A zero-one law for random sentences in description logics

Bernard Ycart(*), Marie-Christine Rousset(**),
(*) UFR Math-Info, 45 rue des Saints-Pères
75270 Paris Cedex 06, France
ycart@math-info.univ-paris5.fr
(**) LRI, UMR 8623 CNRS, University of Paris-Sud
91405 Orsay Cedex, France
mcr@lri.fr

Abstract

We consider the set of sentences in a decidable fragment of first order logic, restricted to unary and binary predicates, expressed on a finite set of objects. Probability distributions on that set of sentences are defined and studied. For large sets of objects, it is shown that under these distributions, random sentences typically have a very particular form, and that all monotone symmetric properties of these random sentences have a probability close to 0 or 1.

1 Introduction

Description logics (DL) are decidable fragments of first order logic with equality, which have recently received significant attention because of their relevance for modeling a wide variety of applications. A description logic is restricted to unary predicates (referred to as *concepts*), representing sets of objects, and binary predicates (called *roles*), representing relationships between objects. Various logical constructors can be used for defining concepts by imposing some restrictions on the number of fillers of a role, or on the concepts to which fillers of a certain role have to belong. Inference problems within DLs have been extensively studied and a whole family of DL systems have been implemented and used for building a variety of applications (see e.g. [16] for a survey on DLs, and http://dl.kr.org/ for the official description logics home page). In this paper, we consider sets of DL ground sentences that are referred to as *Aboxes* in the DL community. Given a set of concept definitions (called a *Tbox*), an Abox asserts facts about objects in the form of *concept facts* and *role facts*. A *concept fact* $C(o)$ states that the object specified by the constant o, is an instance of the concept C, while a *role fact* $R(o, o')$ states that the two objects o, o' are related by the role R.

For example, the following definition of a concept named C models the set of persons such that all of their children are male, and have at most 3 children, in the form of the sentence:

$$C := person \sqcap (\forall child.male) \sqcap (\leq 3\, child).$$

This definition illustrates some standard DL constructors: conjunction, number restrictions and type restrictions. Relatively to that concept definition, we may have an Abox containing the facts $child(Bob, Jacky)$ and $C(Bob)$: $child(Bob, Jacky)$ is a role fact saying that Jacky is Bob's child, and $C(Bob)$ is a concept fact saying that Bob is an instance of the concept C. From those facts, other facts can be logically inferred: for instance, we can infer the new concept fact $male(Jacky)$ as a logical consequence of the definition of concept C.

This study was motivated by the need for generating *random benchmarks* for DL systems. The advantage of generating random benchmarks is twofold. Firstly, it is a way to provide a great number of artificial knowledge bases whose structure and distribution can be tuned in order to possibly account for some known real-world distributions. Secondly, it makes it possible to perform average-case complexity analysis for DL algorithms, also called *probabilistic analysis* [11].

In this paper, we define and study probability distributions on the set of all satisfiable Aboxes associated to a given Tbox and a given set of objects. These distributions are obtained by conditioning products of Bernoulli measures on facts (definition 3.1). In practical applications the number of objects is usually larger by several order of magnitudes than the number of concepts and roles. It is therefore natural to study the asymptotics of our distributions as the number of objects tends to infinity. The concept definitions appearing in a Tbox entail exclusion rules between facts of any satisfiable Abox associated with it. It turns out that exclusion rules have a dramatic effects on the structure of 'typical' random Aboxes, for large object sets. Some concept facts totally disappear while the Abox tends to be almost deterministic, in the sense that all reasonable properties are either true or false with a probability tending to one. This *zero-one law* is the main result of the article. Similar results were known to hold for random graphs or other structures involving a high degree of independence (see for instance [17, 10]).

The paper is organized as follows. In section 2, we define Aboxes and their main logical property, satisfiability. The probability distributions on satisfiable Aboxes are the object of section 3. In section 4, we introduce a particular case that turns out to provide an asymptotic pattern for the general case. This particular case concerns Aboxes for which exclusion rules involve only concept facts, and are therefore called *role-free*. A zero-one law for probability distributions on role-free sets of Aboxes is stated in theorem 4.1, which is our main result. Finally the asymptotic properties of general random Aboxes is studied in section 5. There it will be shown that a typical random Abox with a large number of objects resembles a role-free random Abox, and therefore has almost deterministic properties.

2 Satisfiable Aboxes

We now formally define the syntax and semantics of the Aboxes that we consider in this paper. They are related to a fixed Tbox $\mathcal{T}$ which is a set of concept definitions in the $\mathcal{ALCN}$ description logic which is one of the most expressive decidable implemented description logic.

Concept expressions in $\mathcal{ALCN}$ are defined using the following syntax (A denotes a primitive concept name, C and D represent concept expressions and R denotes a role name):

$$
\begin{array}{lll}
C, D \rightarrow & A \mid & \text{(primitive concept)} \\
& C \sqcap D \mid C \sqcup D \mid & \text{(conjunction, disjunction)} \\
& \neg C \mid & \text{(complement)} \\
& \forall R.C \mid & \text{(type restriction)} \\
& \exists R.C \mid & \text{(existential quantification)} \\
& (\geq n\, R) \mid (\leq n\, R) & \text{(number restrictions)}
\end{array}
$$

A Tbox is a set of *concept definitions*, which are statements of the form $CN := D$, where CN is a concept name and D is a $\mathcal{ALCN}$ concept expression. We assume that a concept name appears on the left-hand side of at most one definition. A concept name A is said to *depend* on a concept name B if B appears in the concept definition of A. We consider concept definitions that do not have cycles in the dependency relation.

The semantics of the constructors is defined via *interpretations* in the standard way of first order logic. An interpretation I contains a non-empty domain $\mathcal{D}^I$. It assigns a unary relation over $\mathcal{D}^I$ to every concept name, and a binary relation over $\mathcal{D}^I \times \mathcal{D}^I$ to every role name. The interpretations of concept expressions are uniquely defined by the following equations: ($\sharp S$ denotes the cardinality of a set S):

$$
\begin{array}{l}
(C \sqcap D)^I = C^I \cap D^I, \\
(C \sqcup D)^I = C^I \cup D^I, \\
(\neg C)^I = \mathcal{D}^I \setminus C^I, \\
(\forall R.C)^I = \{d \in \mathcal{D}^I \mid \forall e[(d,e) \in R^I \rightarrow e \in C^I]\} \\
(\exists R.C)^I = \{d \in \mathcal{D}^I \mid \exists e[(d,e) \in R^I \wedge e \in C^I]\} \\
(\geq n\, R)^I = \{d \in \mathcal{D}^I \mid \sharp\{e \mid (d,e) \in R^I\} \geq n\} \\
(\leq n\, R)^I = \{d \in \mathcal{D}^I \mid \sharp\{e \mid (d,e) \in R^I\} \leq n\}
\end{array}
$$

An interpretation I is a model of a Tbox $\mathcal{T}$ if $C^I = D^I$ for every concept definition $C := D$ in $\mathcal{T}$. We say that C *is subsumed by* D w.r.t. $\mathcal{T}$ if $C^I \subseteq D^I$ in every model I of $\mathcal{T}$.

Given a fixed finite set of objects (constants) $\mathcal{O}$, and given a Tbox $\mathcal{T}$, an Abox is a set of facts of the form $C(o)$ (concept facts) or $R(o, o')$ (role facts), such that $o, o' \in \mathcal{A}$. Given an Abox $\mathcal{A}$, we extend an interpretation with a mapping ψ^I from the objects appearing in $\mathcal{A}$ to $\mathcal{D}^I$. An interpretation I is said to be a *model* of $\mathcal{A}$ if whenever $C(a) \in \mathcal{A}$ then $\psi^I(a) \in C^I$, and whenever $R(a,b) \in \mathcal{A}$ then $(\psi^I(a), \psi^I(b)) \in R^I$.

An Abox is *satisfiable* if it has a model, i.e, there exists an interpretation in which the conjunction of its facts is true. In the rest of the paper, we focus on satisfiable Aboxes since only satisfiable Aboxes are meaningful to capture realistic properties of data.

It is important to note that the concept definitions appearing in a Tbox entail exclusion rules between facts of any satisfiable Abox associated with $\mathcal{T}$. For instance, the presence of the concept definition:

$$C := person \sqcap (\forall child.male) \sqcap (\leq 3child)$$

in a Tbox imposes exclusion rules between the fact $C(o)$, the role facts $child(o, o')$, and possibly the concept facts involving the object o'. For instance, in a satisfiable Abox, we cannot have simultaneously the facts $C(o)$, $child(o, o_1)$, $child(o, o_2)$, $child(o, o_3)$, $child(o, o_4)$. Similarly, the facts $C(o)$, $child(o, o_1)$, $\neg male(o_1)$ cannot appear together in a satisfiable Abox.

Finally, note that since the semantics of DLs is defined in terms of interpretations in the standard way of first order logic, they obey the so-called *open world assumption*. The open world assumption leaves the possibility open that other unknown facts might be true besides those explicitly mentioned in the Abox.

3 Probability distribution on Aboxes

The set of objects $\mathcal{O} = \{o_1, \dots, o_n\}$ is fixed for the rest of this section.

Let c and r be the numbers of concepts and roles in $\mathcal{T}$. ¿From now on, $\widetilde{\mathcal{A}}$ will denote the set of all 2^{cn+rn^2} Aboxes, and $\mathcal{A} \subset \widetilde{\mathcal{A}}$ will denote the set of all *satisfiable* Aboxes, relative to the Tbox $\mathcal{T}$ and the set of objects $\mathcal{O}$. We want to define a family of product-type probability distributions on $\mathcal{A}$. To this end, we associate to each concept C a probability p_C and to each role R a probability p_R, all being strictly positive and lesser than 1. The parameters p_C and p_R may depend on the number of objects n. The set $\widetilde{\mathcal{A}}$ is first endowed with a product of Bernoulli distributions with parameters p_C and p_R, then the distribution on $\mathcal{A}$ is defined by conditioning.

Definition 3.1 *Let us consider on $\widetilde{\mathcal{A}}$ the product of Bernoulli distributions, denoted by $\widetilde{\mathbb{P}}$, and defined for all $A \in \widetilde{\mathcal{A}}$ by:*

$$\begin{aligned} \widetilde{\mathbb{P}}[A] &= \prod_{C,o} p_C^{\mathbb{1}_A(C(o))} (1 - p_C)^{1 - \mathbb{1}_A(C(o))} \\ &\quad \prod_{R,o,o'} p_R^{\mathbb{1}_A(R(o,o'))} (1 - p_R)^{1 - \mathbb{1}_A(R(o,o'))} . \end{aligned} \tag{3.1}$$

We call conditional product distribution (CPD) with parameters p_C and p_R, $C, R \in \mathcal{T}$ the distribution $\mathbb{P}$ on $\mathcal{A}$ defined for all $A \in \mathcal{A}$ by:

$$\mathbb{P}[A] = \frac{\widetilde{\mathbb{P}}[A]}{\widetilde{\mathbb{P}}[\mathcal{A}]} \, .$$

In formula (3.1), the first product extends over all concepts $C \in \mathcal{T}$ and all objects $o \in \mathcal{O}$. The second product extends over all roles $R \in \mathcal{T}$ and all couples of objects $(o, o') \in \mathcal{O} \times \mathcal{O}$. The notation $\mathbb{1}_A(.)$ is standard for the indicator function of the set A (1 if the element belongs to A, 0 else). Consider the event "$C(o) \in A$". Under the distribution $\widetilde{\mathbb{P}}$, it has probability p_C, whatever $o \in \mathcal{O}$. Similarly the event "$R(o, o') \in A$" has probability p_R, whatever (o, o'). Moreover all these events are mutually independent. When exclusion rules are present, $\mathcal{A}$ is only a subset of $\widetilde{\mathcal{A}}$, and independence between facts no longer holds. Notice that due to combinatorial explosion, $\widetilde{\mathbb{P}}[\mathcal{A}]$ cannot be explicitly computed in general. Nevertheless we shall see that explicitly computing $\widetilde{\mathbb{P}}[\mathcal{A}]$ can be avoided in an asymptotic description (section 5).

In the particular case where all the probabilities p_R and p_C are equal to 1/2, then all satisfiable Aboxes have the same probability under $\mathbb{P}$: $\mathbb{P}[A] = 1/\sharp\mathcal{A}$.

4 Role-free sets

This section deals with very particular Aboxes, namely those for which exclusion rules only concern concept facts.

Definition 4.1 *Let $\mathcal{T}$ be a Tbox, and $\mathcal{A}$ be the set of Aboxes associated to $\mathcal{T}$ and a given set of objects $\mathcal{O}$. The set $\mathcal{A}$ is said to be* role-free *if any role fact $R(o, o')$ can be added to a satisfiable Abox, while preserving satisfiability.*

In particular, if $\mathcal{A}$ is role free, then $\mathcal{T}$ cannot contain any concept defined by an upper-bound cardinality restriction ($C = \leq kR$), nor by a type restriction ($C' = \forall RC$). If $\mathcal{A}$ is role-free, then the distribution $\mathbb{P}$ of Definition 3.1 has a particularly simple product-type structure, which is described in Proposition 4.2. The main result, Theorem 4.1 proves that any symmetric monotone property of a random role-free Abox is either true or false with a probability close to 1 as the number of objects tends to infinity. Obviously, role-free sets of Aboxes are not the most interesting from the point of view of applications. However, it will be proved in section 5 that as the number of objects tends to infinity, most CPD's on general sets of Abox tend to resemble those described here in the role-free case.

We first need to define the *decomposition* of an Abox.

Definition 4.2 *Let A be an Abox.*

1. *Let o be an object. We call* object-component of A associated to o *the Abox, denoted by $A^{(o)}$, which is the subset of A containing only those facts of type $C(o)$, for all C.*

2. *Let R be a role. We call* role-component of A associated to R *the Abox, denoted by A_R, which is the subset of A containing only those facts of type $R(o, o')$, for all o, o'.*

Note that in Definition 4.2, some of the components may be empty. Obviously

$$\{ A^{(o)} \, , \, o \in \mathcal{O} \, ; \, A_R \, , \, R \text{ role } \in \mathcal{T} \} \, ,$$

is a partition of A, and A is uniquely defined by the $(r+n)$-tuple of its components. So the mapping that associates to a given Abox the $(r+n)$-tuple of its components is one-to-one, and it will be referred to as *decomposition.* For each object o, denote by $\mathcal{A}^{(o)}$ the set of all satisfiable Aboxes containing only concepts facts of type $C(o)$. For each role R, denote by $\mathcal{A}_R$ the set of all 2^{n^2} Aboxes containing only role facts of type $R(o, o')$. Role-free sets of Aboxes are actually Cartesian products of $\mathcal{A}^{(o)}$'s and $\mathcal{A}_R$'s.

Proposition 4.1 *The set $\mathcal{A}$ is role-free if and only the decomposition maps it onto the Cartesian product of all $\mathcal{A}^{(o)}$'s and $\mathcal{A}_R$'s. Moreover, Let $\phi^{(o,o')}$ be the mapping defined on $\mathcal{A}^{(o)}$, with values in $\mathcal{A}^{(o')}$, defined by:*

$$C(o) \in A \Longleftrightarrow C(o') \in \phi^{(o,o')}(A) \, .$$

If $\mathcal{A}$ is role-free, then all mappings $\phi^{(o,o')}$ are one-to-one.

Regarding CPD's, Proposition 4.1 yields probabilistic independence of components. Proposition 4.2 below describes the structure of CPD's on role-free sets of Aboxes.

Proposition 4.2 *For each object o, denote by $\mathbb{P}^{(o)}$ the conditional probability distribution of $\mathbb{P}$ over $\mathcal{A}^{(o)}$. For each role R, denote by $\mathbb{P}_R$ the conditional probability distribution of $\mathbb{P}$ over $\mathcal{A}_R$. If $\mathcal{A}$ is role-free then:*

1. *If $A \in \mathcal{A}$ is any satisfiable Abox with decomposition $\{A^{(o)} \, , \, o \in \mathcal{O} \, , \, A_R \, , \, R \text{ role } \in \mathcal{T}\}$, then:*

$$\mathbb{P}[A] = \prod_{o \in \mathcal{O}} \mathbb{P}^{(o)}[A^{(o)}] \prod_{R \in \mathcal{T}} \mathbb{P}_R[A_R] \, .$$

2. *Let o and o' be two objects. Then for all $A \in \mathcal{A}^{(o)}$:*

$$\mathbb{P}^{(o)}[A] = \mathbb{P}^{(o')}[\phi^{(o,o')}(A)] \, .$$

3. *Let R be a role. Let $A \in \mathcal{A}_R$ be any Abox containing m facts. Then:*

$$\mathbb{P}_R[A] = p_R^m (1 - p_R)^{n^2 - m} \, .$$

Thus in the decomposition of a random Abox under $\mathbb{P}$, all components are independent random Aboxes (item *1*). All object-components are identically distributed, up to identification of objects (item *2*). In each role-component, any role fact $R(o, o')$ occurs with probability p_R and all these facts are independent (item *3*).

In order to further understand the structure of CPD's on role-free sets of Aboxes, some properties of the marginal distributions $\mathbb{P}^{(o)}$ and $\mathbb{P}_R$ are worth pointing out. Let us begin with $\mathbb{P}^{(o)}$.

Let $\mathcal{C}$ be the set of concepts in $\mathcal{T}$. To an Abox $A \in \mathcal{A}^{(o)}$, one can uniquely associate a boolean vector, indexed by $\mathcal{C}$, that will be denoted by $\mathbb{1}_A$. For all $C \in \mathcal{C}$,

$$\mathbb{1}_A(C) = \begin{cases} 1 & \text{if} \quad C(o) \in A \ , \\ 0 & \text{if} \quad C(o) \notin A \ . \end{cases}$$

Since the distributions $\mathbb{P}^{(o)}$ do not depend on o (item *2* of Proposition 4.2), they all induce the same distribution, denoted by $\overline{\mathbb{P}}$, on boolean vectors.

$$\overline{\mathbb{P}}[\mathbb{1}_A] = \frac{1}{Z} \prod_{C \in \mathcal{C}} p_C^{\mathbb{1}_A(C)} (1 - p_C)^{1 - \mathbb{1}_A(C)} \ ,$$

where Z denotes the normalizing constant, i.e.:

$$Z = \sum_{A \in \mathcal{A}^{(o)}} \prod_{C \in \mathcal{C}} p_C^{\mathbb{1}_A(C)} (1 - p_C)^{1 - \mathbb{1}_A(C)} \ .$$

In the particular case where no exclusion rule involves concepts, all Aboxes are satisfiable, and the distribution $\overline{\mathbb{P}}$ is a product of Bernoulli distributions on $\{0,1\}^{\mathcal{C}}$. In the general case, $\overline{\mathbb{P}}$ is only a truncation of that product of Bernoulli distributions. It has already appeared in a different context as the reversible measure of some resource sharing models (see [7]). Indeed, one can interpret concepts as components in a system, and a boolean vector as a pattern of activity: $\mathbb{1}_A(C) = 1$ if component C is active, $\mathbb{1}_A(C) = 0$ if it is inactive. The parameter p_C should be seen as the proportion of time that component C spends in the active state, in the absence of constraints. However, components need their share of common resources to be active, so that in certain subsets, all components cannot be active at the same time. These constraints are the exclusion rules of the Tbox.

The particular case where all exclusion rules are binary, and all parameters p_C are equal, has been studied in more depth ([18, 8, 9]). It is connected to an interesting combinatorial problem, that of finding stable sets of an undirected graph. By "binary exclusion rules", we mean that two concepts may be mutually exclusive but it is impossible to find $k \geq 2$ concepts, any two of them are compatible but all k are not. Define an undirected graph structure, the vertices of which are the concepts in $\mathcal{T}$. The edges join two concepts if and only if they are mutually exclusive. This graph will be referred to as the exclusion graph. Now if A is a satisfiable Abox, then a concept can be represented in A only if its neighbors on the exclusion graph are not. In other terms, the boolean vector $\mathbb{1}_A$ is a "stable vector" of the graph ([9]). Determining maximal stable vectors (satisfiable Aboxes of maximal cardinality) was proved to be an NP-hard problem by Karp [13]. When all p_C's are equal, the probability distribution $\overline{\mathbb{P}}$ is the reversible measure of the so called Philosophers' Process (cf. [18]) derived from a classical resource sharing model of Dijkstra ([5]). It has been completely characterized in the case of ladder graphs in [8]. In the case where all p_C's are equal to $1/2$, the normalizing constant Z is proportional to the total number of stable sets of the exclusion graph. Recursive combinatorial methods allow to compute it for special classes of graphs (see [9]). Apart from these classes of graphs, very little can be said in general of the distribution $\overline{\mathbb{P}}$.

It is understood in our models that the number of concepts is relatively small compared to the number of objects. It is reasonable to expect that the cardinality of the $\mathcal{A}^{(o)}$'s is not too high, and that it is possible to explicitly compute the distribution $\overline{\mathbb{P}}$, at least numerically. Once this has been done, then due to the independence of components, the probability of most events concerning concept facts can be explicitly computed. As an example, fix a concept C and consider the event "$C(o) \in A$". Let π_C denote its probability under $\mathbb{P}^{(o)}$. Consider now in a random Abox with distribution $\mathbb{P}$, the total number of facts of type $C(o)$, for all $o \in \mathcal{O}$. That number is a random variable. Under the role-free assumption, it is easy to derive from proposition 4.2 that its distribution is binomial, with parameters n and π_C. If n is large, it can be approximated by a Gaussian distribution with mean $n\pi_C$, and variance $n\pi_C(1-\pi_C)$.

Let us now turn to role components. Under the role-free hypothesis, the situation is quite simple. The set $\mathcal{A}_R$ contains all possible 2^{n^2} Aboxes formed with role facts involving R only. It will be interesting to identify such an Abox to the edge set of a directed graph (or *digraph*) with vertex set $\mathcal{O}$. If $A \in \mathcal{A}_R$, and $o, o' \in \mathcal{O}$, define an edge pointing from o to o' if and only if $R(o, o') \in A$. Thus, to each element of $\mathcal{A}_R$ can be associated in a one-to-one manner a directed graph with vertex set $\mathcal{O}$. (Notice that in our case, contrarily to other models, loops are not excluded). Through that correspondence, the distribution $\mathbb{P}_R$ can be read as that of a random digraph, with independent edges and edge probability p_R. Since the founding paper by Erdös and Rényi [6], random undirected graphs have been extensively studied and all sorts of asymptotic results are now available (see Bollobás [1] or Spencer [17] as general references). Random digraphs, in spite of their numerous applications, have aroused lesser interest. Part of the reason is the belief, supported by some general results, that most asymptotic properties known for random graphs can be adapted to random digraphs. One of the most famous examples is the emergence of the so called giant component around $p_R = 1/n$ (see [12] for graphs and [14, 15] for digraphs).

Random graphs and digraphs are well known examples of the so called "$0-1$ laws", according to which most propositions on a large random system with a high degree of independence are either true or false with probability close to 1. At first, properties were restricted to first order logics. The celebrated Glebskii-Fagin theorem states that in a random (undirected) graph with fixed edge probability, any property of first order logics has a probability tending either to 1 or to 0 as n tends to infinity. That theorem was later extended by Shelah and Spencer to random graphs with edge probability $p = kn^{-\alpha}$, if α is any irrational (positive) number.

We shall rather focus here on *monotone* properties for which some recent results by Friedgut and Kalai [10] and Bourgain [3] equally apply to random graphs and digraphs, as well as other problems such as the phase transition in the k-satisfiability of random clauses. Let N be an integer and consider the set $E = \{0,1\}^N$. A subset F of E is said to be monotone if:

$$\eta \in F \Longrightarrow \zeta \in F\ , \quad \forall \zeta \geq \eta\ ,$$

where the ordering is taken componentwise. A property on E is monotone if the set F of those elements of E for which it is true is itself monotone. Translated in terms of Aboxes, a property is monotone if, being true for A, it remains true for any other Abox containing A. Under the open world hypothesis, non-satisfiability is a monotone property.

Another requirement imposed to our properties is *symmetry* with respect to permutations of objects. If σ is such a permutation, and A is an Abox, we define the Abox $\sigma.A$ by:

$$C(o) \in A \Longleftrightarrow C(\sigma(o)) \in \sigma.A$$

and

$$R(o,o') \in A \Longleftrightarrow R(\sigma(o),\sigma(o')) \in \sigma.A\ .$$

A set of Aboxes is said to be *symmetric* if it contains all $\sigma.A$'s as long as it contains A. A property is symmetric if the set of Aboxes for which it is true is symmetric.

Although the recent results of Bourgain [3] seem to indicate that symmetry could be superfluous in some cases, we believe that it is natural enough to impose it for properties of Aboxes. Notice that we do not impose symmetry through permutations of concepts or roles. Here are a few examples of monotone symmetric properties.

1. P_1: All objects are related by R to at least k other objects.
2. P_2: For any two objects o and o', there exists a path of relations of length k linking o to o'.
3. P_3: There exists k objets all related to each other by role R and for which concepts C_1 and C_2 are true.

The zero-one law that will be stated below concerns mainly the case where the parameters p_R and p_C are constant. However in that case random Aboxes contain mostly role facts, and are very unrealistic. It would be of greater interest to let the parameters depend on n. This will be the object of future work. Theorem 4.1 is essentially an extension of Theorem 1.1 of Friedgut and Kalai [10].

Theorem 4.1 *Assume that $\mathcal{A}$ is role free. Let P be any symmetric monotone proposition. For $\epsilon > 0$, denote by V_ϵ the subset of those elements p_C, p_R of $[0,1]^{c+r}$ such that, under the CPD $\mathbb{P}$, the probability that a random Abox satisfies P is between ϵ and $1-\epsilon$. There exists a constant γ, depending only on the Tbox, such that the diameter of the set V_ϵ is lesser than $\gamma \log(1/2\epsilon)/\log n$.*

Since the diameter of the set V_ϵ tends to 0 as n tends to infinity, it turns out that for most values of the parameters, the probability that a random Abox satisfies P will be either close to 0 (lesser than ϵ), or close to 1 (larger than $1-\epsilon$).

5 Random Aboxes with large object sets

The CPD of definition 3.1 depends on parameters associated to the elements of the Tbox, namely the values of the probabilities p_C and p_R. In the Tbox $\mathcal{T}$, some concepts may be defined by cardinality or type restrictions. The cardinality restrictions do not depend on the number of objects. In this section, we shall study the asymptotic properties of a random Abox under distribution $I\!P$, as the number n of objects tends to infinity. Recall from Definition 3.1 that $I\!P$ is the conditional measure on $\mathcal{A}$ of a product of Bernoulli distributions $\widetilde{I\!P}$. For the measure $\widetilde{I\!P}$, the average number of concept facts is $n\sum p_C$, whereas the average number of role facts is $n^2\sum p_R$. This is an indication that role facts should be predominant in a random Abox. The predominance of role facts has dramatic consequences on the structure of random Aboxes. Some concepts, in particular those defined by cardinality restrictions $(\leq kR)$, disappear whereas the others tend to be independent of roles. Actually, the distribution $I\!P$ is close to another CDP, on a particular role-free set of Aboxes. The precise approximation statement is given in Theorem 5.1, which is the main result of this section. Before that, we need to define an intermediary role-free set of Aboxes, between $\mathcal{A}$ and $\widetilde{\mathcal{A}}$ (the notations are those of section 3).

Definition 5.1 *Let $\mathcal{A}$ be a set of Aboxes, $\mathcal{A}^{(o)}$ and $\mathcal{A}_R$ its images through the coordinates of the decomposition operator of Definition 4.2. We call* role-free completion *of $\mathcal{A}$ the image through inverse decomposition of the Cartesian product:*

$$\mathcal{A}' = \times_{o\in\mathcal{O}}\mathcal{A}^{(o)} \ \times_{R\in\mathcal{T}}\mathcal{A}_R \ .$$

The idea of role-free completion is to separate exclusion rules between concepts, from constraints between concepts and roles. Denote by $\mathcal{A}_1$ the subset mapped by decomposition onto the Cartesian product $\times_o\mathcal{A}^{(o)}$: it could have been defined as the set of those elements of $\mathcal{A}$ that contain no role fact. Exclusion rules between concepts reduce $\mathcal{A}_1$ to a strict subset of $\{0,1\}^{cn}$. On the other hand, no restriction appears between roles in the $\mathcal{A}_R$'s: each of them has all possible 2^{n^2} elements.

We first need to identify those concepts that tend to disappear.

Definition 5.2 *A concept C is called* evanescent *if the probability that a random Abox under distribution $I\!P$ contains at least one concept fact $C(o)$ tends to* 0 *as n tends to infinity.*

The following result describes evanescent concepts.

Proposition 5.1

- *Any cardinality-restriction concept $\leq kR$ is evanescent.*
- *Any type-restriction concept $\forall RC$ is evanescent.*
- *If C is evanescent, then $\exists RC$ is evanescent.*
- *If C_1 and C_2 are evanescent, then $C_1 \sqcup C_2$ is evanescent.*
- *If C_1 or C_2 are evanescent, then $C_1 \sqcap C_2$ is evanescent.*

Evanescent concepts tend to disappear from random Aboxes, whereas non evanescent ones tend to be independent of roles. More precisely, the distribution $\mathbb{P}$ becomes close to another CPD $\mathbb{P}'$ on the role free completion of $\mathcal{A}$, in which evanescent concepts are absent.

Theorem 5.1 *Let $\mathcal{T}$ be a Tbox, and p_C, p_R be the parameters of a CPD $\mathbb{P}$ on $\mathcal{A}$. Define new parameters p'_C by:*

$$p'_C = \begin{cases} 0 & \text{if } C \text{ is evanescent,} \\ p_C & \text{else.} \end{cases}$$

Denote by $\mathbb{P}'$ the CPD on the role-free completion $\mathcal{A}'$ of $\mathcal{A}$, with parameters p'_C and p_R. Then as n tends to infinity, the distribution $\mathbb{P}$ tends to $\mathbb{P}'$ in the sense that:

$$\lim_{n\to\infty} \sup_{\mathcal{B}\in\mathcal{A}} |\mathbb{P}[\mathcal{B}] - \mathbb{P}'[\mathcal{B}]| = 0\ .$$

6 Conclusion

A natural probability distribution on sets of Aboxes associated to a given Tbox has been defined. Its parameters are fixed probabilities attached to the elements of the Tbox. As the number of objects tends to infinity, some concepts tend to disappear from random Aboxes, the others become independent of roles. The result is quasi-deterministic, in the sense that any monotone symmetric assertion on the random Abox is either true or false with a probability tending to one.

Under the fixed parameters hypothesis studied here, random Aboxes contain mostly role facts and they are not very realistic from the point of view of applications. A more interesting situation could be obtained by letting the role probabilities p_R tend to 0 as the inverse of the number of objects. This will be the theme of future researches.

References

[1] B. Bollobás. *Random Graphs.* Academic Press, London, 1985.

[2] B. Bollobás and A. Thomason Threshold functions. *Combinatorica 7, 35–38*, 1986.

[3] J. Bourgain. On sharp thresholds of monotone properties. *To appear*, 1999.

[4] G. Chartrand and L. Lesniak. *Graphs and Digraphs. Second ed.* Wadsworth & Brooks/Cole, Mathematics Series, 1986.

[5] E.W. Dijkstra. The structure of the multiprogramming system. *Comm. ACM 11 341–346*, 1968.

[6] P. Erdös and A. Rényi. On the evolution of random graphs. *Mat. Kutató Int. Közl., 5 17–61*, 1960.

[7] F. Forbes, O. François, and B. Ycart. Stochastic comparison for resource sharing models. *Markov Proc. Rel. Fields 2(4) 581–605*, 1996.

[8] F. Forbes and B. Ycart. The Philosophers' process on ladder graphs. *Comm. Stat. Stoch. Models 12(4) 559–583*, 1996.

[9] F. Forbes and B. Ycart. Counting stable sets on Cartesian products of graphs. *Discrete Mathematics*, 186 105–116, 1998.

[10] E. Friedgut and G. Kalai. Every monotone graph property has a sharp threshold. *Proc. Amer. Math. Soc. 124 2993–3002*, 1996.

[11] M. Hofri. *Probabilistic analysis of algorithms.* Springer-Verlag, New York, 1987.

[12] S. Janson, D.E. Knuth, T. Łuczak, and B. Pittel. The birth of the giant component. *Random Structures and Algorithms, 4(3) 233–258*, 1993.

[13] R.M. Karp. Reducibility among combinatorial problems. In *Complexity of computer computations, R.E. Miller and J.W. Thatcher ed., 85–103,* Plenum Press, New-York, 1972.

[14] R.M. Karp. The transitive closure of a random digraph. *Rand. Structures and Algo. 1(1) 73–94*, 1990.

[15] T. Łuczak. The phase transition in the evolution of random digraphs. *J. Graph Theory 14(2) 217–223*, 1990.

[16] D. Nardi, F.M. Donini, M. Lenzerini, and A. Schaerf. Reasoning in description logics. In *Principles of Artificial Intelligence.* G. Brewska (ed.), Springer-Verlag, New York, 1995.

[17] J.H. Spencer. Nine lectures on random graphs. In P.L. Hennequin, editor, *Ecole d'été de probabilités de Saint-Flour XXI-1991*, L.N. in Math. 1541, 293–347. Springer-Verlag, New York, 1993.

[18] B. Ycart. The Philosophers' Process: An ergodic reversible nearest particle system. *Ann. Appl. Probab. 3(2), 356–363*, 1993.