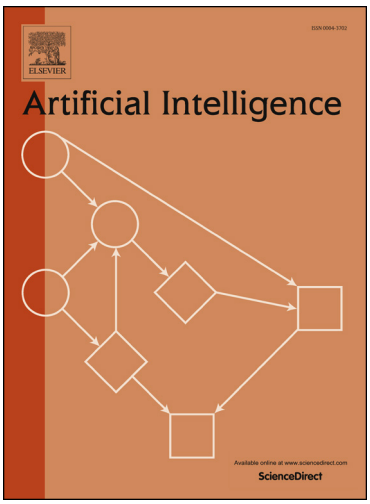
Accepted Manuscript



The Complexity of Bayesian Networks Specified by Propositional and Relational Languages

Fabio G. Cozman, Denis D. Mauá

PII: S0004-3702(18)30304-7
DOI: https://doi.org/10.1016/j.artint.2018.06.001
Reference: ARTINT 3076

To appear in: *Artificial Intelligence*

Received date: 13 June 2017
Revised date: 16 May 2018
Accepted date: 5 June 2018

Please cite this article in press as: F.G. Cozman, D.D. Mauá, The Complexity of Bayesian Networks Specified by Propositional and Relational Languages, *Artif. Intell.* (2018), https://doi.org/10.1016/j.artint.2018.06.001

The Complexity of Bayesian Networks Specified by Propositional and Relational Languages

Fabio G. Cozman[a], Denis D. Mauá[b]

[a] *Escola Politécnica, Universidade de São Paulo, Brazil*
[b] *Instituto de Matemática e Estatística, Universidade de São Paulo, Brazil*

Abstract

We examine the complexity of inference in Bayesian networks specified by logical languages. We consider representations that range from fragments of propositional logic to function-free first-order logic with equality; in doing so we cover a variety of plate models and of probabilistic relational models. We study the complexity of inferences when network, query and domain are the input (the *inferential* and the *combined* complexity), when the network is fixed and query and domain are the input (the *query/data* complexity), and when the network and query are fixed and the domain is the input (the *domain* complexity). We draw connections with probabilistic databases and liftability results, and obtain complexity classes that range from polynomial to exponential levels; we identify new languages with tractable inference, and we relate our results to languages based on plates and probabilistic relational models.

Keywords: Bayesian networks, Complexity theory, Relational logic, Plate models, Probabilistic relational models

1. Introduction

A Bayesian network can represent any distribution over a given set of random variables [33, 69], and this flexibility has been used to great effect in a variety of applications [107]. Many of these applications contain repetitive patterns of entities and relationships. Thus it is not surprising that practical concerns have led to modeling languages where Bayesian networks are specified using relations, logical variables, and quantifiers [46, 109]. Some of these languages enlarge Bayesian networks with plates [47, 83], while others resort to elements of database schema [44, 58]; some others mix probabilities with logic programming [104, 116] and even with functional programming [85, 89, 100]. The spectrum of tools that specify Bayesian networks by moving beyond propositional sentences is vast, and their applications are remarkable.

Yet most of the existing analysis on the complexity of inference with Bayesian networks focuses on a simplified setting where nodes of a network are associated with categorial variables and distributions are specified by flat tables containing