Inference Algorithms for Systems of Medical Diagnosis Aid based on Bayesian Networks

RUDINI MENEZES SAMPAIO[1]
FELIPE LEAL VALENTIM[2]
LEANDRO ALVES DE SOUZA[3]
RICARDO MARTINS DE ABREU SILVA[4]

DCC - Dept. Computer Science
UFLA - Federal University of Lavras
Lavras, MG, Brazil, CEP 372000-000
[1]rudini@dcc.ufla.br
[2]felipe.flv@gmail.com
[3]lsouza@comp.ufla.br
[4]rmas@dcc.ufla.br

Abstract. Several scientific works have modeled medical problems with assistance of Bayesian networks, assisting doctors in the task of diagnosing a disease given the observed symptoms and evaluated exams. This paper aims to present the execution time and convergence analyses for exact and approximate algorithms for probabilistic inference, which allow to apply the Bayesian reasoning in the support to the medical diagnosis. The results of the analyses supply a criterion for the choice of the algorithm to be implemented depending on the resources that are wished to optimize.

Keywords: Bayesian networks, inference algorithms, systems for medical diagnosis aid

(Received August 20, 2007 / Accepted January 04, 2008)

1 Introduction

The evolution of reasoning in the interpretation of natural phenomena across time has brought as a consequence the mathematic bases of scientific thought. In Medicine, it was not different: the observation of biological phenomena, the search for solutions to decrease the impact and the need to prove scientifically the effectiveness of protherapeutic method and of therapeutic procedures opened the doors to what is named evidence-based medicine. Thomas Bayes, an seventeenth century English mathematician left us his theorem which established that the post-test probability of a disease was a function of sensitivity and specificity of the examination and prevalence of the disease in the population (pre-test probability). The physicians in formulating their diagnostic hypotheses, in interpreting laboratorial exams and in prescribing a treatment, intuitively utilize Bayes' Theorem. Today, we live the high technology age in which persons often tend to interpret the positiveness of a sophisticated and expensive exam as a synonymous of disease. We should not forget that all the examinations, without any exception, since the current clinical examination to a computerized tomography, are limited by the sensitivity, specificity and pre-test predictive value. Bayesian networks are oriented acyclic graphs which stand for dependences among variables in a probabilistic model. This approach represents a good strategy to deal with problems which are concerned with uncertainties. As a Bayesian model is constructed in a causal manner (direction of the arcos point from the causes to the effect), the graphic representation of a diagnostic problem is relatively simple, using the knowledge of the specialist of the domain. The fundamental task of a system of support to the medical diagnostic, with a Bayesian structure, is to compute the distribution of a posteriori probability of a set of consultation variables (Variables_Consultation) given the values of another set variables (Variables_Evidence). Nevertheless, in a Bayesian network, any variable can be regarded as Varables_Consultation or Variables_Evidence, allowing four distinct types of consultation:

- Diagnostic (from effects to the cause): for example, given a S1 symptom S1 infer the probability of disease D1, P(D1|S1);
- Causal (from the causes to the effect): for example, given disease D2 find the most likely symptoms, P(S2| D2);
- Inter-causal (between causes of a common effect): for example, given symptom S2 to infer P(D1|S2), but adding the evidence that D2 is present provokes the slope of the probability of D1; in spite of D1 and D2 being independent, the presence of one makes the other less probable;
- Mixed (matches two or more of the previous ones).

In the literature, several proposals for the solving of this problem of great practical importance are found, but, one should be able to choose the best algorithm to optimize the resources and widen the applications of Bayesian networks. We could, for example, choose an algorithm which can be implemented in mobile systems, which possess limited memory and processing resources, for them to be able to be used making a pre-consultation on patients in rows of overworked hospitals. This work presents the analyses of algorithms for probabilistic inference in Bayesian networks, supplying a criterion for choice of the algorithm to be implemented depending on the resources which are wished to optimize.

2 Bayesian networks

Bayesian networks offer an approach to probabilistic reasoning which encompasses graphs theory, for the establishment of the relationships among sentences and still, probability theory, for attribution of trustability levels, contemplating the needs to deal with uncertainty. Bayes' theorem, contemplated in the bases of probability theory, supplies mechanisms to manipulate the probabilities in a Bayesian model.

2.1 Bayes' theorem

To reach Bayes' theorem, one starts form basic principles. So the probability fro one to observe simultaneously an x event and a Y event is defined by the conditional probability. That formulation is presented in several manners which if combined; equation 1 is obtained:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \quad (1)$$

Equation (1) is known as Bayes' theorem (or still Bayes' rule or Bayes' Law). In this equation, P(X|Y) is the probability of evidence X being observed, given that Y is true. Namely, P(X|Y) is the possibility of the occurrence of Y causing evidence Y [7].

In some situations, it is necessary to utilize a more general version of Bayes' theorem, conditioned upon some evidence E known. This more general version is represented by 2:

$$P(Y|X,E) = \frac{P(X|Y,E)P(Y|E)}{P(X|E)} \quad (2)$$

Bayes' theorem is an equation of great importance, for it permits the computation or the utilization of conditional probabilities as a an priori probability function and possibilities.

2.2 Bayesian Reasoning

The Bayesian reasoning can be explained with an medical example taken from [8]. Consider the problem:

1% of the women over 40 years old who took part in routine examinations are carriers of breast cancer. 80% of the women with cancer will have positive results of mammography. 9.6% of the women without the disease will also have positive results at the mammography. A woman at his age comes across with a positive mammography result, which is the probability for her carrying breast cancer?

Second [8], most of the physicians would estimate that the probability for the woman in issue, having breast cancer would be between 70% and 80%. One can settle the problem in a Bayesian manner to estimate the correct probability.

In the first place, in a woman over 40 years old, breast cancer (Cancer) can be either present or absent. Those alternatives, mutually excluding, can be put into a table. One can start the reasoning by the probability of each alternative before we do any test. It is so called a priori probability. Cancer=present or Cancer=absent. Since 1% of the women over 40 years have breast cancer, the a priori probability of Cancer being present is of 0.01 and being absent is of 0.99.

Now, let's incorporate the mammography result. Cancer is present, the conditional probability of Mammography being positive is of 0.80 (80%), and if Cancer is absent, this probability is of 0.096 (9.6%). One can put together that information in a conditional Cancer probability table, as in Table 1.

Table 1: Table of conditional probabilities

Mammography	Cancer	
	present	absent
positive	0,8	0,096
negative	0,2	0,904

Multiplying the a priori probability by the conditional one, we obtain the joint probability of Cancer e Mammography, shown in Table 2:

Table 2: Joint probabilities of Cancer and Mammography

Mammography	Cancer	
	present	absent
positive	0,01 x 0,8 = 0,008	0,99 x 0,096 = 0,09504
negative	0,01 x 0,2 = 0,002	0,99 x 0,904 = 0,89496

To cause the soma of each line of joint probability becomes 1, it is needed to use a normalization: by multiplying each probability by the normalization constant, which is given by 1 divided by the summation of each line of the joint probability table. Obtaining thus the so called the e a posteriori probability shown in Table 3.

Therefore, the Bayesian reasoning led us to conclude that the a posteriori, probability, namely, after the tests, of a women over 40 years of posse de a mammography examination the result of which is positive, that is, a P (Cancer=present | Mammography=positive) is of only 0.07764 (7.764%).

Table 3: A posteriori probabilities

Mammography	Cancer	
	present	absent
positive	0,07764	0,92236
negative	0,00223	0,99777

In [8] it is reported that when that problem was presented to several physicians and medical students, a trend to overestimate the a posteriori probability of a disease and only 46% of the interviewees estimated a probability consistent with the correct answer. That fact points out that the Bayesian reasoning is not intuitive. There seems to be a general trend to ignore the fact that the a priori probability of disease is small.

In the example above, the Bayesian reasoning allowed to quantify the degree in which the positive mammography result fitted an early estimate of chance for a woman to have breast cancer. Under this standpoint, medical test (or evidence) works as a modifier of opinion, updating an early hypothesis (a prioi probability) to generate another (a posteriori probability. This latter encompasses both the previous belief (a priori probability) and the result of the test. The a posteriori probability becomes automatically the a priori probability for subsequent tests [8]. In another example, we can observe a simple domain modeled from a Bayesian network. It is concerned with a hypothetical example, which deals with the diagnostic of hepatitis. It possesses three variables: Fever, Jaundice and Hepatitis (Figure 2.2). It is found that the variables Fever and Jaundice are associated with status table which indicate their a priori probabilities. The a priori probabilities are the element responsible for the indetermination treatment. **When either it cannot be said or not know to say the value of an evidence, the Bayesian network will use in the inference, that a priori probability, which was defined** by the specialist as a species of "standard value".

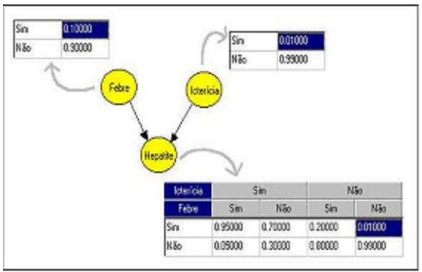

Figure 1: Example of support of diagnostic with Bayesian networks.

2.3 Definition of Bayesian network

A Bayesian network is acyclic directed graph where the knots stand for random variables and the directed arcs stand for direct causal relationships between the connecting knots [7]. It was defined, in Bayesian networks, if there is a arrow of the knot X as far as in the Y, it is said that X is father to Y.

The direction of arcos, in general, stands for the cause-effect relationships among the domain variables. For example, if there is a arc going from a knot A to a knot B, it is denoted that A stands for a cause of B, and it is supposed that A is one of the parents of B, or in the medical case, it could mean that X is a symptom and Y is a disease.

In a Bayesian network, each knot is conditionally independent of any subset of knots which are not their descendents, known the parent knots of Xi (represented by pa(Xi)).

3 Bayesian inference

The basic task which one wishes to perform in a Bayesian network is to compute the distribution of the conditional probability to a set of consultation variables, given the values of a set of evidence variables, that is, to compute the P(variable_consultation|variables_evidence).

That task is called the Bayesian inference and enables to answer to a series of "consultations" on a domain of data. For example, in the medical area, the main task consists in obtaining a diagnostic for a given patient presenting certain symptoms (evidences). This task consists in updating the probabilities of the variables as related to evidence. In the case of medical diagnostic, the probabilities of each of the possible diseases is tried to know, given the symptoms observed in the patient. Those are a posteriori probabilities.

3.1 Exact algorithms of Bayesian inference

One inference algorithm is called exact if it performs the calculation of a posteriori probabilities from the principles of Bayes' theorem, by means of both sum and combinations of values, without any other error except the one of rounding out in the calculation [4].

The general idea of the exact inference methods is evaluating equation 3, derived from Bayes' theorem and of the basic axioms of the probability theory, finding the a posteriori probability of the consultation variable.

$$P(X|e) = \alpha P(X,e) = \alpha \sum_{y} P(X,e,Y) \qquad (3)$$

Where it is denoted that X is a consultation variable, E the set of evidence variables, e the set of values found for E, and Y the unobserved remaining variables (named hidden variables) and α is the normalization constant, which warrants that the sum of the resulting distribution will be equal to 1. The exact inference algorithms implemented were: Enumeration Algorithm [7] and Elimination of Variables [2] [3].

3.2 Approximate Algorithms of Bayesian inference

The algorithms considered within the group of approximate methods utilize distinct simulation techniques to obtain approximate values of probabilities [4].

The most utilized approximate algorithms belong to the group of stochastic simulation algorithms. The main idea of this method is using the model of Bayesian network to simulate the flow of impact or influence of evidence on the other variables [6]. In this type of algorithm, according to the conditional probability tables of the network, a set of randomly selected samples, then, inference is performed, this is, the probabilities of the "consultation" variables are approached by the frequency of their appearances in the sample. The accuracy of the results is to depend on the size of the samples (on the number of simulations which generate the samples) and, differently from the exact methods, the net structure is not relevant in the calculation of inference, that being one of its main advantages.

The approximate inference algorithms implemented were: Forward Sampling [7], Likelihood Weighting [7] [5] and Gibbs Sampling [7] [4] [1].

4 Results

To investigate the implemented algorithms, a computer with a Pentium III 1.2GHz processor with 512MB RAM was utilized. All the algorithms were implemented using the programming language JAVA (J2SE).

4.1 Exact algorithms

To compare the efficiency of exact algorithms, random networks with different numbers of variables (always booleans) were generated and a consultation in the network was performed, always with only a evidence selected, and the execution time of the algorithm was measured. Table 4 presents the measured times.

Table 4: Execution time of the exact algorithms

Execution time (milliseconds)		Number of variables
Enumeration Algorithm	Algorithm of variable elimination	
15	15	3
21	20	24
38	32	5
47	45	7
67	59	8
124365	40215	20

The increase of the size of the net implies into an increase of the execution time in a non-linear form. It is known that the complexity of time of both the implemented exact algorithms is exponential relative to the net size [7]. However, the algorithm of elimination of variables, as foreseen, eliminates repeated calculations supplying again the efficiency of that algorithm in relation to the enumeration algorithm.

One should stress that the ordering of the variables influences the execution time of the Elimination of Variables algorithm [4] [7], and that the implementation of that work always uses the ordering of the leaves to the roots.

4.2 Approximate algorithms

The implemented algorithms were tested in different topology networks, however, Asia, DogProblem and CarDiagnostic networks were chosen to present the results (Figures 4.2, 4.2 and 4.2 respectively).

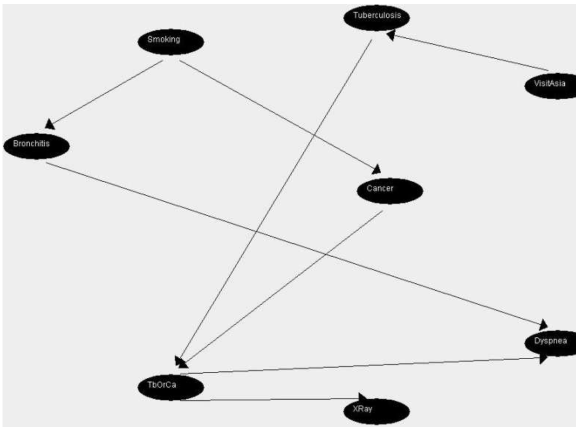

Figure 2: Structure of Asia network

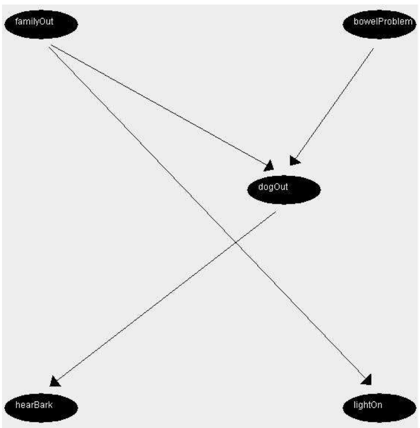

Figure 3: Structure of DogProblem network

Asia network is a multi-connected network associated to the eight booleans variables; DogProblem problem, is a classic example found in the literature, simpler, with five booleans variables; CarDiagnostic network is a more complex network with 20 discrete variables with several layers, which can be applied to the diagnostic of bad functioning of self-propelling vehicles.

The Gibbs Sampling inference algorithm presents some problems inherent to its characteristics; and some suggestions of solutions were implemented. For the problem of not having an optimum initial configuration, the method proposed by [6] was implemented, which is discard from 5 to 10% of the initial configurations. The algorithm with that method will be identified by Gibbs

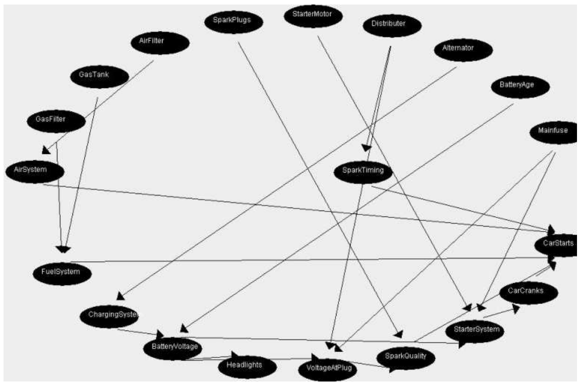

Figure 4: Structure of CarDiagnostic network.

Sampling Burn-in in the figures which present the results.

In coming into contact with the approximate inference techniques, the idea of mixing characteristics of two approximate algorithms which presented more interesting solutions, and e on pretence of validation, the results of that algorithm, baptized Gibbs Weighting, will be included in the comparisons. Gibbs Weighting generates the sample of the events exactly as Gibbs Sampling, but it does the updating of posteriori probability and according to the weighing function of Likelihood Weighting algorithm.

Therefore, the inference algorithms evaluated were: Forward Sampling, Likelihood Weighting, Gibbs Sampling, Gibbs Weighting Sampling, Gibbs Sampling Burn-in, Enumeration and Elimination of Variables.

To compare the implemented inference algorithms convergence analyses were done, depending on the number of interactions and time to evaluate the efficiency of each of them.

The exact results, which served as a comparison base, were obtained from the Enumeration algorithms and Variable Elimination, which showed identical results (ignoring the rounding out of errors).

Figures 4.2, 4.2 and 4.2, showed the convergence curves of the algorithms for the proposed test networks, considering the number o iterations.

It can be observed that all the algorithms converge to the exact result, increasing the number of iterations.

Gibbs Sampling algorithms and their variation with the implemented Burn-in method, proved the most efficient, needing of lower number of iterations to converge to the exact result. One can notice that, in small networks, which need of few iterations to warrant a precise result, the discard of the initial samples influence negatively the result, but, to a greater number of iterations, Burn-in supplies a significant help to the precision of the result.

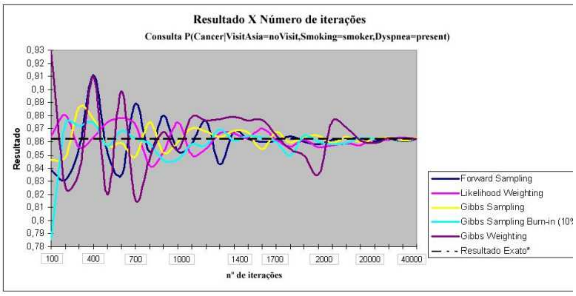

Figure 5: Graphic. Result X Num. of iterations for a consultation in Asia network.

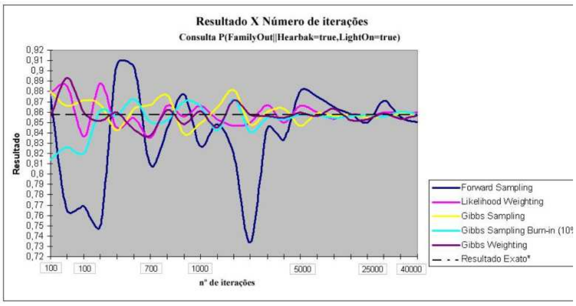

Figure 6: Graphic .Result X Num. of iterations for a consultation in the DogProblem network.

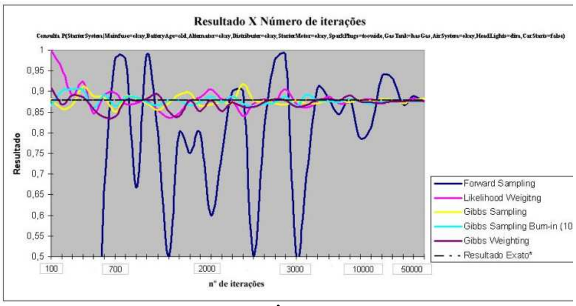

Figure 7: Graphic .Result X NÂº of iterations for a consultation in CarDiagnostic network.

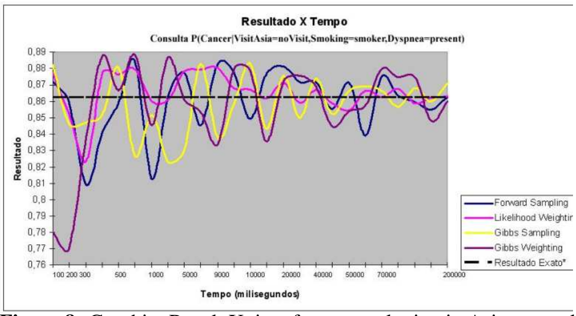

Figure 8: Graphic. Result X time for a consultation in Asia network.

The Likelihood Weighting algorithm proves a powerful method taking into consideration the number of simulations. One can notice that for some cases of the tested networks, the behavior of that algorithm proved similar to that of Gibbs Sampling. One should stress the ease of implementing that algorithm as compared with the most efficient methods.

Forward Sampling algorithm presented a good behavior when a consultation was performed in the Asia network, but for consultations in DogProblem and CarDiagnostic networks, that algorithm seems unviable. This fact can be accounted for by the great number of rejected samples. According to Segundo [7], the portion of rejected samples grows exponentially according to the number of evidence variables does, and so the algorithm is simply useless for complex problems. That fact can be corroborated experimentally, observing the behavior of the algorithm of figure 4.2.

The number of iterations necessary to a convergence for the exact result is a good indicator of efficiency of algorithms. But, each has an average rate of execution per iteration. To compare which algorithm converges faster for the exact result, we stipulated time for the execution instead the number of simulations. Figure 4.2, 4.2 and 4.2 show the behavior for that analysis.

One can observe that Gibbs Sampling algorithm converges faster for the exact result, followed by Likelihood Weighting. That behavior is only stood out in larger networks.

The good amount of random number generated constitutes the secret for the efficiency and good convergence of the approximate inference algorithms. In the implemented algorithms, an efficient generator of random numbers was employed, MersenneTwister [7], which is available on the Internet.

The choice of the ideal inference algorithm depends upon the size of Bayesian network and on the resources one wishes to optimize. The exact inference consumes processing exaggeratedly in small networks and consumes too much time in very large networks, becoming unviable to certain real problems, mainly if we take into account that the algorithms could be intended to solve problems in real time or to be implemented in mobile systems and/or embarked with the restriction of processing and memory.

Out of the approximate algorithms, Gibbs Sampling proved, in general, the most powerful method to accomplish inference, deserving special attention fro possible improvements and adaptations which make inference still more efficient. Likelihood Weighting algorithm also proved a powerful method (having performance similar in several cases to the one of Gibbs Sampling), mainly if we take into account its easy implementation.

Gibbs Weighting algorithm proposed by the author, demonstrated behavior similar in relation to their precursors (Gibbs Sampling and Likelihood Weighting),

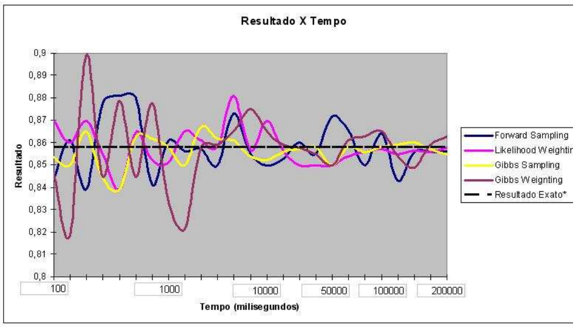

Figure 9: Graphic. Result X time for a consultation in DogProblem network.

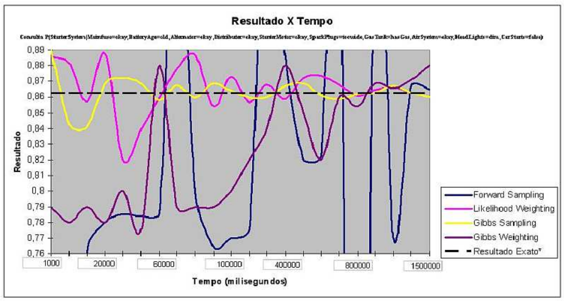

Figure 10: Graphic. Result X time for a consultation in DogProblem network.

presenting a good convergence in relation to the number of iterations, by demanding a lot of time for convergence.

5 Conclusions

From the use of Bayesian networks and the probabilistic reasoning, it is possible to construct systems which act as specialists in the support to the medical diagnostic. The present work demonstrates briefly that Bayesian inference algorithms found in the literature can be easily implemented and adapted to meet systems with restrictions of processing and memory, as embarked systems or web systems, by means of a previous study of the complexity of the problem which will be modeled.

References

[1] Casella, G. and George, E. I. Explaining the gibbs sampler. *The American Statistician*, 46(3):167–174, August 1992.

[2] Cozman, F. Generalizing variable elimination in bayesian networks. *Technical Report of the Escola Politecnica, University of Sao Paulo*, 2000.

[3] Dechter, R. Bucket elimination: a unifying framework for probabilistic inference. pages 75–104, 1999.

[4] Enrique Castillo, J. M. G. and Hadi, A. S. Expert systems and probabilistic network models. *The Knowledge Engineering Review*, 15(3):303–304, September 2000.

[5] Fung, R. M. and Chang, K.-C. Weighing and integrating evidence for stochastic simulation in bayesian networks. pages 209–220, 1990.

[6] Jensen, F. Bayesian graphical models. *Ed. Springer*, 2000.

[7] Wagner Teixeira da Silva, M. L. Mineração de dados em redes bayesianas. In *CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO*.

[8] Yudkowsky, E. An intuitive explanation of bayesian reasoning. 2003.