sqsupset \mathcal{M}) = \epsilon.$$

After straightforward calculations we obtain

$$n_i^{\bar{d}_W} p_\epsilon^{r_W} \leq \ln(1-\epsilon)^{-1/A},$$

so we have an upper bound

$$p_\epsilon(n_i) \leq \frac{\left(\ln(1-\epsilon)^{-1/A}\right)^{1/r_W}}{n_i^{\bar{d}_W/r_W}}. \tag{12}$$

Now from (11),(12) and (10) we have

$$p_\epsilon(n_i)/p_{1/2}(n_i) \leq \left(\frac{\ln(1-\epsilon)^{-1/A}}{\ln(1/2)^{-1/B}}\right)^{1/r_W} = \left(\frac{B\ln(\frac{1}{1-\epsilon})}{A\ln(2)}\right)^{1/r_W} \leq$$

$$\left(\frac{B\ln\left(\frac{1}{1-\left(1-2^{-2^{-r_{\mathcal{M}}}A/B}\right)}\right)}{A\ln(2)}\right)^{1/r_W} = (1/2)^{r_{\mathcal{M}}/r_W} \leq 1/2,$$

therefore $p_\epsilon(n_i)/p_{1/2}(n_i)$ is bounded away from 1.

Corollary 3. *Let $\mathcal{M}$ be a structure such that $R_{\mathcal{M}}$ is non-empty and contains at least one constant tuple. Then the threshold function for the $\mathcal{M}$-containment property is constant and the threshold is coarse.*

Proof. Let $p_\epsilon(n)$ be the tuple probability such that $\mu(M_{n,p_\epsilon} \sqsupset \mathcal{M}) = \epsilon$. Let us show that for every $0 < \epsilon < 1$ we have $\lim_{n\to\infty} p_\epsilon(n) = \epsilon^{1/\hat{r}_{\mathcal{M}}}$. From this, the corollary easily follows.

Let $\hat{\mathcal{M}}$ be a structure with the same domain as $\mathcal{M}$ and the relation $R_{\hat{\mathcal{M}}}$ consisting of all constant tuples of $R_{\mathcal{M}}$; likewise let $\bar{\mathcal{M}}$ be a structure with the same domain as $\mathcal{M}$ and the relation $R_{\bar{\mathcal{M}}}$ consisting of all non-constant tuples of $R_{\mathcal{M}}$. Then we have $R_{\hat{\mathcal{M}}} \cap R_{\bar{\mathcal{M}}} = \emptyset$ and $R_{\hat{\mathcal{M}}} \cup R_{\bar{\mathcal{M}}} = R_{\mathcal{M}}$. It is easy to see that $M_{n,p}$ contains $\mathcal{M}$ if and only if $M_{n,p}$ contains both $\hat{\mathcal{M}}$ and $\bar{\mathcal{M}}$. Also, since $R_{\hat{\mathcal{M}}}$ and $R_{\bar{\mathcal{M}}}$ are disjoint, the containment of $\hat{\mathcal{M}}$ is independent from the containment of $\bar{\mathcal{M}}$ and therefore

$$\mu(M_{n,p} \sqsupset \mathcal{M}) = \mu(M_{n,p} \sqsupset \hat{\mathcal{M}})\mu(M_{n,p} \sqsupset \bar{\mathcal{M}}).$$

Since all tuples in $R_{\hat{\mathcal{M}}}$ constant, we have $\mu(M_{n,p} \sqsupset \hat{\mathcal{M}}) = p^{\hat{r}_{\mathcal{M}}}$, then the formula above gives

$$\mu(M_{n,p} \sqsupset \mathcal{M}) = p^{\hat{r}_{\mathcal{M}}}\mu(M_{n,p} \sqsupset \bar{\mathcal{M}}). \tag{13}$$

There are two possible cases.

1. *The relation $R_{\bar{\mathcal{M}}}$ contains no tuples (that is, all tuples in $R_{\mathcal{M}}$ are constant).* In this case $p_\epsilon(n) = \epsilon^{1/\hat{r}_{\mathcal{M}}}$, so the threshold function is constant and the threshold is coarse.
2. *$R_{\bar{\mathcal{M}}}$ contains at least one tuple.* In this case we can apply Theorem 3 in the following way. First, from (13) we have

$$\epsilon = \mu(M_{n,p_\epsilon} \sqsupset \mathcal{M}) = p_\epsilon^{\hat{r}_{\mathcal{M}}}\mu(M_{n,p_\epsilon} \sqsupset \bar{\mathcal{M}}) \leq p_\epsilon^{\hat{r}_{\mathcal{M}}}.$$

   Therefore

$$\epsilon^{1/\hat{r}_{\mathcal{M}}} \leq p_\epsilon(n) \tag{14}$$

   for all sufficiently large n. From the left-hand side of (2) and (13) we have

$$p_\epsilon^{\hat{r}_{\mathcal{M}}}(1 - \exp\{-\Theta(\Phi_{\bar{\mathcal{M}}})\}) \leq p_\epsilon^{\hat{r}_{\mathcal{M}}}\mu(M_{n,p_\epsilon} \sqsupset \bar{\mathcal{M}}) = \mu(M_{n,p_\epsilon} \sqsupset \mathcal{M}) = \epsilon. \tag{15}$$

   Since $p_\epsilon(n)$ is bounded from below we have $\lim_{n\to\infty}(1 - \exp\{-\Theta(\Phi_{\bar{\mathcal{M}}})\}) = 1$. Define $g(n) = (1 - \exp\{-\Theta(\Phi_{\bar{\mathcal{M}}})\})^{-1/\hat{r}_{\mathcal{M}}}$. It is clear that $\lim_{n\to\infty} g(n) = 1$. From (14) and (15) we have

$$\epsilon^{1/\hat{r}_{\mathcal{M}}} \leq p_\epsilon(n) \leq \epsilon^{1/\hat{r}_{\mathcal{M}}}g(n).$$

   Therefore $\lim_{n\to\infty} p_\epsilon(n) = \epsilon^{1/\hat{r}_{\mathcal{M}}}$. So the threshold function for the $\mathcal{M}$-containment property is constant. The threshold for the $\mathcal{M}$-containment is trivially coarse since for $\epsilon \neq 1/2$ we have $\lim_{n\to\infty} p_\epsilon(n)/p_{1/2} = (2\epsilon)^{1/\hat{r}_{\mathcal{M}}} \neq 1$.

5 Threshold for the Weak Containment

In this section we study the weak containment property for an arbitrary but fixed finite family of structures S, show how to calculate the threshold function for this property, and show that it has a coarse threshold.

It is clear that if there is a structure $M \in S$ such that R_M is empty then the weak S-containment property is trivial since all structures with a sufficiently large domain weakly contain S. Therefore, we assume that for each structure $M \in S$ the relation R_M

is non-empty. Let $\bar{S}$ denote the set of all structures M from S such that R_M contains no constant tuples.

Let $\bar{S} \neq \emptyset$. Define *density of* S to be

$$m(S) = \min_{\mathcal{M} \in \bar{S}} m(\mathcal{M}).$$

Then the following holds.

Corollary 4. *Let S be a finite set of structures such that for each $M \in S$ the relation R_M is non-empty. Suppose that $\bar{S} \neq \emptyset$. Then the threshold function for the weak S-containment property is $n^{-1/m(S)}$ and the threshold is coarse.*

Proof. From Corollary 1 we immediately obtain the following. If $p \ll n^{-1/m(S)}$, then for every structure $\mathcal{M}$ in S we have $\lim_{n\to\infty} \mu(M_{n,p} \sqsupset \mathcal{M}) = 0$ and therefore $\lim_{n\to\infty} \mu(M_{n,p} \sqsupset_w S) = 0$. If $p \gg n^{-1/m(S)}$, then we can take a structure $\mathcal{M}$ from S with the least $m(\mathcal{M})$; then we have $\lim_{n\to\infty} \mu(M_{n,p} \sqsupset \mathcal{M}) = 1$ and therefore $\lim_{n\to\infty} \mu(M_{n,p} \sqsupset_w S) = 1$. So we have proved that the threshold function for the weak S-containment property is $n^{-1/m(S)}$.

Let us prove that this threshold is coarse. The proof will be similar to the proof of Corollary 2. Consider any ϵ such that $0 < \epsilon < 1$. Let $p_\epsilon(n)$ be the tuple probability such that $\mu(M_{n,p_\epsilon} \sqsupset_w S) = \epsilon$. Consider a structure $\mathcal{M}$ from $\bar{S}$. From (2) it follows that for some constants $A_{\mathcal{M}} > 0$ and $B_{\mathcal{M}} > 0$ we have

$$1 - \exp\{-A_{\mathcal{M}} \Phi_{\mathcal{M}}\} \leq \mu(M_{n,p_\epsilon} \sqsupset \mathcal{M}) = \epsilon \leq 1 - \exp\{-B_{\mathcal{M}} \Phi_{\mathcal{M}}\} \qquad (16)$$

for every $0 < \epsilon < 1$ and sufficiently large n.

Let $A = \min_{\mathcal{M}\in\bar{S}} A_{\mathcal{M}}$, $B = \max_{\mathcal{M}\in\bar{S}} A_{\mathcal{M}}$, $E = \max_{\mathcal{M}\in\bar{S}} r_{\mathcal{M}}$, and $L = |S|$ where $|S|$ is the number of structures in S. Consider an ϵ such that

$$0 < \epsilon < 1 - (1 - 1/2L)^{2^{-E} A/B}. \qquad (17)$$

It is straightforward to check that $0 < \epsilon < 1$. Let us show that for this ϵ we do not have $p_\epsilon(n)/p_{1/2}(n) \to 1$. We have

$$\max_{\mathcal{M}\in S} \mu(M_{n,p_\epsilon} \sqsupset \mathcal{M}) \leq \mu(M_{n,p_\epsilon} \sqsupset_w S) = \epsilon \leq \sum_{\mathcal{M}\in S} \mu(M_{n,p_\epsilon} \sqsupset \mathcal{M}) \leq L \max_{\mathcal{M}\in S} \mu(M_{n,p_\epsilon} \sqsupset \mathcal{M}).$$

Let V be a structure in S on which the maximum $\max_{\mathcal{M}\in S} \mu(M_{n,p_\epsilon} \sqsupset \mathcal{M})$ is reached infinitely often. Let $\{n_i\}$ be an infinite subset of $\mathbb{N}$ such that

$$\mu(M_{n_i,p_\epsilon} \sqsupset V) \leq \mu(M_{n_i,p_\epsilon} \sqsupset_w S) = \epsilon \leq L\mu(M_{n_i,p_\epsilon} \sqsupset V). \qquad (18)$$

Let us show that $V \in \bar{S}$. Suppose, by contradiction, that R_V has at least one constant tuple. From the previous we have that the threshold function for the weak S-containment property is $n^{-1/m(S)}$ and therefore $p_\epsilon(n_i) \to 0$. Since R_V has constant tuples we have that the threshold function for the V-containment property is constant (see Corollary 3) and therefore $\mu(M_{n_i,p_\epsilon} \sqsupset V) \to 0$. This contradicts to the inequality on the right in (18).

Now we have that R_V has no constant tuples and we can proceed similarly to Corollary 2. Again we consider a structure $W \sqsubseteq V$ on which the minimum $\Phi_{\mathcal{M}}(p_\epsilon, n_i) = \min_{Q \sqsubseteq V, r_Q > 0} n_i^{\bar{d}_Q} p_\epsilon^{r_Q}$ is reached infinitely often. So let $\{n_{i_j} | i, j \in \mathbb{N}\}$ be an infinite subset of $\{n_i\}$ such that $\Phi_{\mathcal{M}}(p_\epsilon, n_{i_j}) = n_{i_j}^{\bar{d}_W} p_\epsilon^{r_W}$. It is clear that for our goal it is enough to prove that $p_\epsilon(n_{i_j})/p_{1/2}(n_{i_j})$ is bounded away from 1. In the same way as in Corollary 2 we calculate lower and upper bounds for $p_\epsilon(n_{i_j})$ as

$$\frac{\left(\ln(1-\epsilon/L)^{-1/B_V}\right)^{1/r_W}}{n_{i_j}^{\bar{d}_W/r_W}} \leq p_\epsilon(n_{i_j}) \leq \frac{\left(\ln(1-\epsilon)^{-1/A_V}\right)^{1/r_W}}{n_{i_j}^{\bar{d}_W/r_W}}. \tag{19}$$

Now from (19) and (17) we have

$$p_\epsilon(n_{i_j})/p_{1/2}(n_{i_j}) \leq \left(\frac{\ln(1-\epsilon)^{-1/A_V}}{\ln(1-1/2L)^{-1/B_V}}\right)^{1/r_W} = \left(\frac{B_V \ln(\frac{1}{1-\epsilon})}{A_V \ln(\frac{1}{1-1/2L})}\right)^{1/r_W} \leq$$

$$\left(\frac{B_V \ln\left(\frac{1}{1-\left(1-(1-1/2L)^{2^{-E}A/B}\right)}\right)}{A_V \ln(\frac{1}{1-1/2L})}\right)^{1/r_W} = \left(\frac{B_V A 2^{-E}}{A_V B}\right)^{1/r_W} \leq (1/2)^{E/r_w} \leq 1/2.$$

Therefore $p_\epsilon(n_{i_j})/p_{1/2}(n_{i_j})$ is bounded away from 1.

Corollary 5. *Let S be a finite set of structures such that for each $M \in S$ the relation R_M is non-empty. Suppose that $\bar{S} = \emptyset$. Then the threshold function for the weak S-containment property is constant and the threshold is coarse.*

Proof. We have that for all structures $M \in S$ the relation R_M contains at least one constant tuple. Corollary 3 implies that for $p \to 0$ and every structure in S we have $\lim_{n\to\infty} \mu(M_{n,p} \sqsupset \mathcal{M}) = 0$ and therefore $\lim_{n\to\infty} \mu(M_{n,p} \sqsupset_w S) = 0$. This proves that the threshold function is constant.

In order to prove that the threshold is coarse we consider any ϵ such that

$$0 < \epsilon < 2^{-2E-1}/L. \tag{20}$$

where L is the number of structures in S and $E = \max_{\mathcal{M}\in S} \hat{r}_{\mathcal{M}}$. Let $p_\epsilon(n)$ be the tuple probability such that $\mu(M_{n,p_\epsilon} \sqsupset_w S) = \epsilon$. We have

$$\max_{\mathcal{M}\in S} \mu(M_{n,p_\epsilon} \sqsupset \mathcal{M}) \leq \mu(M_{n,p_\epsilon} \sqsupset_w S) = \epsilon \leq \sum_{\mathcal{M}\in S} \mu(M_{n,p_\epsilon} \sqsupset \mathcal{M}) \leq L \max_{\mathcal{M}\in S} \mu(M_{n,p_\epsilon} \sqsupset \mathcal{M}).$$

Let V be a structure in S on which the maximum $\max_{\mathcal{M}\in S} \mu(M_{n,p_\epsilon} \sqsupset \mathcal{M})$ is reached infinitely often. Let $\{n_i\}$ be an infinite subset of $\mathbb{N}$ such that

$$\mu(M_{n_i,p_\epsilon} \sqsupset V) \leq \mu(M_{n_i,p_\epsilon} \sqsupset_w S) = \epsilon \leq L\mu(M_{n_i,p_\epsilon} \sqsupset V). \tag{21}$$

In the same way as in Corollary 3 we calculate lower and upper bounds for $p_\epsilon(n_i)$ as

$$(\epsilon/L)^{1/\hat{r}_V} \leq p_\epsilon(n_i) \leq \epsilon^{1/\hat{r}_V} g(n_i), \tag{22}$$

where $g(n_i) \to 1$. For sufficiently large n_i we can replace $g(n_i)$ in (22) with 2, obtaining

$$(\epsilon/L)^{1/\hat{r}_V} \leq p_\epsilon(n_i) \leq 2\epsilon^{1/\hat{r}_V}.$$

Now using (20), we have

$$p_\epsilon(n_i)/p_{1/2}(n_i) \leq \frac{2\epsilon^{1/\hat{r}_V}}{(1/2L)^{1/\hat{r}_V}} \leq 2^{-2E/\hat{r}_V} 2 \leq 1/2.$$

Therefore $p_\epsilon(n_i)/p_{1/2}(n_i)$ is bounded away from 1.

6 Main Results

We can now put together all the presented results as follows. First we introduce the key notion of *density of a query*. Let φ be a monotone non-recursive datalog$^{\neq}$ query. Then, by Theorem 1 it is equivalent to a sentence in the language $\Sigma_1^{\neq}$. By Theorem 2 such a query is also equivalent to the weak S_φ-structure containment problem for a finite family of structures S_φ. Moreover such a family can be found effectively from φ. If every structure in S_φ contains a constant tuple ($\bar{S}_\varphi = \emptyset$) then the threshold function for the weak S-containment property is constant and the threshold is coarse. Otherwise we define *density* of φ, denoted $m(\varphi)$, to be the density of S_φ, (see Section 5 for the definition of the density for families of structures). From the above it follows that the density of a query can be calculated effectively. Now we are ready to formulate our main theorem.

Theorem 4. *Given a monotone non-recursive datalog$^{\neq}$-query φ one can effectively find the threshold function for the property defined by this query. This threshold function is either constant or has the form $n^{-1/m(\varphi)}$, where $m(\varphi)$ is the density of φ. The density of a query is always a positive rational constant which can be effectively calculated from φ. For every monotone non-recursive datalog$^{\neq}$-query the threshold is coarse.*

Now we give a simple application of Theorem 4 to $0-1$ laws for non-recursive monotone datalog$^{\neq}$. Let us fix a function $0 < p(n) < 1$. We say that non-recursive monotone datalog$^{\neq}$ has a $0-1$ law w.r.t. $p(n)$ if for every boolean query φ expressible in it, $\lim_{n\to\infty} \mu(M_{n,p} \models \varphi)$ equals either to 0 or 1.

Theorem 5. *Monotone non-recursive datalog$^{\neq}$ has a $0-1$ law w.r.t. every probability function $p(n)$ satisfying the following condition: for every rational $q > 0$ either $p \gg n^{-q}$ or $p \ll n^{-q}$ holds.*

Proof. Indeed, from Theorem 4 it follows that for every non-recursive monotone datalog$^{\neq}$ query φ, the threshold function is either constant or has the form n^{-q}, for a rational $q > 0$. Therefore, from the definition of the threshold function follows that for any such query φ, and $p(n)$ as in the statement of the theorem, $\lim_{n\to\infty} \mu(M_{n,p} \models \varphi)$ is either 0 or 1.

For example the theorem holds for $p(n) = n^{-\alpha}$ where $\alpha > 0$ is irrational and also for functions like $ln(n)n^{-t}$ where $t > 0$.

Let us note that $0 - 1$ laws w.r.t. irrational powers of $1/n$ are proved for the full first-order logic on random graphs in [16] see also [17]. For general accounts on $0 - 1$ laws for various logics see [5, 14], for some applications of $0 - 1$ laws to the database theory see, e.g., [15].

Acknowledgements. We are grateful to Evgeny Dantsin for introducing us to the fascinating area of randomness and to Leonid Libkin and Moshe Vardi for providing useful references.

References

1. B. Bollobás. Random graphs. In *Combinatorics (Swansea, 1981)*, volume 52 of *London Math. Soc. Lecture Note Ser.*, pages 80–102. Cambridge Univ. Press, 1981.
2. B. Bollobás. *Random graphs*. Academic Press Inc., London, 1985.
3. B. Bollobás and A. Thomason. Threshold functions. *Combinatorica*, 7:35–38, 1987.
4. M. de Rougemont. The reliability of queries. In *Proceedings of the Fourteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, May 22-25, 1995, San Jose, California*, pages 286–291. ACM Press, 1995.
5. H.-D. Ebbinghaus and J. Flum. *Finite Model Theory*. Perspectives in Mathematical Logic. Springer, 1999.
6. P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Kőzl.*, 5:17–61, 1960.
7. R. Fagin. Probabilities on finite models. *Journal of Symbolic Logic*, 41:50–58, 1976.
8. E. Friedgut. Sharp thresholds of graph properties, and the k-sat problem. *J. Amer. Math. Soc.*, 12(4):1017–1054, 1999.
9. Y. Glebskii, M. Kogan, M. Liogonkii, and V. Talanov. Range and degree of realizability of formulas in the restricted predicate calculus. *Kibernetika*, 5:17–27, 1969. (in Russian); English translation in Cybernetics 5, 142–154, 1969.
10. E. Grädel, Y. Gurevich, and C. Hirsch. The complexity of query reliability. In *Proceedings of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 1-3, 1998, Seattle, Washington*, pages 227–234. ACM Press, 1998.
11. S. Janson, T. Łuczak, and A. Ruciński. *Random graphs*. John Wiley & Sons, Inc., 2000.
12. K. Korovin and A. Voronkov. Random databases and threshold for monotone non-recursive datalog. Preprint, School of Computer Science, The University of Manchester, 2005.
13. R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. The web as a graph. In *Proceedings of the Nineteenth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, May 15-17, 2000, Dallas, Texas, USA*, pages 1–10. ACM, 2000.
14. L. Libkin. *Elements of Finite Model Theory*. Texts in Theoretical Computer Science. Springer, 2004.
15. S. Lifschitz and V. Vianu. A probabilistic view of Datalog parallelization. *Theoretical Computer Science*, 190(2):211–239, 1998.
16. S. Shelah and J. Spencer. Zero one laws for sparse random graphs. *Journal of the AMS*, 1(1):97–115, 1988.
17. J. Spencer. *The Strange Logic of Random Graphs*, volume 22 of *Algorithms and Combinatorics*. Springer, 2001.