

CENTRO UNIVERSITÁRIO FEEVALE

FELIPE SCUCIATTO DOS SANTOS

DESENVOLVIMENTO DE MÓDULO DE REDES BAYESIANAS PARA  
O AMPLIA

Novo Hamburgo, julho de 2009

FELIPE SCUCIATTO DOS SANTOS  
scuciatto@feevale.br

DESENVOLVIMENTO DE MÓDULO DE REDES BAYESIANAS PARA  
O AMPLIA

Centro Universitário Feevale  
Instituto de Ciências Exatas e Tecnológicas  
Curso de Ciência da Computação  
Trabalho de Conclusão de Curso

Professor Orientador: Marta Rosecler Bez  
Professor Co-orientador: Cecília Dias Flores

Novo Hamburgo, julho de 2009

## RESUMO

O software AMPLIA foi desenvolvido pelo Grupo de Inteligência Artificial do Instituto de Informática da Universidade Federal do Rio Grande do Sul (II/UFRGS), em parceria com outras instituições. O AMPLIA é um Ambiente Multiagente Probabilístico Inteligente de Aprendizagem, que possibilita a aprendizagem colaborativa. Dentro do ambiente, o aluno demonstra sua hipótese médica através de um modelo gráfico, construindo uma Rede Bayesiana. Este modelo é baseado em casos clínicos reais e, posteriormente, comparado com modelos criados por especialistas. Visando sanar algumas limitações do projeto inicial, foi proposto um novo projeto, com objetivo de reescrever o AMPLIA, tendo agora a Internet como meio de colaboração, possibilitando assim a colaboração entre várias instituições. Com este trabalho, será desenvolvido o módulo para criação, inferência e avaliação das Redes Bayesianas, dentro do AMPLIA.

**Palavras-chave:** Inteligência Artificial, Redes Bayesianas, Ensino da Medicina.

## ABSTRACT

The AMPLIA software was developed by the Group of Artificial Intelligence of the Institute of Informatics from the Universidade Federal do Rio Grande do Sul (II / UFRGS), in partnership with other institutions. AMPLIA is a Multi-agent Probabilistic Intelligent Learning Environment, which provides collaborative learning. Within the environment, the student demonstrates his medical hypothesis through a graphical model, building a Bayesian Network. This model is based on actual cases and thereafter compared with models created by specialists. Aiming to address some limitations of the initial project, new project was proposed, aiming to rewrite the AMPLIA taking now Internet as a way of collaboration, tenabling the collaboration between several institutions. In this work, will be developed the module for creation, inference and evaluation of Bayesian networks within the AMPLIA software.

**Keywords:** Artificial Intelligence, Bayesian Networks, Medicine Teaching.

## Lista de Figuras

Figura 1	Rede Bayesiana do problema do terremoto/ladrão. Esta rede além da topologia, demonstra a tabela de probabilidades condicionais (RUSSEL; NORVIG, 2004) . . . . .	14
Figura 2	Rede Bayesiana com tabela de probabilidades condicionais . . . . .	15
Figura 3	D-separação entre X e Y . . . . .	16
Figura 4	Rede Bayesiana com suas probabilidades a priori à esquerda (NEA-POLITAN, 2003) . . . . .	19
Figura 5	Grafo antes do processo de moralização . . . . .	23
Figura 6	Grafo após o processo de moralização . . . . .	23
Figura 7	Grafo após o processo de triangularização . . . . .	24
Figura 8	(RUSSEL; NORVIG, 2004) . . . . .	26

## Sumário

<b>INTRODUÇÃO</b> .....	<b>7</b>
<b>1 REDES BAYESIANAS</b> .....	<b>9</b>
1.1 Teoria da Probabilidade . . . . .	9
1.1.1 Cálculo de probabilidades: o axioma base, trabalhando com proba- bilidades incondicionais . . . . .	10
1.1.2 Probabilidade Condicional . . . . .	10
1.1.3 O Teorema de Bayes . . . . .	11
1.1.4 Inferência Bayesiana . . . . .	11
1.2 As Redes Bayesianas . . . . .	13
1.2.1 Formalização Matemática . . . . .	14
1.2.2 Independência condicional nas Redes Bayesianas . . . . .	15
1.2.3 Trabalhos Correlatos . . . . .	16
<b>2 REALIZANDO INFERÊNCIA NAS REDES BAYESIANAS</b> .....	<b>18</b>
2.1 A complexidade dos algoritmos de inferência . . . . .	18
2.2 Exemplificação da Inferência em uma Rede Bayesiana . . . . .	19
2.3 Algoritmos de Inferência Exata . . . . .	19
2.3.1 Inferência por enumeração . . . . .	20
2.3.2 Algoritmo de eliminação de variáveis . . . . .	21
2.3.3 Inferência por árvores de junção . . . . .	22
2.3.4 Transformação topológica e construção da árvore de junção . . . . .	22
2.3.5 Inferência na árvore de junção . . . . .	24

2.4	Inferência aproximada em Redes Bayesianas . . . . .	25
2.4.1	Inferência por simulação de cadeias de Markov . . . . .	25
<b>3</b>	<b>PROPOSTA DE TRABALHO . . . . .</b>	<b>28</b>
3.1	O software UnBBayes . . . . .	28
3.2	Características das Redes Bayesianas no projeto AMPLIA . . . . .	28
	<b>CONSIDERAÇÕES FINAIS . . . . .</b>	<b>29</b>
	<b>Referências . . . . .</b>	<b>30</b>

## INTRODUÇÃO

O AMPLIA<sup>1</sup> é um ambiente computacional multiagente, que objetiva apoiar o desenvolvimento do raciocínio diagnóstico e a modelagem de hipóteses, como um recurso adicional para a formação de novos médicos. No Amplia, o conhecimento é modelado através de Redes Bayesianas, representando assim, de forma gráfica e probabilística, as hipóteses levantadas pelo estudante.

As Redes Bayesianas são uma forma gráfica de representação das relações entre variáveis e suas probabilidades dentro de um escopo. Elas são representadas por grafos acíclicos, nos quais cada nó é uma variável aleatória (CHARNIAK, 1991). Cada variável deve possuir um conjunto limitado de valores (estados) e a cada nó raiz da rede, deve ser atribuída uma probabilidade. As probabilidades dos nós não raiz devem levar em conta as probabilidades dos pais.

Segundo Flores (2003), a abordagem de Redes Bayesianas foi escolhida para tratar o conhecimento incerto por sua rigorosa fundamentação em princípios matemáticos. A abordagem Bayesiana também permite lidar com incertezas, representada por probabilidades, isto é, a probabilidade de variáveis assumir valores específicos, dadas as evidências disponíveis. Ainda segundo Flores, outra razão importante para a escolha desta abordagem é que a mesma permite a modelagem qualitativa e quantitativa do domínio. O enfoque qualitativo é representado pelo conjunto de variáveis e seu relacionamento causal, enquanto o enfoque quantitativo expressa a intensidade desse relacionamento.

No AMPLIA, o aluno desenvolve uma rede bayesiana através do seu raciocínio lógico por meio de um caso clínico apresentado. A rede construída pelo aluno é então comparada com a rede construída pelo especialista, que deve identificar os prováveis conflitos. O resultado gerado é então utilizado, nas estratégias pedagógicas mais convenientes para auxílio ao diagnóstico correto (FLORES, 2005).

Observando as limitações do projeto original, um novo projeto foi iniciado pelo grupo de pesquisas da UFCSPA, com intuito de utilizar tecnologias e recursos de Internet dentro do AMPLIA, visando principalmente a troca de conhecimento entre várias instituições. Esta nova versão mantém todas as funcionalidades já implementadas na versão inicial,

---

<sup>1</sup>Ambiente Multiagente Probabilístico Inteligente de Aprendizagem

porém, o software será reescrito utilizando a linguagem Java, facilitando assim a comunicação através de Web e seu uso em diversas plataformas.

Este trabalho tem como principal objetivo implementar o módulo de desenvolvimento e avaliação de Redes Bayesianas dentro do AMPLIA. Para atingir tal objetivo, técnicas de criação, avaliação e inferência de Redes Bayesianas serão estudadas. Este trabalho utiliza como base o software UnBBayes<sup>2</sup>, utilizado para modelar e avaliar Redes Bayesianas.

Este trabalho está dividido da seguinte forma: no capítulo um, é apresentada a teoria sobre Redes Bayesianas, bem como as fundamentações matemáticas utilizadas pela técnica. No segundo capítulo, os principais algoritmos de inferência sobre redes são apresentados. O capítulo três apresenta a proposta de trabalho para este projeto.

---

<sup>2</sup>Disponível em: <http://unbbayes.sourceforge.net/>

## 1 REDES BAYESIANAS

As Redes Bayesianas são uma forma gráfica de representação das relações entre variáveis dentro de um escopo. É uma representação das probabilidades de um problema. As Redes Bayesianas são representadas por grafos acíclicos, onde cada nó é uma variável aleatória (CHARNIAK, 1991). Utilizando-se Redes Bayesianas, pode-se calcular a probabilidade condicional de cada variável em um problema, apenas observando valores de variáveis próximas.

### 1.1 Teoria da Probabilidade

A Teoria das Probabilidades é o estudo matemático das probabilidades. Por décadas, tem-se usado métodos probabilísticos como base de inferência em sistemas de Inteligência Artificial. Normalmente estes métodos são aplicados a problemas complexos, onde uma característica de determinada entidade tem influência direta sobre outra característica de outra entidade. Por exemplo, a presença ou ausência de uma determinada doença em um ser humano, tem influência direta em um exame que verifica a presença da mesma (NEAPOLITAN, 2003).

A principal vantagem de se trabalhar com métodos probabilísticos é poder lidar com incertezas, podendo assim tomar decisões mesmo com um número reduzido de evidências. Ao se trabalhar com dados incertos, precisa-se utilizar níveis de certeza, e não apenas valores absolutos. Por exemplo, a probabilidade de acertar um número ao se jogar um dado é de aproximadamente 0,2 (20%). Ou então, qual a probabilidade de um paciente fumante apresentar ou não câncer de pulmão?

Neste capítulo serão abordados os pilares da teoria da probabilidade, que servirão de base para assuntos subsequentes. Nos capítulos seguintes estes conceitos serão aplicados no desenvolvimento de um sistema inteligente.

### 1.1.1 Cálculo de probabilidades: o axioma base, trabalhando com probabilidades incondicionais

O axioma base da teoria de probabilidades demonstra o cálculo de probabilidades incondicionais: para qualquer evento  $a$  dado que  $0 \leq P(a) \leq 1$  e  $P(a) = 1$  se e apenas se  $(a)$  ocorrer com certeza.

Simplificando o axioma acima, a probabilidade de um evento qualquer é representada por um número dentro do intervalo  $[0,1]$ . A probabilidade é igual a 1 apenas se o grau de certeza for igual a 100%.

Extendendo o axioma anterior, passa-se a trabalhar com mais eventos: para quaisquer eventos  $a$  e  $b$ , mutuamente exclusivos, a probabilidade de  $a$  ou  $b$  ocorrer é:

$$P(a \vee b) = P(a) + P(b) \quad (1.1)$$

Simplificando, a probabilidade de um evento  $a$  ou  $b$  ocorrer é igual a soma de suas probabilidades.

Exemplificando, considere que a variável aleatória *câncer* denota a possibilidade de um paciente qualquer apresentar câncer. Com isso ao dizer que  $P(\text{câncer}) = 0,01$ , informa-se que a probabilidade deste paciente ter câncer é de 1%, ignorando qualquer outra evidência que possa existir.

Nos exemplos anteriores foram demonstradas probabilidades de um evento, sem que nenhuma outra informação tenha sido observada. Casos como estes são conhecidos por *probabilidade a priori*. Estes casos só ocorrem quando não existem novas informações dentro do escopo, caso contrário, utiliza-se a probabilidade condicional.

### 1.1.2 Probabilidade Condicional

Sempre que uma nova evidência sobre alguma variável aleatória é obtida, a probabilidade atual é condicionada à essa nova evidência. A probabilidade condicional demonstra como um evento se comporta dada determinada evidência. Exemplificando esta situação, pode-se pensar no seguinte exemplo: qual a probabilidade de um paciente possuir câncer dado o fato que ele é fumante?

Segundo Moore (2005), a probabilidade condicional de um evento  $A$  ocorrer, sendo que  $P(A) > 0$  é definida por:

$$P(A|B) = \frac{P(A \wedge B)}{P(A)} \quad (1.2)$$

Os eventos A e B ocorrem de forma distinta, sendo A a probabilidade que está sendo calculada e B a informação repassada.

### 1.1.3 O Teorema de Bayes

O Teorema de Bayes foi criado pelo reverendo Thomas Bayes, sendo publicado póstumamente em 1764 (ALDRICH, 2008). O teorema é a base para a Inferência Bayesiana. No teorema, Bayes demonstra como alterar probabilidades considerando novas evidências, e assim, obtendo novas probabilidades. A seguir, o teorema é demonstrado (NEAPOLITAN, 2003)

Dados dois eventos E e F, sendo que  $P(E) \neq 0$  e  $P(F) \neq 0$ , tem-se:

$$P(E|F) = \frac{P(F|E)P(E)}{P(F)} \quad (1.3)$$

Onde:

- $P(E|F)$  significa a probabilidade do evento E ocorrer dado que F ocorreu, é chamada de probabilidade posterior.
- $P(E)$  é a probabilidade *a priori* do evento E ocorrer. Esta probabilidade foi calculada previamente, antes da evidência (ou evento) F ocorrer.
- $P(F)$  é a probabilidade marginal, ou seja, a probabilidade de testemunharmos o evento F. A probabilidade marginal pode ser calculada através da soma e produto de todas as probabilidades mutuamente exclusivas de um determinado conjunto:  

$$P(F) = \sum_i P(F|H_i) + P(H_i)$$

### 1.1.4 Inferência Bayesiana

Inferência Bayesiana é um método estatístico de inferência. Este método é utilizado quando precisa-se atualizar ou inferir uma probabilidade de determinado evento. Como citado anteriormente, a inferência Bayesiana faz uso constante do Teorema de Bayes.

Realizar este tipo de inferência consiste basicamente em aplicar o Teorema de Bayes

para descobrir determinada probabilidade. Deve-se identificar possíveis evidências, e estimar suas probabilidades iniciais (*a priori*).

Exemplificando a inferência Bayesiana, observe o seguinte caso, extraído de Neapolitan (2003):

Suponha que Joe precisa fazer um exame de Raio-X rotineiro para admissão em um novo emprego. Ao ver o resultado, Joe entra em pânico ao descobrir que seu exame é positivo para câncer de pulmão. Mas qual será a real probabilidade de Joe possuir câncer?

Sabe-se que o exame não é totalmente absoluto. É conhecida uma taxa de 60% de acertos, também sabe-se que em 2% dos casos o exame aponta um falso positivo. Para descobrir qual a real probabilidade de Joe possuir câncer de pulmão, recorreremos ao Teorema de Bayes.

Uma busca mais apurada de informações mostrou a Joe que apenas um em cada mil funcionários possui câncer de pulmão. Logo, a probabilidade de se ter cancer (independente de teste) é de 1%. Com essas informações apuradas, aplica-se o Teorema de Bayes:

$$P(\text{cancer}|\text{positivo}) = \frac{P(\text{pos}|\text{cancer})P(\text{cancer})}{P(\text{pos}|\text{cancer})P(\text{cancer}) + P(\text{pos}|\text{sadio})P(\text{sadio})} \quad (1.4)$$

$$P(\text{cancer}|\text{positivo}) = \frac{P(0,6)P(0,001)}{P(0,6)P(0,001) + P(0,02)P(0,99)} \quad (1.5)$$

$$P(\text{cancer}|\text{positivo}) = 0,23 \quad (1.6)$$

- onde "pos" indica positivo.

Conclui-se, com isso, que a probabilidade de Joe possuir câncer, dado todos os fatos levantados, é de 23%

A inferência Bayesiana vem sendo utilizada em inteligência artificial e sistemas especialistas. A inferência Bayesiana foi aplicada com sucesso em sistemas para filtragem de e-mails indesejados (SAHAMI et al., 1998). Destaca-se, nesta área, o trabalho de Graham, chamado *A Plan for Spam* (GRAHAM, 2002)

## 1.2 As Redes Bayesianas

O uso simples do Teorema de Bayes torna-se difícil ao analisar-se problemas complexos. Problemas de maior complexidade não podem ser resolvidos com uma simples aplicação do Teorema de Bayes. Para estas situações, usa-se Redes Bayesianas, já que elas exploram dependências condicionais entre as variáveis.

Trabalhar com inferência Bayesiana é uma tarefa um tanto quanto simples quando o problema envolve poucas variáveis interrelacionadas. Porém, em situações em que tem-se um número grande de variáveis, cada uma afetando direta ou indiretamente a tomada de decisão, esta abordagem se torna impraticável.

Pode-se usar como exemplo de um problema maior, a clássica situação do ladrão / terremoto, exposta por Pearl (1988): *"Estou no trabalho e recebo uma ligação de meu vizinho 1, ele me avisa que meu alarme está disparando. Porém, meu vizinho 2 não liga avisando. Meu alarme pode ser acionado por um pequeno terremoto. Ou será que um ladrão entrou em minha casa"*. Neste exemplo, demonstrado graficamente na figura 1, um número maior de situações deve ser analisado. Um ladrão pode disparar o alarme; um pequeno terremoto pode disparar o alarme; o alarme disparando pode fazer o vizinho 1 telefonar; o alarme disparando pode fazer o vizinho 2 telefonar.

A situação anterior, apesar de ainda ser um caso simples, mostra uma relação entre vários acontecimentos. Um exemplo bem mais complexo seria um conjunto de variáveis para um diagnóstico médico, composto por várias evidências que podem ou não levar à uma doença.

Utilizando-se ainda do exemplo mostrado na sessão anterior, acrescenta-se a seguinte situação: o fato do paciente ter um histórico de fumo afeta (e quanto) ou não a presença de uma anomalia em seu exame. O fato do indivíduo ser fumante pode levar à presença de bronquite? Este histórico de bronquite pode favorecer a presença de câncer de pulmão?

No caso citado anteriormente, é preciso inferir sobre um grande número de variáveis, muitas das quais não estão relacionadas por influência direta. O uso de Inferência Bayesiana se torna extremamente complexo nestes casos, já que o número de cálculos a ser realizado tende a ser muito grande.

As Redes Bayesianas resolvem este problema, calculando as probabilidades conjuntas de um grande número de variáveis, e efetuando inferência sobre as mesmas. As Redes Bayesianas são representadas por grafos acíclicos dirigidos (*directed acyclic graph*, DAG),

onde cada variável do escopo é associada a um nó do gráfico e as relações entre as variáveis são representadas pelas arestas do grafo.

De acordo com Castillo et al (1998), a representação gráfica de modelos probabilísticos têm a vantagem de mostrar explicitamente as relações entre as variáveis e conservar estas relações de forma qualitativa. Os modelos gráficos são também mais intuitivos.

Russel e Norvig (2004) definiram uma Rede Bayesiana como um grafo orientado, onde cada nó é identificado com informações de probabilidades quantitativa. Neste grafo, um conjunto de variáveis aleatórias constituem os nós da rede. Elas podem ser discretas ou contínuas.

Os nós desta rede são conectados entre si (em pares e sem ciclos). Se houver uma seta do nó  $X$  até o nó  $Y$ , diz-se que  $X$  é pai de  $Y$ .

A cada nó de uma Rede Bayesiana, é atribuída uma tabela de probabilidades condicionais. Esta tabela demonstra as probabilidades do evento ocorrer, dado que seus pais tenham ou não ocorrido. Caso o nó não possua um pai, sua tabela de probabilidades é reduzida à probabilidade incondicional daquele evento ocorrer.

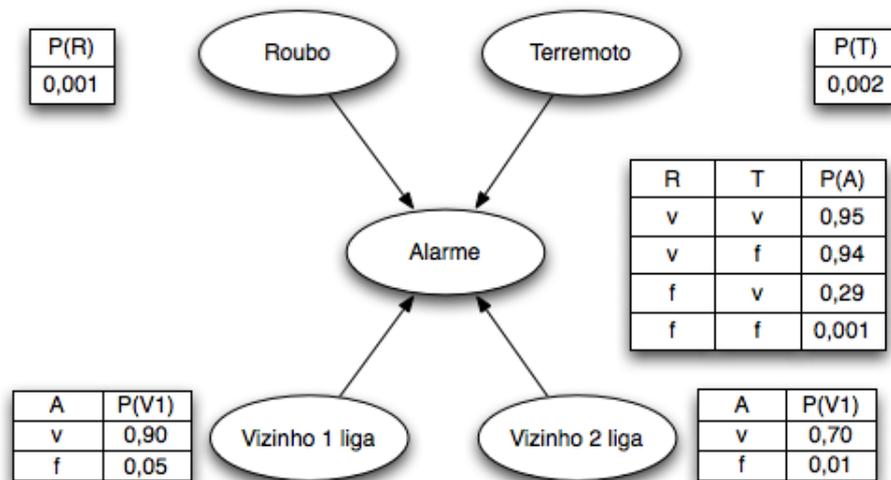


Figura 1: Rede Bayesiana do problema do terremoto/ladrão. Esta rede além da topologia, demonstra a tabela de probabilidades condicionais (RUSSEL; NORVIG, 2004)

### 1.2.1 Formalização Matemática

As Redes Bayesianas fornecem uma descrição total do domínio em estudo. Qualquer entrada na rede pode ser calculada considerando as outras entradas armazenadas na rede. Considerando estas características, chega-se a seguinte fórmula (RUSSEL; NORVIG, 2004):

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{pais}(X_i)) \quad (1.7)$$

Visualizando a fórmula anterior, deduz-se que a probabilidade de cada entrada na rede é representada pelo produto de suas probabilidades dados seus antecessores. Como exemplo, observe a seguinte rede:

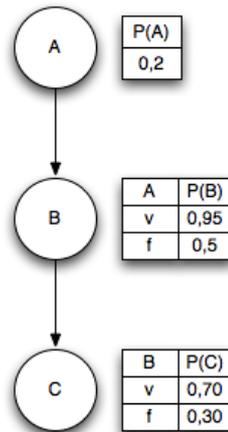


Figura 2: Rede Bayesiana com tabela de probabilidades condicionais

Na rede anterior, caso queira-se calcular a probabilidade do nó C ocorrer, dado que A e B ocorram, deve-se observar as distribuições de probabilidade de todos seus antecessores (pais). A seguir, uma demonstração do cálculo:

$$P(c \wedge b \wedge a) = P(c|b)(Pb|a)P(a) = 0,70 * 0,95 * 0,2 = 0,133 \quad (1.8)$$

### 1.2.2 Independência condicional nas Redes Bayesianas

Segundo Charniak (CHARNIAK, 1991), uma objeção ao uso da teoria da probabilidade é que a especificação completa de uma distribuição requer um número absurdo de números. Por exemplo, se o problema possui  $n$  variáveis aleatórias, sua distribuição completa é representada por  $2^{\sup n}$  probabilidades conjuntas. Ainda segundo Charniak, as Redes Bayesianas incorporam a independência na sua própria estrutura, reduzindo assim consideravelmente o número de variáveis.

Pode-se visualizar as Redes Bayesianas como distribuições compactas para uma série de independências condicionais assumidas em uma distribuição. Esta interpretação nos dá a visão de Rede Bayesiana como provedora de fatorização de uma distribuição. Em outras

palavras, cada nó  $X$  é independente de seus não dependentes, dados seus pais (KOLLER et al., 2007).

A independência que está implícita em uma Rede Bayesiana indica que cada variável é independente de seus não descendentes, dados seus pais (ZHANG; POOLE, 1996). Considerando esta independência, o número de probabilidades requeridas é drasticamente reduzido.

Um procedimento chamado d-separação (separação direta) identifica relações de independência nas Redes Bayesianas. Dois nós em uma Rede Bayesiana estão d-separados sempre que as evidências precisam "passar através" de um nó intermediário, para afetar o destino final. Observe na figura 3 que o nó  $Y$  está d-separado do nó  $X$ .

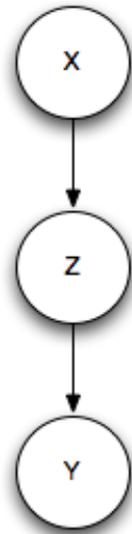


Figura 3: D-separação entre  $X$  e  $Y$

### 1.2.3 Trabalhos Correlatos

As Redes Bayesianas têm sido utilizadas em várias áreas onde tomadas de decisão baseadas em informações incompletas precisam ser realizadas. A área que mais se utiliza destas técnicas é, sem dúvida, a medicina, principalmente para realização de diagnóstico. A seguir, alguns trabalhos utilizando Redes Bayesianas são expostos.

O **Projeto AMPLIA** (FLORES et al., 2003) é uma ferramenta para apoio do desenvolvimento do raciocínio diagnóstico. Ele utiliza Redes Bayesianas para representação do conhecimento.

NasoNet (GALÁN et al., 2001) é um sistema que realiza diagnóstico de câncer nasofaríngeal. Ele utiliza uma Rede Bayesiana composta por mais de cem nós, que indicam hemorragias, infecções, sintomas, etc.

## 2 REALIZANDO INFERÊNCIA NAS REDES BAYESIANAS

Segundo Russel et al (2004), a tarefa básica de inferência em uma rede Bayesiana consiste no cálculo da distribuição de probabilidades posteriores, para um conjunto de variáveis, dado algum evento observado. Basicamente, calcular as probabilidades requer a avaliação de todos os nós da rede, considerando cada evidência que levou ao mesmo.

O principal problema que observa-se na avaliação de Redes Bayesianas é o tempo elevado que uma busca exata requer. Uma série de algoritmos de aproximação foram desenvolvidos para solucionar este problema. No decorrer deste capítulo, diferentes métodos de inferência serão demonstrados e avaliados.

### 2.1 A complexidade dos algoritmos de inferência

O maior desafio no uso de Redes Bayesianas é a complexidade dos algoritmos de inferência. Normalmente, a avaliação exata de uma rede requer um algoritmo de complexidade NP-Difícil (G.F.COOPER, 1991). Problemas de complexidade NP-Difícil (NP-Hard) não podem ser resolvidos em tempo polinomial (GAREY; JOHNSON, 1979).

De acordo com Charniak (1991), há apenas uma classe restrita de redes que podem ser resolvidas de forma exata em tempo aceitável. Estas redes não possuem mais do que um caminho entre dois nós. Estas redes são conhecidas como *singly connected network*. Existem algoritmos que resolvem Redes Bayesianas explorando a estrutura da rede em particular, ou então trabalhando com fatorização das probabilidades conjuntas (ZHANG; POOLE, 1994).

Para evitar o uso de algoritmos NP-Difícil, pode-se optar pelo uso de técnicas de inferência aproximada. Estas técnicas fornecem respostas próximas as exatas, mas evitam a necessidade de um algoritmo NP-Difícil. Na maioria das vezes, estes algoritmos definem valores aleatórios em alguns nós e, após esta definição, calculam probabilidades baseadas nestas valores.

## 2.2 Exemplificação da Inferência em uma Rede Bayesiana

Esta seção demonstra como se dá a inferência em uma Rede Bayesiana. Este conceito básico é aplicado à maioria dos algoritmos.

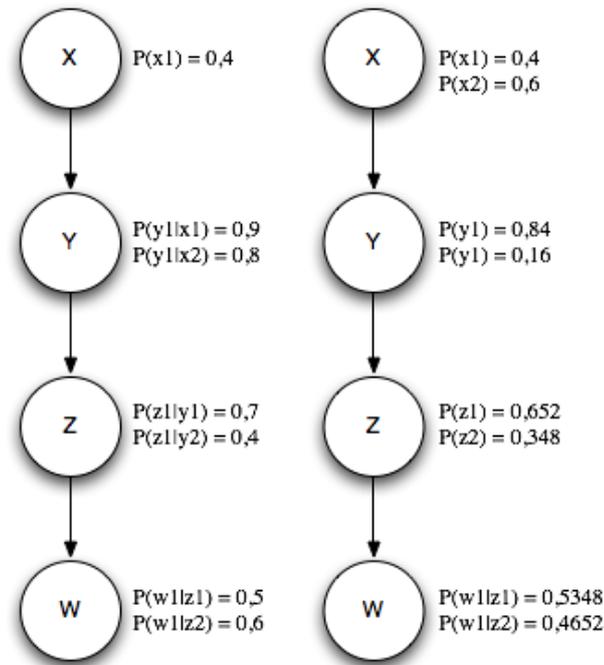


Figura 4: Rede Bayesiana com suas probabilidades a priori à esquerda (NEAPOLITAN, 2003)

Considerando a rede demonstrada na figura 4, para calcular as probabilidades de cada variável, necessita-se de informações determinadas por seus pais.

O cálculo das probabilidades do exemplo anterior se dá da seguinte forma:

$$P(y1) = P(y1|x1)P(x1) + P(y1|x2)P(x2) = 0,84$$

$$P(z1) = P(z1|y1)P(y1) + P(z1|y2)P(y2) = 0,652$$

$$P(w1) = P(w1|z1)P(z1) + P(w1|z2)P(z2) = 0,5348$$

## 2.3 Algoritmos de Inferência Exata

A inferência exata de uma Rede Bayesiana consiste no cálculo da distribuição de probabilidade para um conjunto de variáveis (RUSSEL; NORVIG, 2004). Como a inferência exata, no caso geral, é um problema intratável, apenas um algoritmo é demonstrado neste trabalho. Na sessão seguinte, algoritmos de aproximação serão estudados.

### 2.3.1 Inferência por enumeração

Citando novamente Russel (2004) qualquer probabilidade condicional pode ser calculada através do somatório da distribuição total. Uma consulta  $P(X|e)$  pode ser respondida através da equação 2.1.

$$P(X|e) = \alpha P(X, e) = \alpha \sum_y P(X, e, y) \quad (2.1)$$

Observando a equação anterior, conclui-se que uma consulta à uma Rede Bayesiana pode ser respondida, calculando-se as somas dos produtos de probabilidades condicionais na rede. Considerando a Rede Bayesiana representada na figura 1, demonstrada no capítulo 1, caso queira-se fazer a seguinte busca na rede:  $P(\text{Roubo} | \text{Vizinho 1 liga} = \text{verdadeiro}, \text{Vizinho 2 liga} = \text{verdadeiro})$ , obtém-se a seguinte expressão:

$$P(R|v1, v2) = \alpha P(R, v1, v2) = \alpha \sum_e \sum_a P(R, e, a, v1, v2) \quad (2.2)$$

Onde:

- R = roubo
- v1 = vizinho 1
- v2 = vizinho 2

No algoritmo, demonstrado a seguir, a complexidade de processamento cresce linearmente de acordo com as variáveis acrescentadas à Rede Bayesiana, porém, a complexidade em tempo de processamento cresce na ordem de  $2n$ , onde  $n$  é o número de variáveis.

Abaixo, uma demonstração do algoritmo em linguagem estruturada (RUSSEL; NORVIG, 2004):

```
entradas: X, a variável de consulta
          e, valores observados para as variáveis
          rb, uma Rede Bayesiana
```

```
função ENUMERAÇÃO(X, e, rb) retorna uma distribuição sobre X
  Q(X) <- uma distribuição sobre X, inicialmente vazia
```

```

para cada valor x1 em X faça
    estender e com valor x1 para X
    Q(x1) <- ENUMERAR_TOOS(VARS(rb,e))

```

```

retornar NORMALIZAR(Q(X))

```

```

função ENUMERAR-TODOS(vars,e) retorna um número real
    se VAZIO(vars) então retornar 1
    Y <- PRIMEIRO(vars)
    se Y tem valor y em e então
        retornar (Py | pais(Y)) * ENUMERAR-TODOS(RESTO(vars),ey)

```

### 2.3.2 Algoritmo de eliminação de variáveis

Segundo Zhang e Poole (1996), a chave para uma inferência mais eficiente está no conceito de fatorização. A fatorização de uma probabilidade conjunta é a lista de fatores (funções), das quais as probabilidades conjuntas podem ser construídas.

Nesta sessão, será demonstrado o algoritmo de eliminação de variáveis (*Variable Elimination*) descrito em Zhang e Poole (1996). Este algoritmo deriva do algoritmo conhecido como *Bucket Elimination*, proposto por Dechter (1996). Este algoritmo tem esse nome pois ele soma as variáveis de uma lista de fatores um-a-um. Uma ordem de eliminação das variáveis deve ser informada, ela é chamada de ordem de eliminação (*elimination ordering*).

Russel e Russel (2004), apresentam o algoritmo de forma mais simplificada. De acordo com os autores, a eliminação de variáveis funciona avaliando expressões (equações) da direita para a esquerda. Os resultados intermediários são então armazenados, e os somatórios, são efetuados apenas para as partes da expressão que dependem de determinada variável. Além disso, toda variável que não é ancestral das variáveis de consulta ou evidência podem ser removidas, já que elas são irrelevantes à pesquisa.

A seguir, uma demonstração do algoritmo em linguagem estruturada (RUSSEL; NORVIG, 2004):

```

entradas: X, a variável de consulta
          e, vtores observados para variáveis E

```

rb, uma Rede Bayesiana

```

função ELIMINAÇÃO(X, e, rb) retorna uma distribuição sobre X
  fatores <- []
  vars <- REVERTER(VARS[rb])
  para cada var em vars faça
    fatores <- CRIAR-FATOR(var,e)
    se var é uma variável oculta então
      fatores <- SOMAR(var,fatores)

  retornar NORMALIZAR(PRODUTO(fatores))

```

### 2.3.3 Inferência por árvores de junção

A inferência em Redes Bayesianas pode ocorrer de forma eficiente ao se trabalhar com estruturas computacionais auxiliares. Estas estruturas são conhecidas como árvores de junção, ou *junction trees*. Esta abordagem foi apresentada por Jensen e Jensen (1994).

Segundo Ladeira (1999), uma árvore de junção  $T$  de um grafo  $G$  é uma árvore cujos nós são agregados de  $G$  e:

1. para cada nó  $X$  de  $G$ , existem nós em  $T$  que contém  $X$ .
2. para cada par de agregados (cliques)  $C_i, C_k$ , todos os agregados entre eles contém  $C_i \cap C_k$ .
3. cada arco é rotulado com as variáveis comuns aos agregados que os une (separador).

### 2.3.4 Transformação topológica e construção da árvore de junção

Antes de aplicar-se o algoritmo de inferência, uma transformação topológica deve ser realizada na Rede Bayesiana, transformando-a em um árvore de junção. Esta transformação se dá em quatro passos: moralização, triangularização, identificação de cliques e construção da árvore.

A moralização consiste em adicionar arcos entre cada par de pais de cada nó do grafo original. Em seguida, deve-se eliminar a orientação dos arcos. Abaixo, demonstração de um grafo antes e após a moralização.

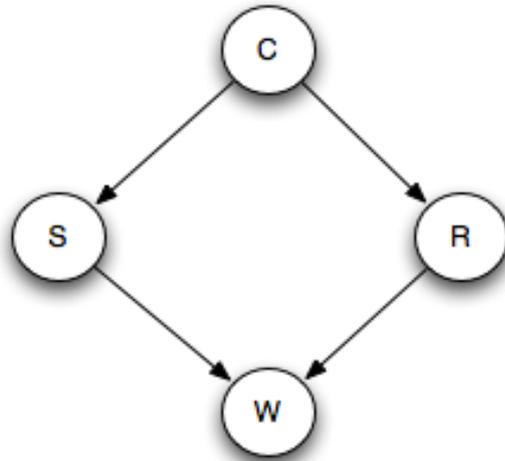


Figura 5: Grafo antes do processo de moralização

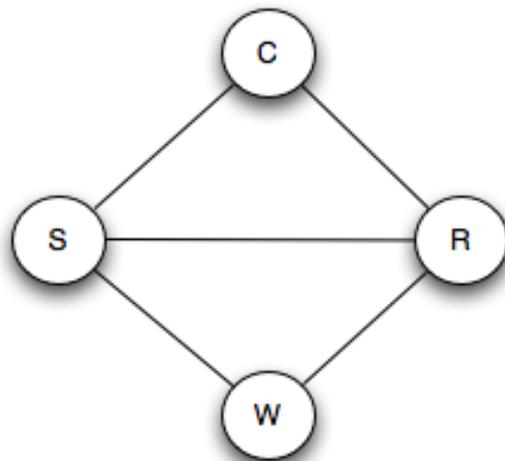


Figura 6: Grafo após o processo de moralização

O passo seguinte é conhecido como triangularização. Ele consiste na inserção de arcos em *loops* com mais de três nós. Cada arco inserido decompõe o *loop* em dois *loops* menores (LADEIRA; VICARI, 1999).

Ainda, segundo Ladeira 1999, um grafo não orientado é triangular se, e somente se, todos os seus nós podem ser eliminados, um a um, sem a adição de qualquer arco. Abaixo, um exemplo de um grafo após o processo de triangularização. Os arcos pontilhados foram incluídos pelo processo.

Seja  $G_t$ , um grafo tringular,  $A_1, \dots, A_n$ , uma seqüência de eliminação de seus nós, e  $C_i$ , o conjunto contendo  $A_i$ , e seus vizinhos no instante de sua eliminação (i.e os vizinhos do

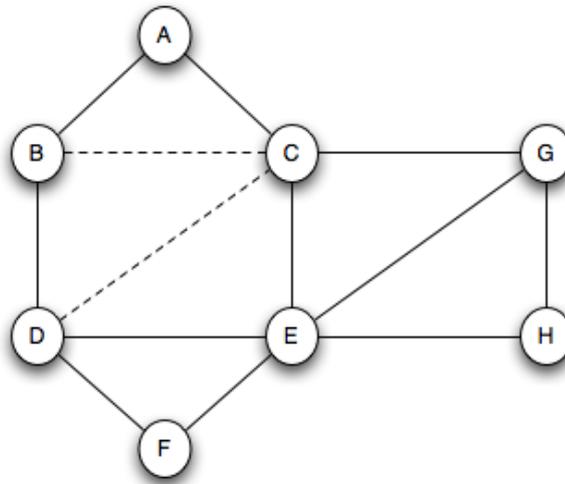


Figura 7: Grafo após o processo de triangularização

nó eliminado, com numeração  $A$ , maior do que a numeração  $A$  do nó sendo eliminado). Então  $C_i$  é um clique de  $G_t$  se  $C_i$  é máximo, se não está contido em nenhum outro clique (LADEIRA; VICARI, 1999 apud JENSEN, 1996).

No algoritmo de construção da árvore de junção, deve-se definir um índice para cada vértice da rede. Este índice, é o inverso da ordem de eliminação da fase de triangularização. Para cada conjunto de cliques no grafo, cria-se uma lista de vértices adjacentes. Se a lista de vértices adjacentes possui um vértice com índice menor do que o vértice de maior índice no clique, então este vértice recebe o valor do maior índice. O clique com índice 1, deve ser tornar a raiz da árvore.

Após a etapa anterior ser realizada, deve-se criar os separadores entre os cliques. Chama-se separador, o conjunto de vértices que contém a intersecção de vértices entre dois cliques. Caso um clique possua uma intersecção com dois cliques, o clique com menor índice é usado como separador.

### 2.3.5 Inferência na árvore de junção

Após a construção da árvore de junção, o algoritmo de inferência segue basicamente 5 passos (KJAERULFF; MADSEN, 2005). São eles:

1. Cada item de evidência deve ser incorporado aos potenciais da árvore de junção. Para cada ítem de evidência, uma função de evidência é multiplicada no clique apropriado.

2. O clique raiz deve ser selecionado. Ele é conhecido como o clique raiz da propagação.
3. Mensagens são passadas por todos os separadores da árvore de junção, até o clique raiz. Estas mensagens, atualizam os potenciais dos clique e dos separadores da árvore de junção.
4. Mensagens são passadas novamente, porém, na direção contrária. Esta fase é conhecida como distribuição de informação.
5. Neste momento, a árvore é dita em equilíbrio. A probabilidade  $P(X|e)$ , pode ser computada à partir de qualquer clique ou separador que contenha  $X$ .

## 2.4 Inferência aproximada em Redes Bayesianas

Como foi tratado anteriormente, o uso de inferência exata pode levar à situações intratáveis. O alto tempo de execução destes algoritmos, pode muitas vezes, tornar a tarefa de inferência em uma Rede Bayesiana impossível.

Para reduzir o tempo de execução, pode-se utilizar algoritmos de inferência aproximada nas Redes. Estes algoritmos fazem uma aproximação das distribuições conjuntas de algumas variáveis da rede (KOLLER et al., 2007). Estas instâncias são conhecidas como amostras. Estas amostras representam parte das probabilidades do problema.

Segundo Castillo (1998), a idéia básica dos algoritmos de inferência aproximada é gerar uma amostra de tamanho  $N$ , a partir da função de probabilidade conjunta das variáveis. Estas amostras são utilizadas para calcular valores aproximados das probabilidades de certos nós, dado a evidência.

Ainda, segundo Castillo, os algoritmos de inferência aproximada podem ser classificados em dois tipos: métodos de simulação estocástica e métodos de busca determinística. Os métodos estocásticos geram as amostras tomando como base a função de probabilidade conjunta, e também, usam mecanismos aleatórios. Já os métodos determinísticos, geram as amostras de forma sistemática.

### 2.4.1 Inferência por simulação de cadeias de Markov

Este método foi proposto por Pearl (1987). Segundo Castillo (1998), este método consiste em designar aos nós de evidências seus respectivos valores e, logo após, simular estocasticamente a rede resultante.

O algoritmo gera cada evento fazendo uma mudança aleatória no evento precedente. Cada estado da rede é gerado por amostragem aleatória de um valor em variáveis que não são de evidência. O algoritmo trabalha invertendo uma variável de cada vez, mas mantém fixas as variáveis de evidência.

Toma-se como exemplo a rede exposta por Russel em RUSSEL; NORVIG 2004, demonstrada aqui na figura que segue.

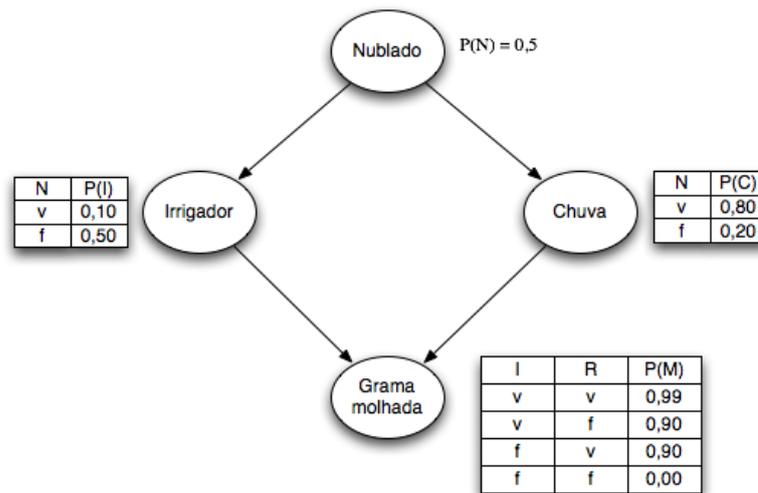


Figura 8: (RUSSEL; NORVIG, 2004)

Ainda seguindo o exemplo demonstrado por Russel, executa-se a seguinte consulta na rede:  $Chuva/Irrigador = v$ ,  $GramaMolhada = v$  aplicado a rede demonstrada anteriormente. As variáveis de evidência, neste caso  $Irrigador$  e  $GramaMolhada$  são definidas com seus respectivos valores observados. As variáveis ocultas, que neste caso são  $Nublado$  e  $Chuva$ , são inicializadas com valores aleatórios (para este exemplo considera-se os valores verdadeiro e falso, respectivamente). As seguintes etapas são então executadas repetidamente:

A variável  $Nublado$  é amostrada, levando em conta suas variáveis pais. Observando o exemplo anterior, a amostra é dada por  $P(Nublado/Irrigador=v, Chuva=f)$ . Suponha, para este exemplo que o resultado da amostragem seja  $Nublado = falso$ .

Cada amostra gerada contribui para o resultado da variável de consulta, neste caso  $Chuva$ . Se o algoritmo visitar  $X$  estados em que a variável possui valor verdadeiro e  $Y$  estados em que seu valor é falso, a resposta à consulta efetuada será a normalização dos valores de  $X$  e  $Y$ .

Castillo (1998) resume o algoritmo de simulação de cadeias de Markov em três passos

básicos. São eles:

1. Deve-se atribuir a cada uma das variáveis não evidenciais um valor aleatório qualquer.
2. Deve-se criar uma ordenação aleatória para selecionar os nós não evidenciais. Para cada variável da lista, se gera um valor aleatório. Este valor é gerado à partir da distribuição de probabilidades conjuntas das variáveis. Normaliza-se então as probabilidades.
3. Repete-se a etapa dois para  $N$  extrações

A seguir, uma demonstração do algoritmo em linguagem estruturada (RUSSEL; NORVIG, 2004):

```

função MARKOV(X,e,rb,N) retorna uma estimativa de P(X|e)
  variáveis: N[X]: um vetor de contagens sobre X (inicia em zero)
             Z: as variáveis não de evidência em rb
             x: o estado atual da rede, copiado de e

  inicializar x com valores aleatórios para as variáveis em Z
  para j=1 até n faça
    N[x] <- N[x] + 1
    para cada variável em Z faça
      fazer a amostragem de Z em x

  retornar NORMALIZAR(N[X])

```

A amostragem, citada no algoritmo demonstrado anteriormente, se dá através da computação de todas as variáveis da *cobertura de Markov*, da variável em evidência naquele momento. A cobertura de Markov de uma variável, consiste em seus pais, filhos e pais dos filhos (RUSSEL; NORVIG, 2004).

### 3 PROPOSTA DE TRABALHO

Neste capítulo, será apresentada a proposta para a criação de um módulo de Redes Bayesianas que será integrado ao projeto AMPLIA. Este módulo, como tratado em capítulos anteriores, possibilita ao aluno representar seu conhecimento de forma gráfica e probabilística. Este módulo também permitirá a comparação de duas redes, uma criada pelo estudante e outra, pelo especialista.

#### 3.1 O software UnBBayes

O UnBBayes (UNBBAYES, 2009) é um software, de código fonte livre, para modelagem, aprendizagem e avaliação de redes probabilísticas. Além das Redes Bayesianas, o UnBBayes trabalha também com diagramas de influências, que poderão ser utilizados em trabalhos futuros.

Por ser software livre e desenvolvido em Java, o UnBBayes atende às necessidades básicas do projeto AMPLIA. O UnBBayes também traz um editor de Redes Bayesianas e uma série de algoritmos de inferência, que servirão como base no projeto AMPLIA.

#### 3.2 Características das Redes Bayesianas no projeto AMPLIA

As Redes Bayesianas, dentro do projeto AMPLIA, possuem uma série de características próprias. Cada nó da rede, além das propriedades básicas, como probabilidades iniciais e tabelas de probabilidades condicionais, deverão possuir informações como links, imagens, informações adicionais, diagnósticos associados à ela e custo de uma decisão.

Estas características serão implementadas estendendo as funcionalidades existentes no UnBBayes. Estas novas características podem influenciar no resultado final da rede. Sendo assim, os algoritmos de inferência que agem na rede precisarão ser adaptados à essa nova realidade.

## CONSIDERAÇÕES FINAIS

Inicialmente foram estudados todos os aspectos teóricos das Redes Bayesianas. O primeiro capítulo deste trabalho se aprofunda nas teorias das Redes Bayesianas, tratando sobre a teoria das probabilidades, teorema de Bayes, inferência bayesinas e também toda a fundamentação matemática envolvida.

Com base no conhecimento adquirido no capítulo um, o segundo capítulo deste trabalho aborda métodos de inferência em Redes Bayesianas. Estes métodos são utilizados para propagar evidências dentro da rede, atualizando assim probabilidades e permitindo a tomada de decisões baseadas nos resultados obtidos.

O capítulo três traz a proposta de trabalho dentro do projeto AMPLIA. Neste capítulo, é proposto a utilização do software UnBBayes. Como o UnBBayes já trabalha com redes Bayesianas, serão necessárias algumas modificações em sua estrutura, a fim de atender as necessidades do AMPLIA.

Considerando todo o conhecimento adquirido, e a ótima proposta que o projeto AMPLIA traz, este trabalho é totalmente viável e promissor. Acredita-se também que este projeto traga grande contribuição para o meio acadêmico, já que o mesmo será utilizado para aprimorar a formação de novos profissionais da saúde.

Como continuação deste trabalho, pretende-se implantar, de forma efetiva, o módulo de construção e avaliação de Redes Bayesianas dentro do software AMPLIA. Outros módulos do projeto AMPLIA já estão em desenvolvimento em outros trabalhos.

## Referências

- ALDRICH, J. R. a. fisher on bayes and bayes theorem. *Journal of the International Society for Bayesian Analysis*, v. 1, n. 1, p. 161–170, 2008.
- CASTILLO, E.; GUTIÉRREZ, J. M.; HADI, A. S. *Sistemas Expertos e Modelos de Redes Probabilísticas*. [S.l.]: Academia Espanola de Ingenieria, 1998.
- CHARNIAK, E. Bayesian networks without tears. *AI Magazine*, 1991.
- DECHTER, R. Bucket elimination: A unifying framework for probabilistic inference. 1996.
- FLORES, C. et al. Projeto amplia - uso de informática na educação médica. 2003.
- FLORES, C. D. *Negociação Pedagógica Aplicada a um Ambiente Multiagente de Aprendizagem Colaborativa*. Tese (Doutorado em Ciência da Computação) — PPGC / UFRGS, 2005.
- GALÁN, S. F. et al. Nasonet, joining bayesian networks and time to model nasopharyngeal cancer spread. In: *AIME '01: Proceedings of the 8th Conference on AI in Medicine in Europe*. London, UK: Springer-Verlag, 2001. p. 207–216.
- GAREY, M. R.; JOHNSON, D. S. *Computers and Intractability. A guide to the Theory of NP-Completeness*. [S.l.]: W. H. Freeman, 1979.
- G.F.COOPER. *Probabilistic Inference Using Belief Networks is NP-Hard*. [S.l.], 1991.
- GRAHAM, P. A plan for spam. <http://www.paulgraham.com/spam.html>, 2002.
- JENSEN, F. V. *An Introduction to Bayesian Networks*. [S.l.]: UCL Press, 1996.
- JENSEN, F. V.; JENSEN, F. Optimal junction trees. In: *In UAI*. [S.l.]: Morgan Kaufmann, 1994. p. 360–366.
- KJAERULFF, U. B.; MADSEN, A. L. *Probabilistic Networks - An introduction to Bayesian Networks and Influence Diagrams*. [S.l.: s.n.], 2005.
- KOLLER, D. et al. Graphical models in a nutshell. In: GETOOR, L.; TASKAR, B. (Ed.). *Introduction to Statistical Relational Learning*. [S.l.]: MIT Press, 2007.
- LADEIRA, M.; VICARI, R. M. *Algoritmos de Inferência em Redes Probabilísticas Baseados em Árvore de Junções*. [S.l.], 1999.
- MOORE, D. S. *A Estatística Básica e Sua Prática*. [S.l.]: LTC, 2005.
- NEAPOLITAN, R. E. *Learning Bayesian Networks*. 1. ed. [S.l.]: Prentice Hall, 2003.

- PEARL, J. *Evidential Reasoning Using Stochastic Simulation of Causal Models*. [S.l.], 1987.
- PEARL, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. [S.l.]: Morgan Kaufmann, 1988.
- RUSSEL, S.; NORVIG, P. *Inteligencia Artificial*. 2nd. ed. [S.l.]: Editora Campus, 2004.
- SAHAMI, M. et al. A bayesian aproach to filering junk e-mail. *AAAI'98 Workshop on Learning for Text Categorization*, 1998.
- UNBBAYES. Junho 2009. Disponível em: <[unbbayes.sourceforge.net/](http://unbbayes.sourceforge.net/)>.
- ZHANG, N.; POOLE, D. A simple approach to bayesian network computations. In: *Proceedings of the Tenth Canadian Conference on Artificial Intelligence*. [S.l.: s.n.], 1994. p. 171–178.
- ZHANG, N. L.; POOLE, D. Exploiting causal independence in bayesian network inference. *Journal of Artificial Intelligence Research*, v. 5, p. 301–328, 1996.