



**CENTRO UNIVERSITÁRIO LUTERANO DE PALMAS**

COMUNIDADE EVANGÉLICA LUTERANA "SÃO PAULO"  
Recredenciado pela Portaria Ministerial nº 3.607 - D.O.U. nº 202 de 20/10/2005

DOUGLAS JOSÉ RAMALHO ARAUJO

**MODELO DE INFERÊNCIA DE PREFERÊNCIAS DE CLIENTES EM  
SHOPPING CENTERS**

Palmas - TO

2015

**DOUGLAS JOSÉ RAMALHO ARAUJO**  
**MODELO DE INFERÊNCIA DE PREFERÊNCIAS DE CLIENTES EM**  
**SHOPPINGS CENTERS**

Trabalho de Conclusão de Curso (TCC) elaborado e apresentado como requisito parcial para obtenção do título de bacharel em Sistemas de Informação pelo Centro Universitário Luterano de Palmas (CEULP/ULBRA).

Orientador: Prof. M.Sc. Jackson Gomes de Souza.

Palmas – TO

2015

**DOUGLAS JOSÉ RAMALHO ARAUJO**  
**MODELO DE INFERÊNCIA DE PREFERÊNCIAS DE CLIENTES EM**  
**SHOPPINGS CENTERS**

Trabalho de Conclusão de Curso (TCC) elaborado e apresentado como requisito parcial para obtenção do título de bacharel em Sistemas de Informação pelo Centro Universitário Luterano de Palmas (CEULP/ULBRA).

Orientador: Prof. M.Sc. Jackson Gomes de Souza.

Aprovada em \_\_/\_\_/2015.

**BANCA EXAMINADORA**

---

Prof. M.Sc. Jackson Gomes de Souza  
Centro Universitário Luterano de Palmas

---

Prof. M.Sc. Parcilene Fernandes Brito  
Centro Universitário Luterano de Palmas

---

Prof. Dr. Edeílson Milhomem da Silva  
Centro Universitário Luterano de Palmas

Palmas – TO  
2015

## AGRADECIMENTOS

Não posso começar meus agradecimentos sem antes falar sobre os principais responsáveis por esta minha conquista, que são meus pais. Não é possível para mim retribuir todo o carinho e amor que recebi da minha mãe dona Cida e do meu pai o senhor José Ramalho, eu os amo profundamente. Uma das formas que tenho de demonstrar o quão grato eu sou por tudo o que vocês fazem por mim é me dedicando aos estudos e superando todos os obstáculos, pois depois de tanto trabalho, eu devo este sucesso a vocês. E por falar em obstáculos, não foram poucos os que encontrei durante o desenvolvimento desta pesquisa, mas noite após noite, todos foram resolvidos graças também a ajuda do meu orientador, o professor Jackson, a quem sou extremamente grato por aceitar este desafio de me orientar, e que também é uma das minhas grandes fontes de inspiração pelo excelente trabalho que executa. Professor, o senhor teve um papel fundamental na minha vida. E para concretizar este trabalho foram meses de ausência do encontro dos meus amigos, que também agradeço pela compreensão de saber que estive em um período difícil, mas mesmo assim, sempre estiveram do meu lado, em especial o grande doutor Giovani, magnânimo do direito, e também o Raimundo de Sagitário, Matheus, Lucas e toda a galera. Posso dizer agora que você está dispensada do cargo de Psicóloga, Cristiane Cunha, por todas as vezes que o desespero batia, mas eu sempre pude contar com uma conversa extremamente agradável que me acalmava. E falando em me aguentar, eis alguém que é profissional em aguentar os meus problemas, Fayad Neto, que há muito tempo superou o que compreendo por definição de amigo, irmão e família, obrigado cara!. E não posso citar família sem mencionar o Flávio e o Joabe, que são meus irmãos e que tanto me ajudaram e me inspiraram sempre que foi necessário, que a nossa amizade supere quaisquer barreiras que o tempo possa colocar, muito obrigado. A minha vitória também dedico a duas pessoas que estiveram comigo durante toda a minha vida, mas que Deus decidiu que levaria para perto de si antes do tempo, Pedro e Gustavo. Pedro, eu sei que você está feliz por mim, eu sei que você está na formatura comigo, e também sei que vai continuar batendo cabeça comigo pois assim que nós fazemos, mesmo que do céu, você está comigo. É difícil segurar dentro de mim as emoções que tenho em relação a perder você em vida, mas nós continuamos juntos meu amigo. E Gustavo, você também me criou, e me amou como um filho, e como filho eu fico aqui orando por você, sempre. E falando em amor, Ruth meu bem, você chegou e ocupou no meu coração um lugar que será sempre seu, sua companhia me acalma, e lutarei para te fazer feliz sempre. E assim, com alegria e amor, com todas as emoções possíveis, agradeço a todos que participam da minha vida, novamente, obrigado.

## RESUMO

O crescimento do número de consumidores que circulam em shoppings centers aumenta a cada ano, isso faz com que comerciantes busquem novas formas de alcançar efetivamente clientes potenciais para conseguir vencer a concorrência que cresce em diversos setores. Com a evolução das Tecnologias da Informação, novas ferramentas e técnicas foram desenvolvidas, de forma que as empresas alcancem o propósito de serem mais eficientes na captação de clientes. Este trabalho busca desenvolver um modelo de inferência, em que seja possível calcular a probabilidade de um cliente estar interessado em alguma categoria de produtos comercializada por determinada loja a partir de um conjunto de características que modelem o interesse do consumidor. Para este objetivo o trabalho busca desenvolver um modelo de inferência para identificar o quão interessado o cliente pode estar em uma categoria de produto, e utiliza como estados das variáveis as características do cliente analisando, e em como elas afetam as lojas que ele pode ou não visitar, e assim, conforme a probabilidade de uma loja aumenta, inferir a categoria em que está disposto a consumir.

**PALAVRAS-CHAVE:** Redes Bayesianas, Tecnologias da Informação, Metodologia.

## LISTA DE FIGURAS

Figura 1: Cenário de aplicação da pesquisa.....	13
Figura 3: Diagrama de Venn para representação da Proposição A união com B, e A intersecção com B .....	19
Figura 4: Probabilidades para a retirada de uma bola azul ou vermelha na primeira tentativa	22
Figura 5: Evento de retirada da segunda bola.....	22
Figura 6: Grafo do problema de tráfego .....	31
Figura 7: Distribuição conjunta em uma Rede Bayesiana.....	32
Figura 8: Relacionamento de dependência de A e B.....	33
Figura 9: Relacionamento de independência entre A e C .....	33
Figura 10: Dependência de A e C caso B seja desconhecida .....	34
Figura 11: Independência de A e C caso B seja conhecido.....	34
Figura 12: Cadeia de dependência entre A e B .....	34
Figura 13: Cadeia de dependência entre A e C .....	35
Figura 14: Relação de Dependência de B.....	35
Figura 15: Rede Bayesiana complexa .....	36
Figura 16: Grafo da Rede Bayesiana para Probabilidades de calçada escorregadia .....	37
Figura 17: Trocando arcos direcionados por não direcionados.....	38
Figura 18: Desenhando arcos entre os pais .....	39
Figura 19: Passos para criação da árvore de junção .....	40
Figura 20: Mapa de trajeto de um cliente no shopping (YAN, 2010).....	42
Figura 21: Modelagem de uma rede bayesiana no UnBBayes.....	47
Figura 23: Relação de lojas e suas variáveis .....	54
Figura 24: Tela inicial da pesquisa web .....	57
Figura 25: Confirmação da primeira fase da pesquisa .....	58

Figura 26: Tela de seleção de lojas.....	58
Figura 27: Fase dois da pesquisa eletrônica .....	59
Figura 28: Diagrama de Banco de Dados.....	60
Figura 29: Tela principal do UnBBayes .....	62
Figura 30: Tela de modelagem da Rede Bayesiana no UnBBayes .....	62
Figura 31: Adicionando e renomeando estados no UnBBayes .....	63
Figura 32: Variáveis da Rede Bayesiana.....	64
Figura 33: Grupo de variáveis gerais.....	65
Figura 34: Relação de Conforto e Atratividade.....	65
Figura 35: Relação entre as variáveis Atratividade e Marcas .....	66
Figura 36: Relação entre Poder de Compra e Marcas .....	66
Figura 37: Grupo de variáveis gerais com seus relacionamentos.....	67
Figura 38: Grupo de variáveis de Lojas.....	67
Figura 39: Relacionamento da variável Marcas com as Demais lojas .....	68
Figura 40: Relação de Idade e Gênero com as demais variáveis.....	68
Figura 41: Grupo de variáveis de categorias .....	68
Figura 42: Rede Bayesiana Completa .....	69
Figura 43: TPC da variável Moda_Unisex .....	69
Figura 44: TPC da variável Conforto com os valores .....	70
Figura 45: Rede da Seção 3.2 Compilada .....	72
Figura 46: Alterando a probabilidade do Estado ClasseAB .....	73

## LISTA DE TABELAS

Tabela 1: Distribuição Conjunta de $X$ e $Y$ (MARQUES; DUTRA, 2000) .....	23
Tabela 2: Distribuição Conjunta total para as variáveis <i>dordedente</i> , <i>cárie</i> e <i>boticão</i> (RUSSEL; NORVIG, 1995).....	24
Tabela 3: Contribuição de pesquisas para trabalho .....	43
Tabela 4: Tabela de fontes.....	51
Tabela 5: Lojas por Categorias .....	55
Tabela 6: Cenários aplicados na Rede Bayesiana .....	73
Tabela 7: Probabilidades do cenário 1 .....	74
Tabela 8: Probabilidades do cenário 1 com gênero feminino .....	76
Tabela 9: Alteração de Poder de Compra para ClasseAB no cenário 1 .....	78
Tabela 10: Probabilidades do cenário 2.....	79



## SUMÁRIO

<b>1</b>	<b>REFERENCIAL TEÓRICO.....</b>	<b>15</b>
1.1	Teoria da Probabilidade .....	15
1.1.1	Raciocínio probabilístico.....	16
1.1.2	Axiomas da Probabilidade.....	18
1.1.3	Probabilidade Incondicional.....	19
1.1.4	Probabilidade Condicional .....	21
1.1.5	Distribuição Conjunta.....	23
1.1.6	Independência Condicional .....	25
1.1.7	Teorema de Bayes.....	26
1.2	Redes Bayesianas.....	29
1.2.1	Conceitos e terminologia .....	29
1.2.2	Modelos gráficos direcionados: D-Separação .....	33
1.2.3	Inferência em Redes Bayesianas .....	36
1.3	Trabalhos Relacionados.....	40
<b>2</b>	<b>MATERIAIS E MÉTODOS .....</b>	<b>46</b>
2.1.	Local e Período de Realização da Pesquisa.....	46
2.2.	Materiais .....	46
2.2.1.	<i>UnBBayes</i> .....	46
2.2.2.	<i>Microsoft SQL Server</i> .....	47
2.2.3.	<i>Microsoft .NET Framework</i> .....	48
2.3.	Método de Trabalho.....	48
2.3.1.	Pesquisa sobre comportamento do consumidor em shoppings .....	48
2.3.2.	Coleta de Dados .....	49
2.3.3.	Definição dos estados possíveis de cada variável.....	50
2.3.4.	Construir a estrutura da rede, conectando as variáveis para formar o grafo	50
2.3.5.	Definir a tabela de probabilidade condicional para cada nó da rede.....	50
2.3.6.	Aplicação de Cenários com base nos dados coletados.....	51
<b>3</b>	<b>RESULTADOS E DISCUSSÃO .....</b>	<b>53</b>
3.1.	Variáveis.....	53
3.1.1.	Comodidade .....	53
3.1.2.	Lojas.....	54

3.1.3.	Categorias.....	55
3.1.4.	Gênero .....	55
3.1.5.	Poder de Compra.....	55
3.1.6.	Atratividade .....	56
3.1.7.	Marca.....	56
3.1.8.	Idade .....	56
3.2.	Pesquisa e tratamento dos dados .....	57
3.3.	Construção da Rede Bayesiana com UnbBayes .....	61
3.4.	Aplicação de cenários.....	71
3.4.1.	Cenário 1 .....	74
3.4.2.	Cenário 2 .....	79
5.	CONSIDERAÇÕES FINAIS.....	82
6.	REFERÊNCIAS BIBLIOGRÁFICAS .....	85

## INTRODUÇÃO

Segundo dados da Associação Brasileira de *Franchising* em relatório divulgado em 2014, o faturamento dos *Shopping Centers* vai crescer a uma taxa de 15% ao ano (NEGÓCIOS., 2014). Dados da Associação Brasileira de *Shopping Centers* (ABRASCE) informam que 415 milhões de pessoas trafegam mensalmente pelos *shoppings*. Devido à grande popularidade dos *shoppings* mediante o fluxo de clientes dentro deles, entender o comportamento dos consumidores durante o tempo em que estão dentro do *shopping* pode tornar os lojistas mais eficientes na captação de clientes.

Uma forma de analisar esta interação entre clientes e lojistas é por meio de inferência estatística. Na inferência estatística o objetivo é identificar o quão provável alguma afirmação pode ser, por exemplo, se um cliente está interessado em uma categoria de produto. Para identificar a probabilidade é necessário um conjunto de valores representativo, que é chamado de amostra do universo, que no escopo desta pesquisa trata-se das características da interação dos clientes com as lojas, que são transformadas em um conjunto amostral do universo que representa a relação entre clientes e lojas.

Uma das técnicas que podem ser empregadas no processo de inferência estatística são as Redes Bayesianas. Segundo Marques (2000, p.7), Redes Bayesianas “constituem um modelo gráfico que representa de forma simples as relações de causalidade das variáveis de um sistema”. Dessa forma as Redes Bayesianas trabalham realizando relações entre diversas variáveis de um sistema de forma a calcular probabilidades condicionais ou incondicionais para cada uma delas, sem que seja necessário o conhecimento prévio de todos os valores de cada variável, ou sem um grande conjunto de informações.

As características de uma Rede Bayesiana envolvem a associação de um conjunto de variáveis entre si, criando relações de causalidade entre elas e permitindo o cálculo de probabilidades de acordo com o grau de confiança que o sistema tem em relação aos valores de entrada para as variáveis e seus relacionamentos (RUSSEL; NORVIG, 1995).

Trabalhos como os de Yan (2010) e Yim (2010) classificam o comportamento do cliente em uma série de variáveis como sexo, idade, trajetos, tempo de permanência na loja e tempo de permanência no *shopping*. Estas variáveis também foram utilizadas em trabalhos que buscaram estudar a influência sobre o fluxo de pessoas em lojas de marcas famosas em relação a outras menos conhecidas, mas que estavam próximas (SANDALL, 2007). Um modelo hipotético é apresentado na Figura 1.

Figura 1: Cenário de aplicação da pesquisa



Fonte: <http://www.gazetabragantina.com.br/> editada pelo autor

Na Figura 1 é apresentado o contexto em que se aplica esta pesquisa, o modelo de inferência é constituído por variáveis que representam características do ambiente como a loja, a categoria da loja e também características do cliente como idade, gênero, poder de compra, por meio destas variáveis e seus valores calcular a probabilidade de interesse deste indivíduo por uma categoria.

O presente trabalho busca desenvolver um modelo de inferência para identificar o quão interessado o cliente pode estar em uma categoria de produto, e utiliza como amostragem as características do cliente analisado, e em como elas afetam as lojas que ele pode ou não visitar. Deste modo conforme a probabilidade de um indivíduo visitar uma loja aumenta, pode-se inferir a categoria que ele esteja disposto a consumir. Não está no escopo deste trabalho a validação do modelo.

A seção 1 apresenta o referencial teórico do trabalho, dividido nas seções 1.1 sobre teoria da probabilidade, 1.2 Redes Bayesianas e trabalhos relacionados na seção 1.3. A seção 2 apresenta as ferramentas utilizadas e a metodologia de desenvolvimento do trabalho. A seção 3 apresenta os resultados obtidos na pesquisa, e a seção seguinte se dedica as considerações finais do projeto.



# 1 REFERENCIAL TEÓRICO

Este capítulo tem como objetivo investigar os diversos temas relevantes para o desenvolvimento do trabalho, que são as estruturas, definições e conceitos que regem o estudo das Redes Bayesianas, e a análise de trabalhos publicados que possuem como foco a análise de comportamentos e características de clientes em *shoppings centers*.

A seção 1.1 apresenta uma revisão de literatura sobre os fundamentos da Teoria da Probabilidade e as regras que fundamentam o Teorema de Bayes e que constituem os fundamentos necessários para a construção do raciocínio sobre Redes Bayesianas. A seção 1.2 apresenta o referencial com os conceitos principais de Redes Bayesianas. A seção também demonstra alguns algoritmos de inferência, como o algoritmo de Propagação da Árvore de *Cliques*, e, por fim, é apresentado um conjunto de ferramentas utilizadas para modelar soluções de problemas usando as Redes Bayesianas.

A seção 1.3 reúne alguns trabalhos acadêmicos que tiveram como objetivo explorar a análise de características e comportamentos de clientes em *shoppings centers*, buscando atribuir valores ao processo de compra. Em seguida os clientes são avaliados de acordo com uma série de critérios (variáveis), que extraem informações da relação dos clientes com as lojas e servem como influência para o desenvolvimento da presente pesquisa.

## 1.1 Teoria da Probabilidade

A Probabilidade é o ramo da Matemática que pretende modelar fenômenos não determinísticos (VIALI, 2009). Isto significa que a Probabilidade pode ser usada para resolver problemas em que os fatores envolvidos estão relacionados com o acaso. Pode-se exemplificar este conceito utilizando uma partida de futebol, em que um torcedor esteja interessado em saber qual a probabilidade de seu time iniciar a partida jogando, dado o resultado do lançamento de uma moeda. Não se sabe previamente qual time irá começar o jogo, pois o evento está associado ao resultado de cara ou coroa na moeda que será lançada pelo juiz. Desta forma, a Teoria da Probabilidade será apresentada nas próximas seções com o fim de solidificar o raciocínio probabilístico, para fornecer a base necessária para a exploração das Redes Bayesianas. O estudo da probabilidade neste trabalho conclui-se com a demonstração do Teorema de Bayes e, em seguida, são explorados os conceitos relacionados a Redes Bayesianas.

### 1.1.1 Raciocínio probabilístico

Os primeiros trabalhos envolvendo probabilidade surgiram da popularidade dos jogos de azar, que provocaram os filósofos matemáticos Blaise Pascal e Pierre de Fermat, a desenvolverem os primeiros problemas probabilísticos (GRINSTEAD; SNELL, [s.d.]).

A Teoria da Probabilidade clássica afirma que é possível definir a probabilidade de um evento  $A$  ocorrer dividindo a quantidade de elementos contidos em  $A$  pelo número de eventos possíveis no espaço amostral  $S$ , sendo que  $S$  representa todos os eventos possíveis do domínio (LUNA, 2004). Essa abordagem é apresentada na seguinte forma:

$$P(A) = \frac{n_A}{n_S} \quad (1)$$

Onde:

- $P(A)$ : Probabilidade do evento  $A$  ocorrer;
- $n_A$ : Número de elementos no evento  $A$ ; e
- $n_S$ : Numero de elementos no espaço amostral  $S$ .

A Equação 1 pode ser exemplificada utilizando-se a seguinte situação: no lançamento de um dado, quais as chances do número da face que fica para cima ser par? Pode-se organizar (ou modelar) o problema utilizando o seguinte entendimento:

- $S$ : Conjunto de valores possíveis (número de cada face): {1, 2, 3, 4, 5, 6}
- $A$ : Conjunto de valores favoráveis (faces de números pares): {2, 4, 6}

Usando a Equação 1, tem-se:

$$P(A) = \frac{n_A}{n_S} = \frac{3}{6} = 0,5 = 50\%$$

Logo, no lançamento de um dado, há 50% de chances de que o número da face que fica para cima ser par.

Apesar de satisfazer algumas classes de problemas mais simples como o apresentado acima, em que é possível obter ou enumerar todo o espaço amostral ( $S$ ), esta abordagem apresenta algumas limitações quando aplicada em cenários maiores, como a exigência de contagem de todos os elementos pertencentes aos grupos de eventos existentes no espaço amostral  $S$  e também contidos em  $A$ , ou seja, há casos em que não é possível uma abordagem

qualitativa (LUNA, 2004). Neste sentido, a Teoria da Probabilidade clássica possui características que podem ser limitadores em alguns cenários (LUNA, 2004), tais como:

- Dificuldade em enumerar os casos favoráveis e possíveis em experiências onde os resultados precisem de abordagem qualitativa, ou não se é possível efetuar contagens;
- Não depende de experiência para que se afirme o que é idealizado, por exemplo: para o caso “Qual a probabilidade de chover amanhã?” há: eventos possíveis: 2; eventos favoráveis: 1. Assim, aplicando a Equação 1, a chance é de 50% (de chover ou de não chover), o que não se pode afirmar; e
- Espaço amostral com conjunto infinito de elementos não possibilita a inferência de  $n_S$ .

Mediante estas características da Teoria da Probabilidade clássica, ela pode não se aplicar quando se deseja resolver problemas envolvendo incertezas probabilísticas, mas sim, quando há ambientes onde se tem o comportamento ideal em condições ideais (LUNA, 2004), como, por exemplo, o lançamento de um dado ou uma moeda sem a manipulação de pessoas ou do ambiente. Concordando com estas limitações, as restrições envolvendo a teoria clássica da probabilidade podem ser resumidas em (GNERI, 2014):

- Todos os casos possíveis devem ter a mesma probabilidade; e
- Deve existir um número finito de casos possíveis.

Assim, o grande problema na Teoria da Probabilidade clássica vem de que nem sempre é possível determinar a probabilidade de cada um dos eventos do conjunto avaliado (SÁ, 2012), por exemplo, não é possível calcular a probabilidade de um cliente entrar em uma loja apenas dividindo o número de clientes que entraram pela quantidade total de clientes que passaram na porta.

A incerteza probabilística pode ser demonstrada pelo seguinte exemplo, extraído de (RUSSEL; NORVIG, 1995), ao considerar um sistema de diagnóstico em que é aplicável a situação: um paciente chega ao consultório com dor no dente. Pode-se adotar a regra que:

$$\forall p \text{ Sintoma}(p, \text{DorDente}) \rightarrow \text{Doenca}(p, \text{Carie})$$

Em outras palavras, o formalismo indica que se o paciente apresenta sintoma de dor de dente, então a doença é cárie. Entretanto, a regra é incorreta, uma vez que não é possível afirmar que todos os pacientes com sintoma de dor no dente tenham cárie, pois em alguns casos a dor pode



vir de impacto contra o dente, doenças na gengiva ou uma série de outros problemas. A regra, então, poderia ser remodelada:

$$\forall p \text{ Sintoma}(p, \text{DorDente}) \rightarrow \text{Doenca}(p, \text{Carie}) \vee \text{Doenca}(p, \text{Gengiva}) \vee \text{Doenca}(p, \text{Impacto})$$

Neste caso, o formalismo indica que se o sintoma do paciente é dor de dente, a doença pode ser uma entre estas: cárie, problemas na gengiva ou impacto contra o dente. Assim, seria possível afirmar que a representação está mais completa, pois está considerando mais fatores. Entretanto, para fazer com que a regra esteja mais próxima de uma verdade, foi necessário adicionar uma lista de causas possíveis. O que gera o entendimento de que, para diminuir a incerteza, seria necessário enumerar mais doenças causadas pelo sintoma “dor de dente”.

Ainda, é possível tornar o formalismo em uma regra causal, usando a seguinte forma:

$$\forall p \text{ Doenca}(p, \text{Carie}) \rightarrow \text{Sintoma}(p, \text{DorDente})$$

Assim, a regra representa que, se houver cárie, o sintoma deverá ser dor de dente. Entretanto, a representação do diagnóstico ainda continua falha, uma vez que nem toda cárie causa dor. O exaustivo trabalho para se identificar todas as variáveis do problema, a falta de embasamento concreto sobre as variáveis, e a impossibilidade de se realizar todos os testes para identificar todos os cenários possíveis, descrevem o que é a incerteza (RUSSEL; NORVIG, 1995).

As próximas subseções continuam a explorar os conceitos que constituem o raciocínio probabilístico.

### 1.1.2 Axiomas da Probabilidade

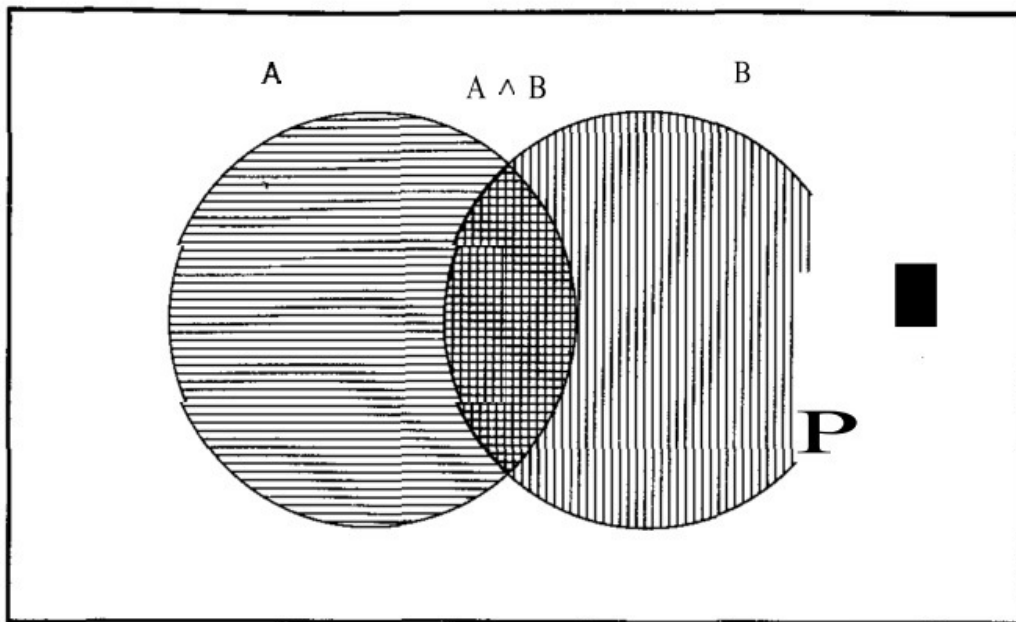
Existem três regras que definem as atribuições que um sistema responsável pelos cálculos de probabilidade pode tomar para quaisquer proposições (RUSSEL; NORVIG, 1995), e para isso são definidos os seguintes axiomas:

- Todas as probabilidades estão entre 0 e 1;
- São verdades todas as proposições de probabilidade 1, e falsas todas as de probabilidade 0, bem como a soma de todas as variáveis para cada um de seus estados resulta em 1; e
- A probabilidade de uma disjunção é dada por:  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

As duas primeiras regras estabelecem a escala da probabilidade, enquanto a terceira é uma referência ao diagrama de Venn (RUSSEL; NORVIG, 1995), que pode ser visualizado na Figura 2.

Figura 2: Diagrama de Venn para representação da Proposição A união com B, e A intersecção com B

**Verdade**



Fonte: (RUSSEL; NORVIG, 1995)

A Figura 2 mostra a intersecção entre dois conjuntos de dados A e B, onde a disjunção é dada pela soma dos conjuntos menos a sua intersecção, que caso não fosse somada, iria resultar em números duplicados para o cálculo da probabilidade. A probabilidade de uma disjunção, apresentada na terceira regra, também é conhecida por **princípio de inclusão-exclusão** (RUSSEL; NORVIG, 1995). Em linhas gerais, o princípio da inclusão-exclusão diz que uma vez observados os casos em que A é válido, somados aos casos em que B é válido, consegue-se obter todos os casos em que  $(A \cup B)$  são válidos. Entretanto, somar os conjuntos A e B faz com que sua intersecção seja contada duas vezes, por isso a necessidade de se subtrair  $(A \cap B)$ .

Estes axiomas estarão presentes em todas as probabilidades consideradas nas próximas seções.

### 1.1.3 Probabilidade Incondicional

Probabilidade incondicional é aquela existente antes de qualquer evidência, e também é chamada de probabilidade *a priori*. Luger (2013, p.149) define probabilidade *a priori* da seguinte forma:

A probabilidade *a priori*, geralmente uma probabilidade incondicional, de um evento é a probabilidade atribuída com base em todo o conhecimento que dá suporte a sua ocorrência ou ausência, ou seja, a probabilidade do evento antes de qualquer evidência nova. A probabilidade *a priori* de um evento é simbolizada por:  $P(\text{evento})$ .

Dessa forma se entende como *a priori* o grau de confiança que é dado a uma hipótese na ausência de quaisquer outras informações (RUSSEL; NORVIG, 1995). Por exemplo, diga-se que a probabilidade de um paciente ter “Cárie” é 10% (0,1), sua representação (na forma de proposição) é:

$$P(\text{Carie}) = 0,1$$

Outra forma de representar esta notação é usar:

$$P(\text{Variável} = \text{Valor}) = \text{Número}$$

Nos exemplos a seguir a variável é “Clima” e os valores possíveis são “Calor”, “Chuvoso”, “Nublado” e “Neve”:

$$\begin{aligned} P(\text{Clima} = \text{Calor}) &= 0,7 \\ P(\text{Clima} = \text{Chuvoso}) &= 0,2 \\ P(\text{Clima} = \text{Nublado}) &= 0,08 \\ P(\text{Clima} = \text{Neve}) &= 0,02 \end{aligned}$$

As proposições que possuem igualdades são conhecidas como variáveis aleatórias, assim como os valores também podem ser expressas como *booleanos* (sim/não). Os valores definidos para as proposições podem ser compreendidos como distribuição das probabilidades, quando um conjunto de valores é atribuído a uma proposição e denota todos os valores adotados por ela (RUSSEL; NORVIG, 1995). Dessa forma, o exemplo anterior também pode ser expresso da seguinte forma:

$$P(\text{Clima}) = \{0,7, 0,2, 0,08, 0,02\}$$

É possível adicionar operadores lógicos para construir sentenças mais complexas e também expressar as variáveis levadas em consideração pela distribuição, como por exemplo:

$$P(\text{Carie} \wedge \neg \text{DorDente}) = 0,06$$

Esta sentença informa que a probabilidade de um paciente ter cárie, mas não ter dor de dente, é de 6%. É possível construir uma expressão como  $P(\text{Clima}, \text{Carie})$  e extrair a

probabilidade para todas as combinações de uma lista de variáveis aleatórias. Quando uma nova evidência é encontrada, a probabilidade incondicional não é mais útil, e nesse momento passa a se considerar a probabilidade *a posteriori*, ou probabilidade condicional.

#### 1.1.4 Probabilidade Condicional

Probabilidade Condicional, ou probabilidade *a posteriori*, é a probabilidade da ocorrência de um evento dada uma nova informação ou restrição sobre esse evento (LUGER, 2013). Para determinar a probabilidade Condicional, Luger (2013, p.149) propõe a seguinte definição:

A probabilidade *a posteriori* (após o fato), geralmente de uma probabilidade condicional de um evento, é a probabilidade de um evento dada uma evidência nova. A probabilidade *a posteriori* de um evento dada uma evidência é simbolizada por:  $p(\text{evento} | \text{evidência})$ .

Por exemplo, caso seja expressar a probabilidade de fazer calor, dado a observação da estação do ano, e considerando que a estação seja Verão, a notação atribuindo-se o evento na relação do clima e estação do ano pode ser expresso da seguinte forma:

$$P(\text{Calor} | \text{estação} = \text{Verão}) = 0,8$$

A expressão é lida da seguinte forma: a probabilidade de haver calor, dada a evidência de que estamos no verão, é 0,8, e a evidência previamente conhecida é a estação do ano. Ou seja, a notação  $P(A|B)$  significa a probabilidade de o evento  $A$  ocorrer, dado que tudo que se conhece é  $B$  (a evidência).

A probabilidade condicional é calculada através do conhecimento das probabilidades incondicionais em relação às variáveis que participam do evento, ou seja, são assim definidas em termos de probabilidade incondicional (RUSSEL; NORVIG, 1995).

A equação é dada por:

$$P(A | B) = \frac{P(A \wedge B)}{P(B)} \quad [1]$$

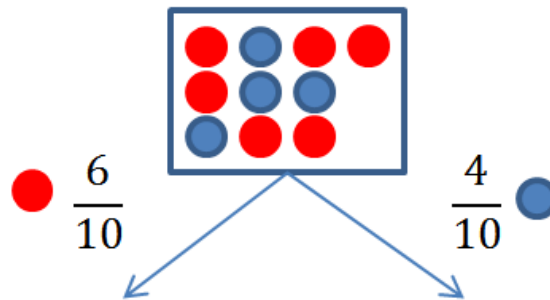
Onde:

- $P(a | b)$ : Probabilidade de  $b$  considerando que se conhece o valor de  $a$ ;
- $P(a \wedge b)$ : Probabilidade incondicional para a verdade de que  $a$  e  $b$  sejam verdadeiros; e
- $P(b)$ : Probabilidade de  $b$ .

Para racionalizar sobre a probabilidade condicional considere o seguinte exemplo: em uma caixa contendo dez bolas, onde seis são vermelhas e quatro são azuis, qual a probabilidade

de que alguém retire uma bola vermelha na segunda tentativa, sendo que na primeira vez ele retirou uma bolinha azul? A probabilidade condicional, como informado anteriormente, é a probabilidade de um evento B, dado o acontecimento de um evento A. Então, primeiro deve-se verificar a probabilidade do evento de retirar uma bola azul na primeira tentativa. A Figura 3 apresenta as probabilidades para os possíveis eventos A.

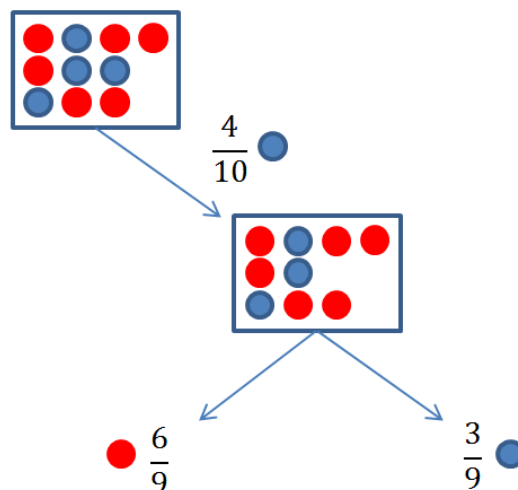
Figura 3: Probabilidades para a retirada de uma bola azul ou vermelha na primeira tentativa



Fonte: Elaborada pelo autor

A probabilidade de que na primeira tentativa seja retirada uma bola vermelha é de 0,6, enquanto que a probabilidade de se retirar uma bola azul é de 0,4. Considerando que a primeira bola retirada foi a azul, pode-se verificar a probabilidade de que a segunda bola retirada seja vermelha. A Figura 4 exibe o próximo evento de retirada da segunda bola.

Figura 4: Evento de retirada da segunda bola



Fonte: Elaborada pelo autor

O fato de ter retirado no primeiro momento a bola azul, alterou as chances de que seja retirada a bola vermelha no segundo momento. A probabilidade de retirar uma bola vermelha no segundo momento é de 0,66 enquanto que a probabilidade de retirar uma segunda bola azul é de 0,33.

Segundo a notação da probabilidade condicional apresentada anteriormente, pode-se expressar o cálculo anterior como:

$$P(A) = \frac{4}{10}$$

$$P(B | A) = \frac{6}{9}$$

Uma vez com os valores de  $P(A)$  e  $P(B | A)$  definidos, é possível calcular então a probabilidade do evento A e B acontecerem, para isso basta reescrever a equação da probabilidade condicional [1] na forma da regra do produto ficando na seguinte forma:

$$P(A \wedge B) = P(A) \times P(B | A)$$

$$\frac{4}{10} \times \frac{6}{9} = \frac{24}{90} = 0,2$$

Isso significa que a probabilidade do evento A e B acontecerem é de 0,2. A regra do produto é a reescrita da regra da probabilidade condicional, aplicada para se calcular a probabilidade da ocorrência de dois eventos A e B. Ela pode assumir duas formas:

$$P(A \wedge B) = P(A) \times P(B | A)$$

E ainda;

$$P(A \wedge B) = P(B) \times P(A | B)$$

A próxima seção demonstra como é realizado o processo de inferência em probabilidade, através da distribuição da probabilidade conjunta.

### 1.1.5 Distribuição Conjunta

A Distribuição Conjunta, ou Distribuição da Probabilidade Conjunta, é dada por uma variável  $X$  e o conjunto de todos os seus estados possíveis, representado por  $(X_1, \dots, X_m)$ , em relação às demais variáveis do sistema, por exemplo, uma segunda variável  $Y$  e seu respectivo conjunto de estados  $(Y_1, \dots, Y_n)$ . Dessa forma, a Distribuição Conjunta representa a probabilidade das variáveis  $X$  e  $Y$  ocorrerem no mesmo tempo em cada um de seus estados, resultando em uma tabela de  $m \times n$  probabilidades (MARQUES; DUTRA, 2000). Por exemplo, dados os estados de  $X (a_1, a_2)$  e  $Y (b_1, b_2, b_3)$  a tabela resultante possui: 2 linhas, 3 colunas e  $2 \times 3$  elementos (cruzamentos linha-coluna, ou seja, entre as variáveis  $X$  e  $Y$ ), e é representada pela Tabela 1.

Tabela 1: Distribuição Conjunta de  $X$  e  $Y$  (MARQUES; DUTRA, 2000)

	$b_1$	$b_2$	$b_3$
$a_1$	0,2	0,6	0,4
$a_2$	0,8	0,4	0,6

Dessa forma, a Tabela 1 indica que a chance de ocorrerem os eventos  $a_1$  e  $b_2$  é de 60%. Assim, a Distribuição Conjunta pode ser classificada como a computação de probabilidades posteriores a partir de uma evidência observada (RUSSEL; NORVIG, 1995). Ou seja, para que o evento  $P(a_1 \wedge b_2)$  aconteça, deve ser observada a presença condicional tanto de  $a_1$  quanto de  $b_2$ .

A Distribuição Conjunta pode ser observado no seguinte exemplo (RUSSEL; NORVIG, 1995): em um processo de diagnóstico dentário no qual, para um determinado paciente, deseja-se construir a tabela de distribuição conjunta completa para o relacionamento entre as variáveis “cárie” e “dordedente”, onde a dor tem relação ainda com uma terceira variável chamada “boticão”. Para construir a tabela de Distribuição Conjunta é necessário incluir também os casos de exclusão de cada uma das variáveis, ou seja, como as variáveis são *booleanas*, seus valores possíveis representam “sim” ou “não”.

Tabela 2: Distribuição Conjunta total para as variáveis *dordedente*, *cárie* e *boticão* (RUSSEL; NORVIG, 1995)

	<i>dordedente</i>		$\neg$ <i>dordedente</i>	
	<i>boticão</i>	$\neg$ <i>boticão</i>	<i>boticão</i>	$\neg$ <i>boticão</i>
<i>cárie</i>	0,108	0,012	0,072	0,008
$\neg$ <i>cárie</i>	0,016	0,064	0,144	0,576

A distribuição conjunta total apresenta  $2 \times 2 \times 2$  valores para a probabilidade incondicional de cada estado, além de que pode ser observado o respeito ao axioma de que o somatório de todas as distribuições conjuntas resulta em 1. Analisando a tabela, é possível, por exemplo, encontrar a probabilidade incondicional de:

$$P(\textit{cárie}) = 0,108 + 0,012 + 0,072 + 0,008 = 0,2;$$

$$P(\neg\textit{cárie}) = 0,016 + 0,064 + 0,144 + 0,576 = 0,8; \text{ e ainda}$$

$$P(\textit{cárie} \wedge \textit{dordedente}) = 0,108 + 0,012 = 0,12.$$

É possível calcular a probabilidade condicional em razão da distribuição conjunta, utilizando-se para isso primeiro a regra do produto, e em seguida avaliando a expressão a partir da distribuição conjunta total (RUSSEL; NORVIG, 1995). Por exemplo, o cálculo da probabilidade condicional  $P(\textit{cárie} | \textit{dordedente})$  de acordo com o exemplo anterior, encontrado em (RUSSEL; NORVIG, 1995), é desenvolvido da seguinte forma:

$$P(\textit{boticao} | \textit{carie}) = \frac{P(\textit{boticao} \wedge \textit{carie})}{P(\textit{carie})}$$

$$P(\textit{boticao} | \textit{carie}) = \frac{0,108 + 0,072}{0,108 + 0,012 + 0,072 + 0,008} = 0,9$$

Como o valor de  $P(\text{boticao} | \text{carie}) = 0,9$ , pode-se dizer então que  $P(\neg\text{boticao} | \text{carie}) = 0,1$ , uma vez o valor da soma da distribuição conjunta deve ser 1,0. O cálculo a seguir demonstra a afirmação de que  $P(\neg\text{boticao} | \text{carie}) = 0,1$ .

$$P(\neg\text{boticao} | \text{carie}) = \frac{P(\neg\text{boticao} \wedge \text{carie})}{P(\text{carie})}$$

$$P(\neg\text{boticao} | \text{carie}) = \frac{0,012 + 0,008}{0,108 + 0,012 + 0,072 + 0,008} = 0,1$$

Analisando o desenvolvimento das duas probabilidades condicionais, é possível observar um padrão na fórmula, que pode ser dado por:

$$\frac{1}{P(\text{carie})} = \frac{1}{0,108 + 0,012 + 0,072 + 0,008} = 0,5$$

Ou seja, o valor de  $\frac{1}{P(\text{carie})}$  atua como constante de normalização (RUSSEL; NORVIG, 1995), assegurando que a soma de ambas as distribuições seja sempre 1. Dessa forma é possível então calcular o valor para *boticao* e  $\neg\text{boticao}$  sem ter conhecimento do valor de  $P(\text{carie})$ , utilizando para isso a soma dos valores de em que *boticao* e  $\neg\text{boticao}$  são válidos, para em seguida dividir cada um dos dois resultados pela soma de ambos (*boticao* e  $\neg\text{boticao}$ ). O procedimento é exemplificado da seguinte forma:

$$\text{boticao} = 0,108 + 0,072 = 0,18$$

$$\neg\text{boticao} = 0,012 + 0,008 = 0,02$$

$$P(\text{boticao} | \text{carie}) = \frac{0,18}{0,18 + 0,02} = \frac{0,18}{0,2} = 0,9$$

$$P(\neg\text{boticao} | \text{carie}) = \frac{0,02}{0,18 + 0,02} = \frac{0,02}{0,2} = 0,1$$

Trabalhar com distribuições conjuntas requer um custo muito alto, ao passo em que a construção das tabelas de conjunção cresce em escala, fazendo com que não seja uma ferramenta prática para a construção de sistemas de raciocínio (RUSSEL; NORVIG, 1995). A seção 1.1.7 apresentará o Teorema de Bayes o qual, mediante o desenvolvimento das equações da regra de produto visto até agora, é considerado a base dos sistemas de inferência probabilísticos modernos (RUSSEL; NORVIG, 1995).

### 1.1.6 Independência Condicional

Construir tabelas de distribuição conjunta total pode se tornar uma atividade muito complexa dependendo da quantidade de variáveis de um sistema (RUSSEL; NORVIG, 1995). Como a distribuição calcula probabilidades entre todas as variáveis do sistema, podem existir relações que talvez não interfiram umas nas outras, em razão de variáveis que não possuem



relação condicional serem empregadas para calcular uma probabilidade. Isso é dado, segundo Luger (2013, p.151), pela definição de eventos independentes:

Dois eventos A e B são independentes um do outro se, e somente se,  $P(A \cap B) = P(A)P(B)$ . Quando  $P(B) \neq 0$ , isso é o mesmo que dizer que  $P(A) = P(A|B)$ . Ou seja, saber que B é verdadeiro não afeta a probabilidade de A ser verdadeiro.

Ao examinar a proposição  $P(A|B)$  é incorreto afirmar que se B for verdade, então  $P(A)$  é 0,8, pois  $P(A)$  informa apenas a probabilidade incondicional de A, e não seu valor depois de dada alguma evidência. Dada a definição anterior de eventos independentes, para cada nova informação C adicionada, é necessária calcular  $P(A|B \cap C)$ .

### 1.1.7 Teorema de Bayes

Como visto nas seções 1.1.4 e 1.1.5, a regra do produto é dada por:

$$P(A \wedge B) = P(A|B)P(B)$$

Que também pode ser escrita na forma:

$$P(A \wedge B) = P(B|A)P(A)$$

Sendo assim, podem-se igualar as duas expressões e desenvolver a equação da seguinte forma:

$$\begin{aligned} P(B|A)P(A) &= P(A|B)P(B) \\ P(B|A) &= \frac{P(A|B)P(B)}{P(A)} \end{aligned}$$

O desenvolvimento da regra do produto em relação as suas equivalências forma o que é conhecido por teorema de Bayes. A regra de Bayes permite calcular uma probabilidade desconhecida a partir de uma evidência conhecida, e para isso são necessárias três variáveis, sendo uma probabilidade condicional e duas probabilidades incondicionais (RUSSEL; NORVIG, 1995).

O teorema de Bayes na sua forma mais geral é dado por:

$$P(Y|X) = \frac{P(X|Y) \times P(Y)}{P(X)} \quad (3)$$

onde:

- $P(Y)$  e  $P(X)$  são probabilidades incondicionais; e
- $P(X|Y)$  é a probabilidade condicional.

As aplicações que utilizam o teorema de Bayes trabalham fortemente com as relações que existem entre efeito e causa, onde frequentemente busca-se compreender o efeito, para que seja possível calcular a probabilidade da causa (LUGER, 2014). Russel (1995, p435) exemplifica esta relação de causa e efeito no teorema de Bayes da seguinte forma:

$$P(\text{causa} | \text{efeito}) = \frac{P(\text{efeito} | \text{causa}) P(\text{causa})}{P(\text{efeito})}$$

Exemplificando o raciocínio da causa e efeito, pode-se mencionar o diagnóstico médico, na qual é provável saber a probabilidade de  $P(\text{sintomas} | \text{doenças})$ , de modo que necessita calcular a probabilidade de  $P(\text{doenças} | \text{sintomas})$ . Por exemplo, um médico sabe que uma doença como meningite causa torcicolo em 50% dos casos. O médico sabe também que a probabilidade de um paciente ter meningite é 1/50.000 casos, e a probabilidade de um paciente qualquer ter torcicolo é 1/20. Atribuindo a preposição  $S$  para os pacientes que têm torcicolo e  $M$  para os pacientes que têm meningite, obtém-se:

$$\begin{aligned} P(S|M) &= 0,5 \\ P(M) &= \frac{1}{50000} \\ P(S) &= \frac{1}{20} \\ P(M|S) &= \frac{P(S|M) \times P(M)}{P(S)} = \frac{0,5 \times 1/50000}{1/20} = 0,0002 \end{aligned}$$

Isso significa que a cada 5000 pacientes apenas um paciente com torcicolo deve ter meningite (RUSSEL; NORVIG, 1995). Quando se trabalha com um conjunto limitado de dados como no exemplo anterior, ou seja, poucas relações entre evidências e eventos, o Teorema de Bayes se mostra um método seguro e eficiente para cálculo probabilístico.

O teorema de Bayes não está vinculado ao uso de apenas uma evidência para se realizar o cálculo de uma probabilidade. Existem diversos casos em que os cálculos envolvem um conjunto muito maior de evidências. Voltando ao exemplo dado na seção 1.1.5 sobre o diagnóstico dentário, pode-se calcular a probabilidade de Cárie na existência de duas evidências, sendo elas  $P(\text{dordedente})$  e  $P(\text{boticão})$ . Assim, deseja-se calcular  $P(\text{Cárie} | \text{dordedente} \wedge \text{boticão})$ , onde a fórmula para o cálculo está na seguinte forma:

$$P(\text{Cárie} | \text{dordedente} \wedge \text{boticão}) = \frac{P(\text{dordedente} \wedge \text{boticão})P(\text{Cárie})}{P(\text{dordedente} \wedge \text{boticão})}$$

Para desenvolver a probabilidade de  $P(\text{Cárie} | \text{dordedente} \wedge \text{boticão})$  será necessário calcular a probabilidade condicional de  $P(\text{dordedente} \wedge \text{boticão})$  para cada valor de  $P(\text{Cárie})$ , e com os valores obtidos calcular a distribuição conjunta, e por fim substituir no teorema de Bayes. Isso pode ser feito sem muito esforço para  $P(\text{dordedente} \wedge \text{boticão})$  uma vez que irá derivar apenas em função de  $P(\text{Cárie})$ . Mas, o que fazer quando o cálculo da probabilidade envolve um conjunto muito maior de evidências? A escala limita o uso das distribuições conjuntas e do teorema de Bayes, pois as tabelas de distribuição e a quantidade de variáveis utilizadas no teorema podem crescer de forma exponencial dependendo do domínio estudado.

Uma forma de tentar reduzir o esforço para calcular probabilidades a partir de um conjunto extenso de evidências é com o uso da Probabilidade Incondicional visto na seção 1.1.3. Entretanto  $P(\text{dordedente} \wedge \text{boticão})$  não se trata de uma probabilidade Incondicional, pois ambas  $P(\text{dordedente})$  e  $P(\text{boticão})$  estão relacionadas entre si para que exista a probabilidade condicional.

Para o processo de simplificação do cálculo da probabilidade no teorema de Bayes, pode-se então utilizar uma derivação da Probabilidade Incondicional chamada de Independência Condicional (RUSSEL; NORVIG, 1995). A independência condicional procura tornar dois eventos condicionalmente independentes dada a presença de um terceiro evento (LUGER, 2014). Sendo assim,  $P(\text{dordedente})$  e  $P(\text{boticão})$  estão condicionalmente vinculados à existência de  $P(\text{Cárie})$ , tornando-os assim independentes um do outro dadas à presença ou não de  $P(\text{Cárie})$ . A equação adquire a seguinte forma:

$$P(\text{Cárie} | \text{dordedente} \wedge \text{boticão}) = P(\text{dordedente} | \text{Cárie}) P(\text{boticão} | \text{Cárie}) P(\text{Cárie})$$

Assim, pode-se concluir que Cárie separa  $P(\text{dordedente})$  e  $P(\text{boticão})$  porque é uma consequência direta de ambas, e o uso da independência condicional na distribuição da probabilidade como método de simplificação do cálculo é chamado de modelo bayesiano ingênuo (RUSSEL; NORVIG, 1995).

Caso não seja adotado o modelo bayesiano ingênuo, conforme aumenta a complexidade das relações e quando existe um grande conjunto de evidências e outro de eventos, a manutenção do sistema pode requerer um custo de processamento muito alto, além de elevar bastante a complexidade da solução. Se considerar um sistema com 200 doenças e 2.000 sintomas, a quantidade de relações existentes é superior a 800.000.000.

## 1.2 Redes Bayesianas

Na seção 1.1.5 foi apresentado um método de inferência probabilística, onde a partir de uma tabela de probabilidades era possível calcular qualquer conjunto de outras probabilidades derivadas com referência na distribuição conjunta total. O uso da independência e da independência condicional apresentados nas seções 1.1.6 e 1.1.7 permitem tornar menos trabalhoso e complexo o cálculo de probabilidades usando o teorema de Bayes. Entretanto, existem cenários onde as relações entre as variáveis do sistema podem ser expressas no formato de uma rede, e no formato de rede é possível uma representação mais fiel da realidade e demonstra de forma mais explícita as relações entre as variáveis do sistema (RUSSEL; NORVIG, 1995). As próximas seções apresentam os conceitos e características relacionadas a Redes Bayesianas, que possui sua base fundamentada nas seções apresentadas anteriormente.

### 1.2.1 Conceitos e terminologia

Como visto na seção 1.1.7 existe uma dificuldade de escala ao se trabalhar com o Teorema de Bayes em sistemas complexos ou que possuam uma natureza de manutenção (tabelas de conjunção que são atualizadas continuamente). Um exemplo disso são os sistemas de diagnósticos médicos, em que novas descobertas são realizadas continuamente e assim demandam mais capacidade de processamento computacional para se atualizar a tabela de distribuição conjunta total.

Para resolver estes problemas de complexidade na atualização constante de sistemas de natureza dinâmica, ou seja, que mudam com o tempo, são empregados os modelos gráficos probabilísticos, também conhecidos por Redes Bayesianas (outros títulos aceitos são: Redes Bayesianas de Crença, Rede Probabilística, Rede Causal ou Mapa de Conhecimento) (RUSSEL; NORVIG, 1995). As Redes Bayesianas de Crença possibilitam a suposição de que outros eventos sejam condicionalmente independentes ou que a correlação seja tão pequena que seja possível ignorar (LUGER, 2013), assim a definição de Rede Bayesiana de Crença é dada por (RUSSEL; NORVIG, 1995):

- Cada nó está associado a uma variável aleatória.
- Um conjunto de vínculos orientados ou setas conecta pares de nós, e estabelece uma relação de Pai para Filho entre os nós conectados. Se um nó X vai até outro nó Y, entende-se que X é pai de Y.
- Cada nó  $X_i$  tem uma distribuição de probabilidade condicional  $P(X_i | P(\text{Pais}(X_i)))$  que qualifica os pais do nó.

A base das Redes Bayesianas é a união entre a Teoria da Probabilidade e a Teoria dos Grafos, utilizando-se da modularidade oferecida pelo modelo gráfico apoiado pelo tratamento da incerteza oferecido pela Teoria da Probabilidade (2013, JORDAN apud LUGER, 1999, p. 303). Marques (2000, p.7) complementa o significado de Redes Bayesianas mostrando que “constituem um modelo gráfico que representa de forma simples as relações de causalidade das variáveis de um sistema”.

Dessa forma, as Redes Bayesianas tratam conjuntos menores de eventos e evidências – ainda que as mais relevantes – apresentando um modelo computacional que trabalha empregando o Teorema de Bayes para calcular as probabilidades de eventos dado um conjunto de evidências, e estas evidências estão no contexto das relações causais, ou seja, relacionadas entre si e dentro do domínio de um problema (LUGER, 2013).

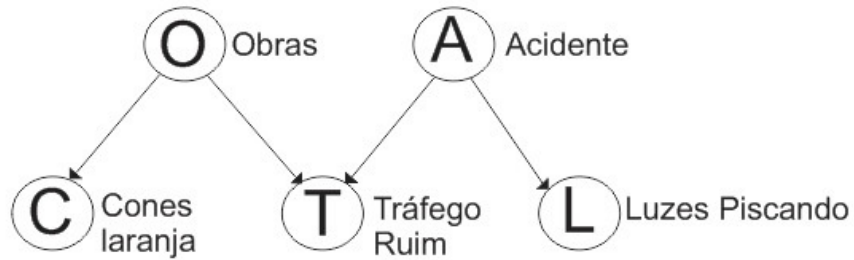
É possível ainda, além do benefício da modularidade dos grafos, que especialistas do domínio do problema (médicos em sistemas de diagnósticos clínicos, meteorologistas em aplicações de controle do clima etc.) possam selecionar empiricamente fenômenos específicos, para obter as probabilidades e influências com o foco apenas neste grupo de eventos. Essa dependência apenas do nó em relação aos seus pais em Redes Bayesianas de Crença é dada pela representação de probabilidades condicionais por meio da influência causal, em que está implícita a suposição de que essas influências são diretas, ou seja, a presença de um evento causa outros eventos na rede (LUGER, 2013).

O seguinte exemplo, desenvolvido por Luger (2013, p.303), demonstra a utilização de uma Rede Bayesiana de Crença: considere a análise do tráfego de uma rodovia, em que o motorista observa a redução da velocidade na pista e a presença de um congestionamento. Qual pode ser a causa da situação? Para modelagem do problema, adota-se a formalização:

- O: obras na estrada;
- A: acidente;
- C: presença de cones laranja;
- T: tráfego ruim; e
- L: luzes piscando.

O grafo utilizado para representar o problema de tráfego é apresentado na Figura 5.

Figura 5: Grafo do problema de tráfego



Fonte: Elaborada pelo autor

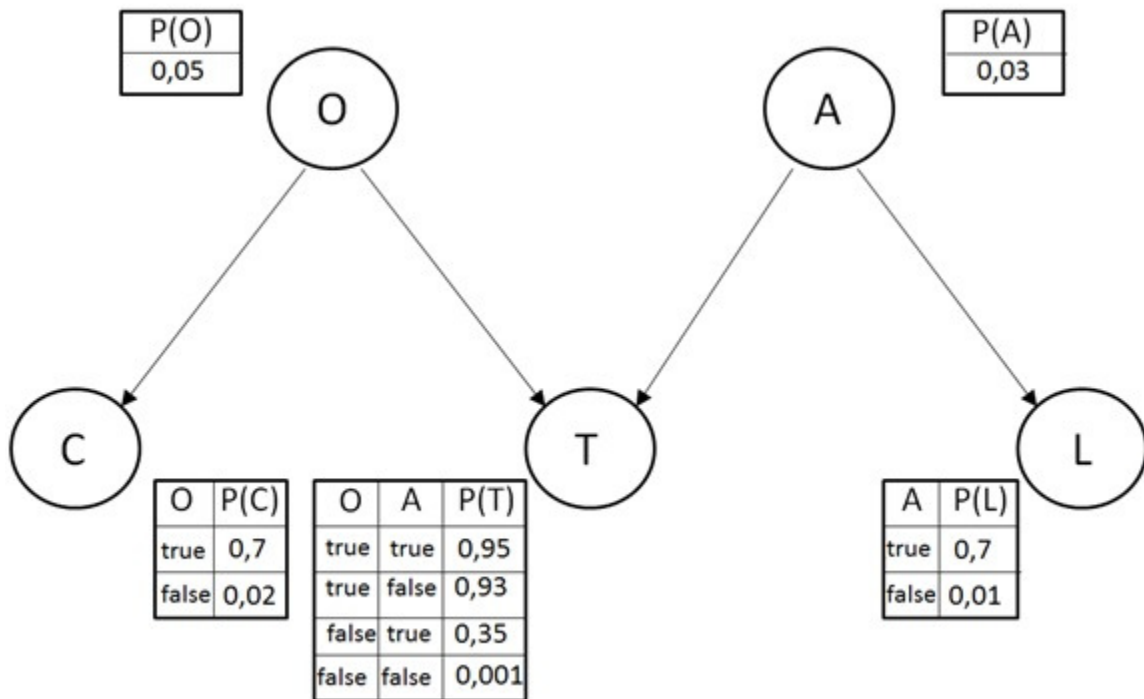
As relações causais no grafo podem ser lidas da seguinte forma: obras não são um efeito causal para Acidente; Cones laranja são causados apenas por obras; e Tráfego ruim é causado por Obras ou Acidentes; Luzes piscando são causadas apenas por Acidentes. Aqui é empregado o mesmo conceito de independência condicional visto na seção 1.1.7.

Utilizando o conceito de Redes Bayesianas é necessário identificar o conjunto de estados possíveis em cada uma das variáveis ( $O$ ,  $A$ ,  $C$ ,  $T$ ,  $L$ ), e uma vez identificados, deve-se desenvolver a Distribuição Conjunta (vista na seção 1.1.5) para identificar a probabilidade das relações causais existentes, que é representada por:

$$P(O, A, C, T, L) = P(O) \times p(A|O) \times P(C|O, A) \times P(T|O, A, C) \times P(L|O, A, C, T)$$

A Distribuição Conjunta é demonstrada da seguinte forma:

Figura 6: Distribuição conjunta em uma Rede Bayesiana



Fonte: Elaborado pelo autor

Os dados da Figura 6 foram definidos no próprio exemplo e não obtidos de outros lugares. A Distribuição Conjunta foi então a atribuição das probabilidades para cada estado possível nas variáveis aleatórias. A quantidade de valores na tabela para cada variável pode ser mensurada pelo número de pais que o nó tenha, ou seja, a quantidade de entradas para cada variável é de  $2^k$ , onde  $k$  representa a quantidade de pais do nó. Como em redes Bayesianas os nós são independentes de todos os descendentes, dado o conhecimento de sua relação causal, pode-se reduzir a expressão para:

$$p(O, A, C, T, L) = p(O) \times p(A) \times p(C|O) \times p(T|O, A) \times p(L|A)$$

Observa-se que o termo  $p(C|O, A)$  foi reduzido para  $p(C|O)$ , pois parte-se do princípio que a existência de cones (C) não é um efeito causal para acidentes (A). Da mesma forma, a presença de cones (C) não é uma causa para tráfego ruim (T), mas obras (O) e acidentes (A), sim, reduzindo o termo  $p(T|O, A, C)$  para  $p(T|O, A)$ . Por fim, a presença de luzes piscando (L) é causada apenas por acidentes (A), o que reduz o termo  $p(L|O, A, C, T)$  para  $p(L|A)$ . Pelo modelo gráfico apresentado na Figura 5 foi possível então visualizar o conjunto de variáveis que são dependentes ou não entre si. Este método de determinar a relação de causa entre as variáveis por meio de um modelo gráfico é conhecido por Separação-D e será abordado na próxima seção.

### 1.2.2 Modelos gráficos direcionados: D-Separação

D-Separação é a propriedade que permite, a partir da presença ou ausência de informações, particionar o modelo de modo a controlar sua complexidade de busca (LUGER, 2013) e determinar a relação de causa entre as variáveis. A definição para D-Separação é dada por:

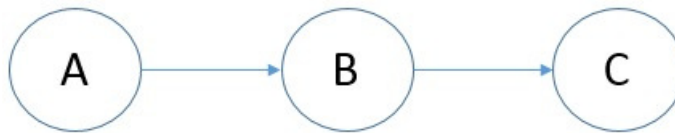
Dois nós (A e B) em um grafo direcionado acíclico são d-separados se todo caminho entre eles estiver bloqueado. Um caminho é qualquer série contínua de conexões no grafo. Um caminho está bloqueado se houver um nó intermediário V no caminho com uma das seguintes propriedades:

A conexão é serial ou divergente e o estado V é conhecido ou

A conexão é convergente e nem V nem outro de seus filhos possuem evidências (LUGER, 2013).

Qualquer cadeia representada forma exemplificada pela Figura 7:

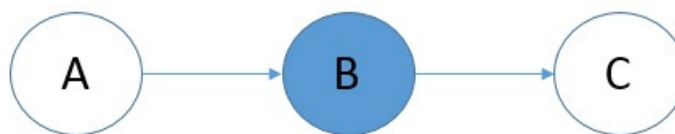
Figura 7: Relacionamento de dependência de A e B.



Fonte: Elaborado pelo autor

Torna a variável A e a variável C dependentes se todas as variáveis são desconhecidas. Se a cadeia apresenta a variável intermediária conhecida, terá uma representação conforme a Figura 8:

Figura 8: Relacionamento de independência entre A e C

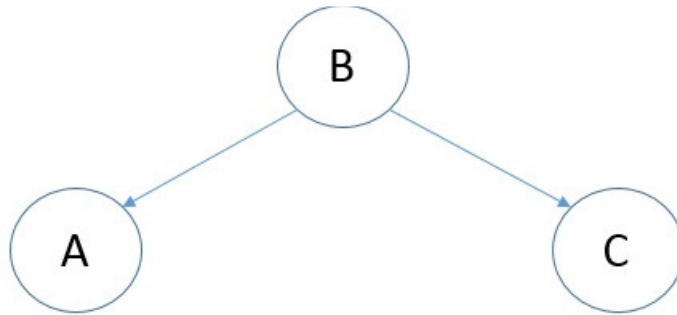


Fonte: Elaborado pelo autor

A variável A e a variável C serão independentes dado o conhecimento de B, como apresentado na Figura 8. Uma cadeia que apresenta a seguinte forma:



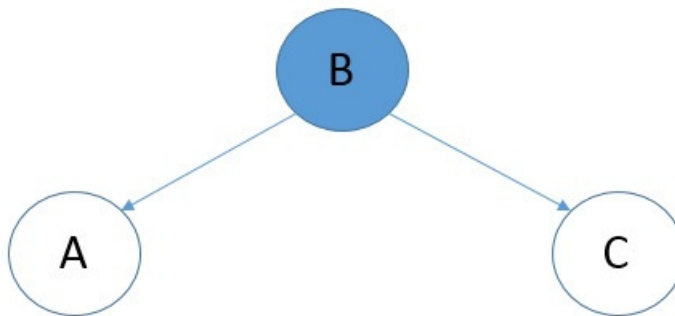
Figura 9: Dependência de A e C caso B seja desconhecida



Fonte: Elaborado pelo autor.

Na figura as variáveis A e C são dependentes uma vez que não se conhece a variável B. Se a variável B for conhecida, então as variáveis A e C serão independentes dado o conhecimento de B, como mostra a Figura 10.

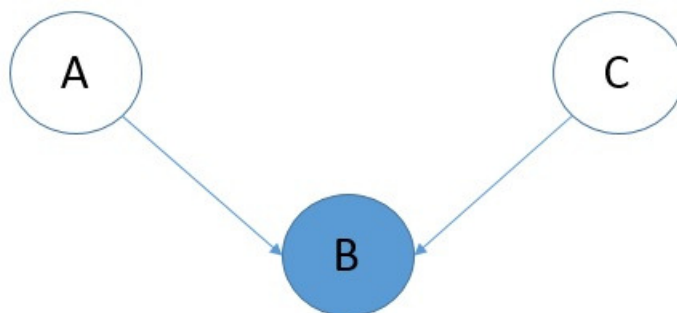
Figura 10: Independência de A e C caso B seja conhecido



Fonte: Elaborado pelo autor

Em uma cadeia distribuída conforme a Figura 11:

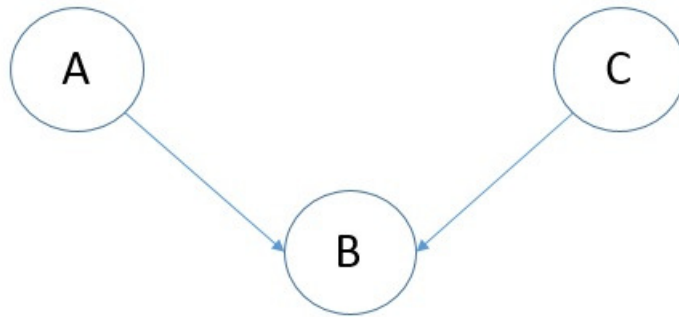
Figura 11: Cadeia de dependência entre A e B



Fonte: Elaborado pelo autor

Se a variável C for conhecida, para a cadeia constituída na Figura 11, as variáveis A e B serão dependentes. Se todas as variáveis forem desconhecidas, como exemplificada na cadeia da imagem abaixo.

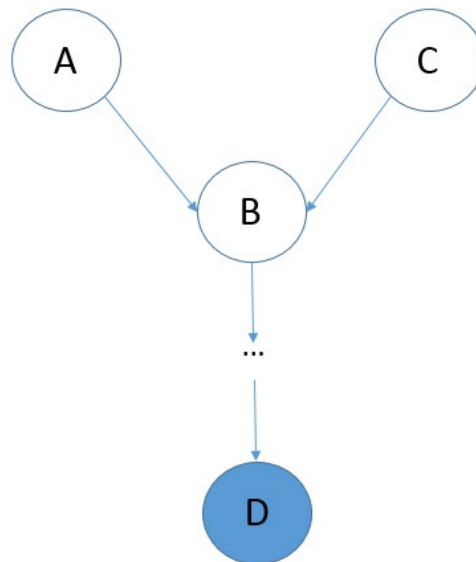
Figura 12: Cadeia de dependência entre A e C



Fonte: Elaborado pelo autor

Então as variáveis A e B serão independentes, como apresentada na cadeia acima. Se a cadeia possuir o formato da Figura 13:

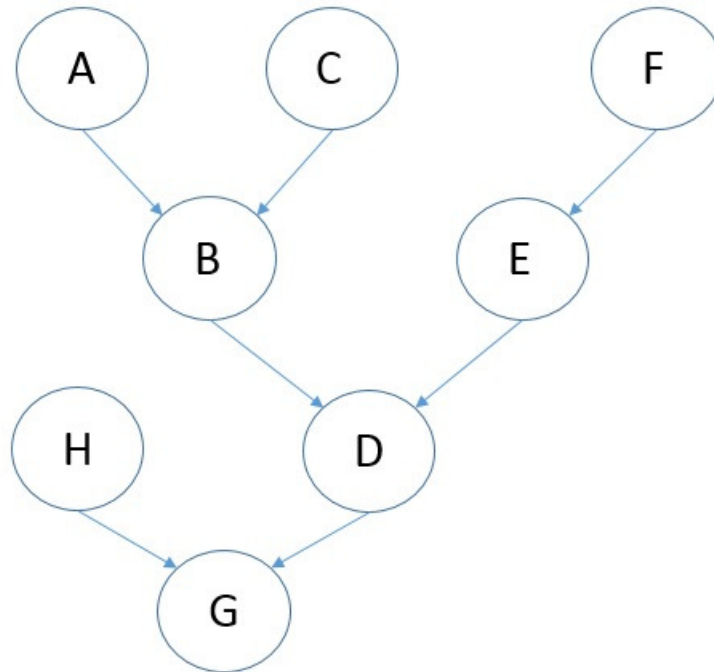
Figura 13: Relação de Dependência de B



Fonte: Elaborado pelo autor

Então as variáveis A, B e C serão dependentes dado o conhecimento de D. Considerando as definições apresentadas nesta seção, considere a Rede Bayesiana distribuída da seguinte forma:

Figura 14: Rede Bayesiana complexa



Fonte: Elaborado pelo autor

É possível afirmar que a variável F é independente de A? Sim, pois segundo das regras de separação, F é dependente de D assim como A é dependente de D, e caso D não seja conhecido, então não é possível deduzir qualquer dependência entre A e F. Entretanto, uma vez conhecida a variável D, F e A se tornam dependentes. Isto acontece porque B e A são dependentes dado o conhecimento de D, e essa dependência é repassada para A e F que possuem a variável D em comum. Agora, se a variável H e A forem conhecidas F seria independente? A resposta é sim, pois H pode dar algum conhecimento sobre a variável G, mas H não deduz nada sobre a variável D, e assim não é possível a partir de H identificar o valor para D, que é obrigatório para se chegar até F. A partir da Separação-D é possível então identificar quaisquer relações de dependência e independência condicional em uma Rede Bayesiana.

### 1.2.3 Inferência em Redes Bayesianas

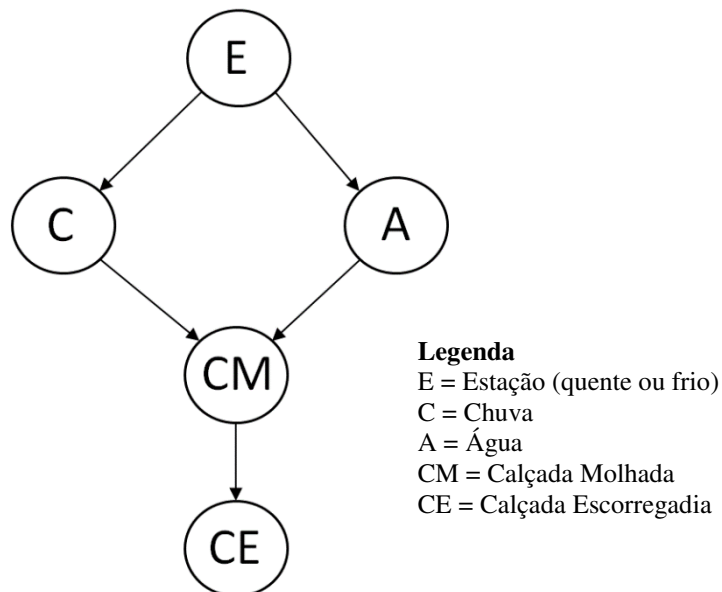
Segundo Russel e Norvig (1995, p.445), “a tarefa básica de qualquer sistema de inferência probabilística é computar a distribuição da probabilidade posterior para um conjunto de variáveis de consulta”. Dessa forma, a inferência em Redes Bayesianas trata do processo de calcular as probabilidades de novas evidências, para assim inseri-las na rede calculando as variáveis com os nós já existentes e assim propagando a informação pela rede. Existem diversos

algoritmos desenvolvidos para essa função, entre eles o algoritmo de Propagação da Árvore de Cliques (DARWICHE, 2009a).

O algoritmo Propagação da Árvore de Cliques realiza o trabalho de propagar variáveis conforme novas evidências surgem, utilizando uma árvore de junções que é formada pela união de diversos Cliques e que tem por função simplificar a quantidade de relacionamentos entre os nós, uma vez que os Cliques se relacionam apenas com os vizinhos diretamente conectados a eles (LUGER, 2013).

Cliques são subconjuntos formados pelas variáveis que compõem a rede. No grafo da seção 1.2.1 existem dois cliques: um formado pelas variáveis ( $O, A$ ) e outro com as variáveis ( $C, T e L$ ). O exemplo a seguir, proposto por Luger (2013, p.308), demonstra o uso do método de Propagação da Árvore de Cliques: considere-se que a estação do ano ( $E$ ), determina a probabilidade de chuvas em certa região ( $C$ ), além de também influenciar na probabilidade de água em um sistema aspersor ( $A$ ). A calçada molhada ( $CM$ ) correlacionada com a chuva ou com o sistema aspersor. A calçada escorregadia ( $CE$ ) estará relacionada com a calçada molhada. O grafo da Figura 15 exemplifica as relações mencionadas.

Figura 15: Grafo da Rede Bayesiana para Probabilidades de calçada escorregadia



Fonte: Elaborado pelo autor

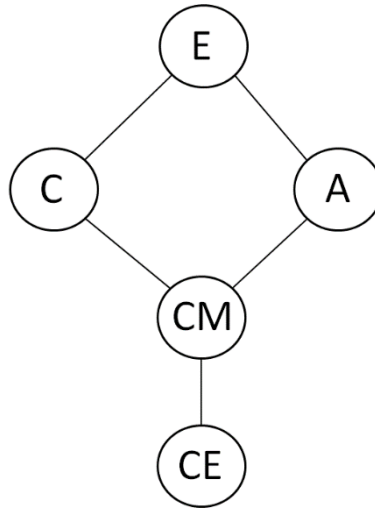
Para se descrever a probabilidade de  $CM$  é necessário calcular a correlação completa das variáveis  $C$  e  $A$ , além de suas correlações com  $E$ , uma vez que as duas causas de  $CM$  não são mutuamente independentes, conforme foi explicado na seção 1.2.2. Calculando então um valor possível  $x$  através da distribuição da probabilidade conjunta (vista na seção 1.1.5),  $x$  é dado pela soma das probabilidades conjuntas para cada estado de  $E$ (quente ou frio).

$$x = p(E)p(C|E = quente) p(A | E) + p(E)p(C|E = frio) p(A|E)$$

O grande problema dessa distribuição é que, conforme cresce o número de pais do estado, maior a complexidade do cálculo da probabilidade conjunta. O clique é justamente o valor  $X$ , que vai ter sua complexidade reduzida substituindo a propagação de restrições do gráfico direcionado acíclico descrito anteriormente. Isso é feito a partir do algoritmo de propagação de cliques desenvolvido por Lauritzen e Spiegelhalter (2013, LAURINTZEN apud LUGER, 1988, p. 308) através das seguintes etapas:

1. Trocar todos os arcos direcionados para não direcionados, resultando no grafo da Figura 16.

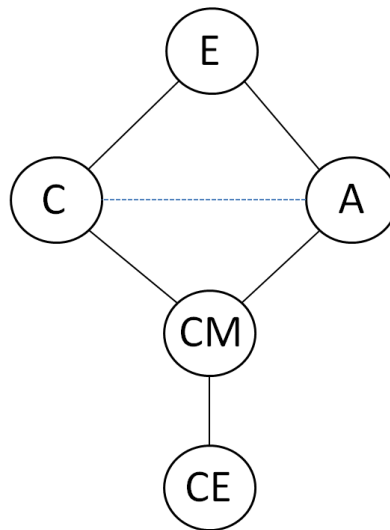
Figura 16: Trocando arcos direcionados por não direcionados



Fonte: Elaborado pelo autor

2. Para cada par de nós, desenhar arcos entre seus pais, conforme a Figura 17.

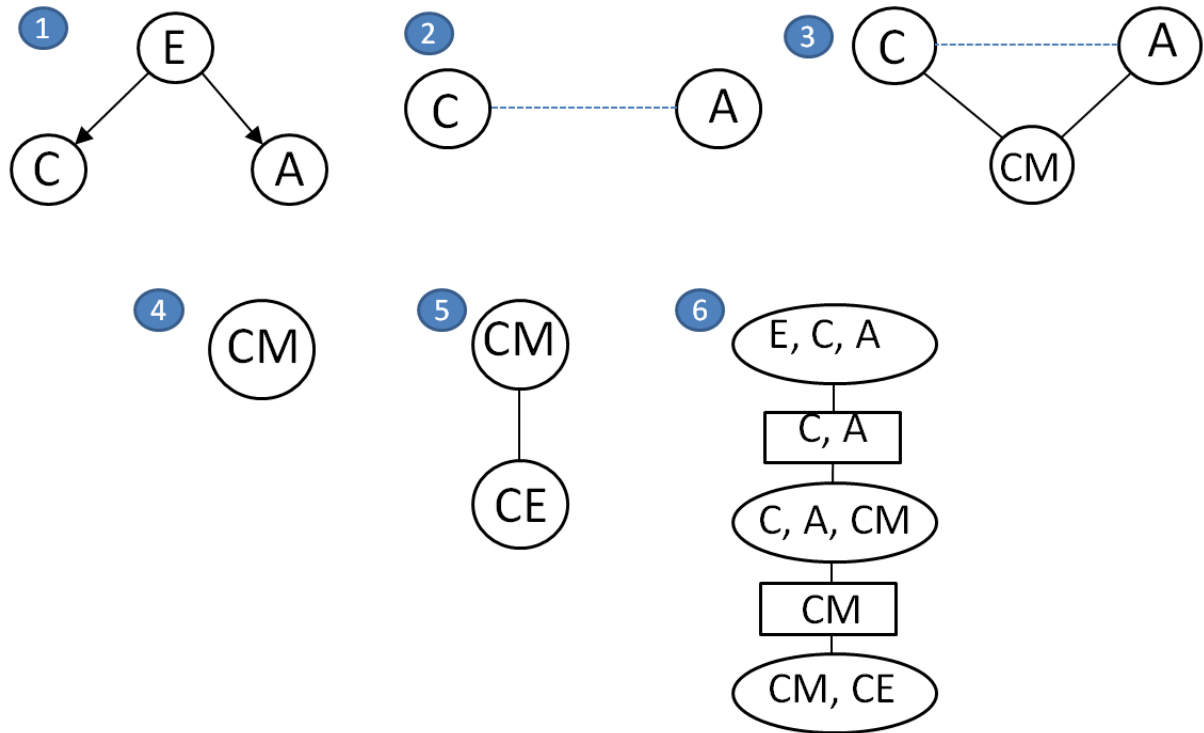
Figura 17: Desenhando arcos entre os pais



Fonte: Elaborado pelo autor

3. Identificar os ciclos de comprimento maior que três e acrescentar novos arcos para que os ciclos tenham comprimento não mais que três. No exemplo em questão, não é necessário, pois o maior ciclo não ultrapassa este valor.
4. Construir a árvore de junção com a estrutura triangulada resultante. Para isso é necessário identificar os cliques maximais (aquelas que são subgrafos completos e não subgrafos de um clique maior). As variáveis contidas nestes cliques são agrupadas em junções, que são a conexão de duas junções quaisquer que compartilham ao menos uma variável. A Figura 18 ilustra os passos deste processo.

Figura 18: Passos para criação da árvore de junção



Fonte: Elaborado pelo autor

Os passos 1, 2, 3, 4 e 5 representam as separações realizadas a partir dos cliques, o passo 6 já exibe a árvore de junção resultante.

### 1.3 Trabalhos Relacionados

A evolução das Tecnologias da Informação nas últimas duas décadas proporcionou uma mudança substancial na forma como os mercados e as organizações reagem, lidam e trabalham com a informação no relacionamento com seus clientes (CHEN, 2010). Essa evolução permitiu que ferramentas cada vez mais sofisticadas fossem desenvolvidas, fazendo com que as empresas pudessem analisar dados em tempo real, deixando assim de atuarem passivamente, ou seja, esperando que o consumidor tome a iniciativa (CHEN, 2010).

Diversas áreas foram desenvolvidas para trabalharem com este grande conjunto de que surgiram com o uso de softwares que monitoram os clientes e seus comportamentos, com o objetivo desenvolver habilidades, tecnologias, aplicações e práticas que apoiem a tomada de decisão das empresas (TEIXEIRA, 2009). Buscando atingir estes objetivos, de fornecer apoio à tomada de decisão pelas empresas, cada vez mais trabalhos concentraram esforços em analisar o ambiente em que o cliente está inserido (CHEN, 2010). Um dos passos realizados para a análise é a realização da coleta de dados de posicionamento e trajetos dos clientes em algum local específico, como shoppings centers, supermercados e lojas de departamentos, para em

seguida analisar estes dados com o propósito de identificar padrões e relações entre os locais visitados.

Além dos trajetos realizados pelos clientes, a forma como eles interagem com o ambiente também é alvo de pesquisas, com a finalidade de se identificar padrões de comportamento precorrentes a aproximação a alguma loja, ou ação de compra. As primeiras tentativas de se analisar comportamentos utilizando tecnologia ocorreram ainda no século XIX, com o rastreamento de chamadas de telefone; no século XX o comportamento passou a ser avaliado a partir de aplicações com áudio e vídeo utilizando monitoramento de câmeras, e a partir da década de 1980 surgiram as primeiras aplicações com aparelhos de GPS (*Global Position Systems*) (MICHAEL; CLARKE, 2013).

Hoje em dia grande parte dos dispositivos mobile possuem componentes GPS que permitem acessar dados de localização em tempo real. Por meio destes dispositivos, diversas aplicações como *Google Maps*<sup>1</sup>, *Nike Fuel*<sup>2</sup> e *Caledos*<sup>3</sup> apresentam mapas com as estatísticas em tempo real sobre a posição da pessoa e diversas outras informações do ambiente em que o usuário se encontra, como restaurantes próximos, bares, shoppings, velocidade de corrida, percurso e etc. Estes mesmos dados de localização são suficientes para aplicações que busquem prever trajetos, inferindo o posicionamento em um futuro próximo de acordo com associações a comportamentos, propósitos e intenções identificadas a partir do histórico de trajetos do usuário (MICHAEL; CLARKE, 2013).

Uma vez possível à predição de trajetos, pesquisas então surgem com o objetivo de avaliar estes dados sobre históricos de trajetos e realizar processos de inferência com base nas informações coletadas. Algumas destas pesquisas avaliam a intenção de compra do consumidor (YAN, 2010), acompanhadas de pesquisas que buscam entender e calcular taxas de conversão de pessoas que trafegam por *shoppings* em consumidores de algum produto ou serviço (YIM YIU; CHEONG NG, 2010).

As novas tecnologias de GPS e BI permitem alterar abordagens que antes eram realizadas (e ainda são em sua maioria) por meio de questionários e entrevistas, em métodos de avaliação da informação sem qualquer interferência ou influência do pesquisador (YIM YIU; CHEONG NG, 2010). A aplicação de métodos de inferência probabilística (estocástica) em trabalhos dentro do contexto de BI foi empregada como meio de quantificar as informações, calculando probabilidades para as ações que o usuário pode tomar mediante a presença ou

---

<sup>1</sup> <https://maps.google.com.br/>

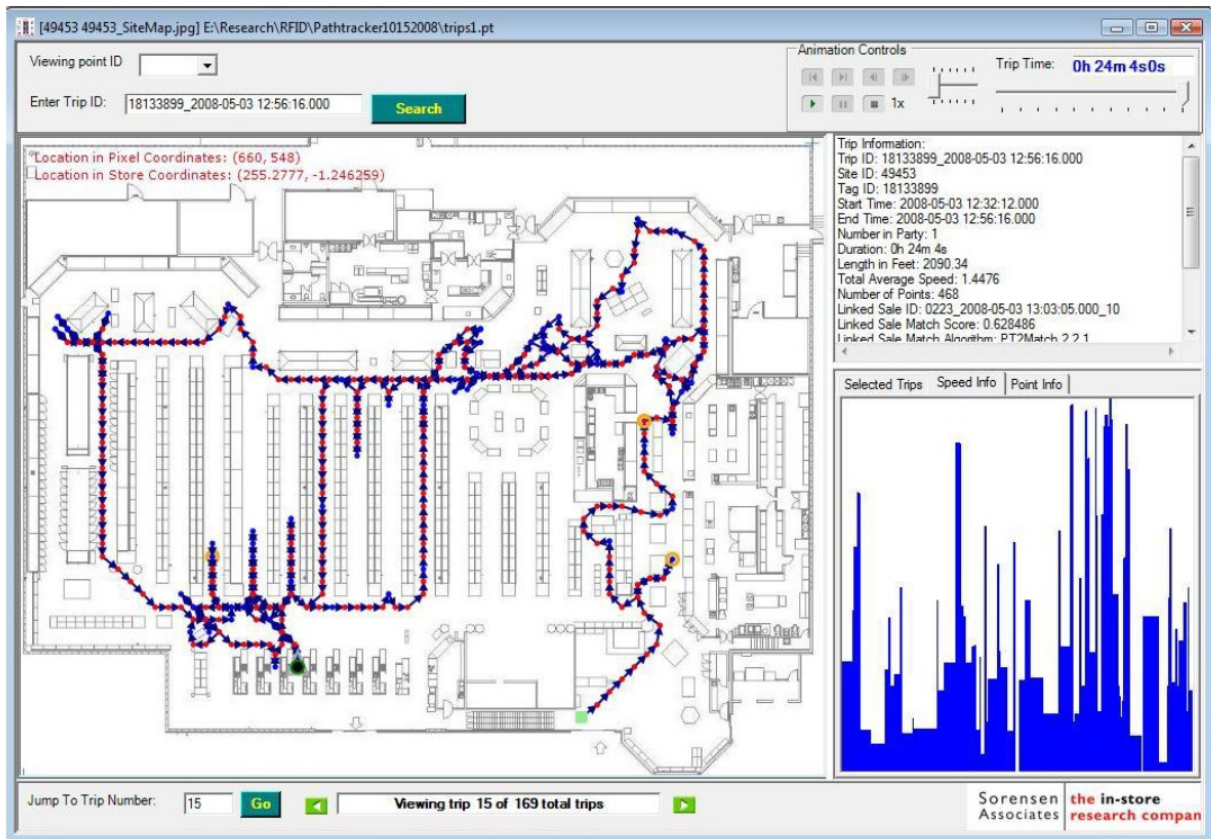
<sup>2</sup> <https://secure-nikeplus.nike.com>

<sup>3</sup> <http://www.windowsphone.com/pt-br/store/app/caledos-runner/>



ausência de alguma variável. A Figura 19 demonstra o recurso de construção de trajetória de clientes, que significa capturar todos os locais por onde o cliente passou dentro da loja para posteriormente utilizar os dados no processo de inferência probabilística.

Figura 19: Mapa de trajeto de um cliente no shopping (YAN, 2010)



Fonte: (YAN, 2010)

A Figura 19 apresenta a captura dos dados do trajeto realizado por um cliente dentro da loja. A captura foi realizada utilizando um software que acompanha o cliente por meio de uma etiqueta RFID anexada ao carrinho de compras, os dados obtidos são disponibilizados pelo software e contém o tempo total do percurso, os pontos e o tempo em cada ponto. Cada cliente pode ter seu percurso armazenado e computado por meio dos dados obtidos com o uso do software, em seguida são desenvolvidas redes que relacionam os diversos valores encontrados para cada variável (um exemplo de variável pode ser tempo de percurso total, e pontos de parada em seções).

Os algoritmos e métodos de inferência, em especial Redes Bayesianas e cálculo de Bayes, são empregados em conjunto com um uma série de variáveis que delimitam o universo estudado. Estas variáveis atuam no processo de compreensão do comportamento de compra do consumidor (YAN, 2010). É por meio da atribuição de valores a cada uma das variáveis de comportamento, relacionando-as e assim criando uma rede, que são formadas tabelas de

distribuição de conjunções e também as Redes Bayesianas, que servem para responder a diversas questões no que tange o universo estudado. Este conceito é utilizado no trabalho de (YIM YIU; CHEONG NG, 2010) para responder a seguinte pergunta: “qual a proporção de clientes do *shopping* que em meio a um passeio casual, se transformaram em consumidores de alguma loja?”.

A pergunta anterior foi respondida por meio da definição de algumas variáveis de comportamento como aproximação, tempo de permanência, trajetos realizados e categorias de produtos visitadas, onde foi concluído que algumas seções, por exemplo, o de eletrônicos, consegue converter até 30% de clientes do *shopping* em consumidores da loja (YIM YIU; CHEONG NG, 2010).

Como já foi mencionado, variáveis podem ser utilizadas como meios de se definir comportamento de consumo. Estas variáveis foram alvo de estudo em diversos trabalhos que buscam construir modelos e padrões de comportamento de compra (SANDALL, 2007). Elas buscam descrever não apenas o ambiente, mas também o cliente. Alguns exemplos de variáveis são:

- Sexo;
- Idade;
- Tempo de permanência na loja;
- Tipos de loja visitada;
- Dia na semana; e
- Produtos comprados.

A construção de redes de relacionamento causa entre variáveis de comportamento (as variáveis aleatórias da rede bayesiana), e a avaliação disso com métodos de probabilidade aplicando algoritmos de inferência, pode permitir o avanço de pesquisas que buscam compreender o que pode levar um cliente ao ato de comprar, ou então qual são seus interesses.

A tabela apresenta uma síntese das contribuições de cada trabalho supracitado para o desenvolvimento desta pesquisa.

Tabela 3: Contribuição de pesquisas para trabalho

Título do Trabalho	Referência	Pesquisa
--------------------	------------	----------

Business and Market Intelligence 2.0	(CHEN, 2010)	Artigo que buscou fazer um apanhado sobre as revoluções que a tecnologia está promovendo no mercado de negócios.
Redes de Valor e de Cooperação - evoluindo com a Inteligência Empresarial	(TEIXEIRA, 2009)	Artigo sobre <i>Business Intelligence</i> que trabalha com diversos aspectos da BI e como sua aplicação transforma as empresas atualmente.
Location and tracking of mobile devices: Überveillance stalks the streets	(MICHAEL; CLARKE, 2013)	Explora as diversas formas de rastreabilidade das novas tecnologias, como redes sem fio, RFID ( <i>Radio Frequency Identification</i> ) e NFC ( <i>Near Field Communication</i> ).
SPATIAL-TEMPORAL DATA ANALYTICS AND CONSUMER SHOPPING MODELING	(YAN, 2010)	Uso de Redes Bayesianas Dinâmicas para cálculo da probabilidade de consumidores em relação ao caminho percorrido por eles em seções de uma loja de Departamentos.
Buyers-to-shoppers ratio of shopping malls: A probit study in Hong Kong	(YIM YIU; CHEONG NG, 2010)	Utiliza Redes Bayesianas para calcular a distribuição da probabilidade conjunta para taxas de conversão de pessoas em clientes.
Comportamentos Precorrentes de Aproximação de Lojas: Efeitos do nível de diferenciação de marca e localização sobre taxas de	(SANDALL, 2007)	Dissertação sobre os efeitos de uma marca sobre o fluxo de clientes em lojas vizinhas. Explorou bastante o uso de

conversão em um shopping center Comportamentos Precorrentes de Aproximação de Lojas:		variáveis de comportamento para classificar clientes.
--	--	--

Os trabalhos relacionados na Tabela 3 foram muito importantes para o desenvolvimento deste trabalho. Cada pesquisa na Tabela 3 contribuiu com diversos elementos utilizados no decorrer deste projeto e são fonte de informação para as decisões e caminhos percorridos pelo pesquisador. Foram através deles que o autor desta pesquisa teve acesso ao emprego de redes bayesianas em cenários reais, conciliando os dados com tecnologias de posicionamento como GPS e RFID entre outras informações que permitiram o andamento desta pesquisa.

## 2 MATERIAIS E MÉTODOS

Esta seção tem como objetivo apresentar os instrumentos que serão utilizados para o desenvolvimento do presente trabalho, bem como descrever os procedimentos realizados e a análise dos dados, justificando a constituição destes elementos com base na fundamentação teórica apresentada na seção 1.

### 2.1. Local e Período de Realização da Pesquisa

A pesquisa foi desenvolvida ao longo dos semestres 2014-2 e 2015-1, e ocorreu no ambiente do CEULP/ULBRA e também da residência do próprio pesquisador. Ambos os locais foram utilizados durante todo o tempo de duração da pesquisa em períodos alternados de acordo com as reuniões realizadas com os orientadores. Foram realizadas observações no ambiente de um *shopping* local com a intenção de investigar o ambiente em que será aplicada a pesquisa, de modo a registrar eventos que foram aplicados nesta pesquisa. Esta observação foi feita de modo não participativo, em que não ocorreu nenhum tipo de interação entre o pesquisador e o objeto de estudo (clientes em um *shopping*).

### 2.2. Materiais

Os materiais utilizados para realização desse trabalho são em sua totalidade trabalhos científicos, coletados em grande parte por meio eletrônico. Entre os trabalhos utilizados encontram-se livros, artigos, monografias, teses e dissertações. Além desta documentação foi realizado do seguinte conjunto de ferramentas: *UnBBayes* utilizado para a modelagem da Rede Bayesiana, *Microsoft SQL Server* versão 2012, como Sistema de Gerenciamento de Banco de Dados, e a plataforma de desenvolvimento *Microsoft .NET Framework* com a linguagem C#. Estas ferramentas são descritas nas próximas subseções.

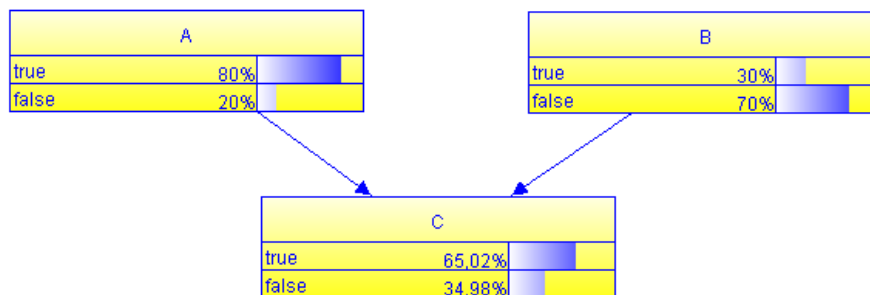
#### 2.2.1. *UnBBayes*

*UnBBayes* é um *framework* para construção de redes probabilísticas com suporte a Redes Bayesianas, desenvolvido na plataforma *Java* (CARVALHO, 2010). O *framework* é composto por uma API (*Application Programming Interface*) e uma interface gráfica, em que é possível adicionar objetos à tela e visualmente modelar uma rede bayesiana. Estes objetos são

nós, que podem ser conectados por setas a fim de definir o grafo que representa a rede e as relações entre os nós.

O *UnBBayes* foi desenvolvido pelo Grupo de Inteligência Artificial da Universidade de Brasília e encontra-se disponível<sup>4</sup> desde o ano 2000 (MATSUMOTO, 2008). A Figura 20 apresenta uma rede composta por 3 variáveis.

Figura 20: Modelagem de uma rede bayesiana no UnBBayes



Fonte: UnBBayes editado pelo Autor

A rede ilustrada na Figura 21 é composta pelas variáveis A, B e C, onde C tem como pais as variáveis A e B. Todas as variáveis possuem os estados *true* e *false*, e a probabilidade para a variável C é calculada em razão das probabilidades de seus pais. A imagem mostra a Rede Bayesiana com as probabilidades já calculadas através das tabelas de probabilidade condicional que foram inseridas para cada variável.

### 2.2.2. Microsoft SQL Server

*Microsoft SQL Server* é um Sistema de Gerenciamento de Banco de Dados fornecido e desenvolvido pela *Microsoft* (MICROSOFT, 2015a). A partir do *Microsoft SQL Server* é possível criar e administrar bancos de dados relacionais, que podem ser acessados de diversos ambientes (intranet, internet). O *Microsoft SQL Server* foi desenvolvido em 1989, e posteriormente teve incorporado ao seu código a base do Sybase SQL Server em 1998. O SQL Server é continuamente desenvolvido e hoje encontra-se na versão 12.0, conhecida como SQL Server 2014.

<sup>4</sup> <http://sourceforge.net/projects/unbbayes/>

### 2.2.3. Microsoft .NET Framework

O *Microsoft .NET Framework* é desenvolvido pela *Microsoft* para funcionar inicialmente na plataforma *Windows* (MICROSOFT, 2015b). O *framework* contém um conjunto de bibliotecas de classe que tem como objetivo prover interoperabilidade entre uma série de linguagens de desenvolvimento que fazem parte do ambiente do .NET Framework, entre estas linguagens está o C#. A linguagem de programação C# apresenta uma sintaxe bem similar ao Java, além de tratar-se de uma linguagem orientada a objetos.

O .NET Framework foi lançado em sua versão final em 2002, e desde então está em contínuo desenvolvimento pela Microsoft.

## 2.3. Método de Trabalho

Esta seção tem como objetivo apresentar as etapas que foram elaboradas para o desenvolvimento da pesquisa. As etapas que dividiram o trabalho foram elaboradas com base no próprio modelo de construção de uma Rede Bayesiana já especificado na seção 1.2, com o acréscimo de algumas etapas que coletem dados para serem utilizados na criação da rede. As etapas foram elaboradas e executadas na seguinte ordem:

1. Pesquisa sobre comportamento do consumidor em *shoppings*;
2. Coleta de Dados;
3. Definição dos estados possíveis de cada variável;
4. Construção da estrutura da rede, conectando as variáveis para formar o grafo;
5. Definição da tabela de probabilidade condicional para cada nó da rede; e
6. Aplicação de Cenários com base nos dados coletados (experimentos).

A seguir, estas etapas serão apresentadas em detalhes.

### 2.3.1. Pesquisa sobre comportamento do consumidor em shoppings

Para identificar as variáveis que fazem parte da rede, primeiro foi necessário saber o que pode influenciar o comportamento do consumidor. Para isso foram consultados trabalhos e pesquisas que tivessem envolvimento com análise de consumidores em *shoppings centers*. Dos trabalhos pesquisados aqueles que mais contribuíram para esta pesquisa foram apresentados na seção 1.3.

Durante a pesquisa por características do comportamento do consumidor, diversas variáveis foram encontradas, como estado civil, distância do *shopping* até a residência, tempo

médio de permanência no *shopping*, gênero entre outras. Entretanto, para o presente trabalho, foi selecionado um número limitado de variáveis, segundo o critério: para fazer parte da Rede Bayesiana criada no trabalho, a variável deve ter dados quantificáveis, ou seja, deve ter valores reais possíveis de serem alcançados pelo pesquisador, por exemplo o gênero pode ser observado como masculino ou feminino, ou idade, que pode ser de 16 a 34 anos, e de 35 anos ou mais.

### 2.3.2. Coleta de Dados

De acordo com as variáveis que foram selecionadas na etapa anterior, foi elaborada uma pesquisa *online* com o objetivo de coletar dados que sejam suficientes para a etapa futura de definição de tabelas de probabilidade condicional e *a priori* na etapa 5. Por dados suficientes entende-se por volume de pessoas que respondam ao questionário. A quantidade de respostas foi calculada observando-se o fluxo de pessoas no *shopping* Capim Dourado em um intervalo de 1 hora. Para identificar a quantidade necessária foram utilizadas informações disponíveis no site do *shopping*, que informa a quantidade de visitas anuais, e deste número foi feita a média mensal e em seguida a média diária. A Coleta de Dados também ajudou na seleção de cenários que foram aplicados na etapa 6.

Dados coletados no site oficial do *shopping* Capim Dourado informam que anualmente o *shopping* recebe 4.947.114 milhões de visitas. Com esta quantidade de visitas pode-se chegar ao número 1374 visitas por hora. Este valor por hora pode ser obtido calculando-se a média dos meses e dividi-la pelo valor correspondente a um dia de funcionamento. Dado a quantidade de visitas por hora, para que seja possível simular um fluxo de pessoas são necessárias 110 pessoas, que corresponde a quantidade de indivíduos do espaço amostral necessária para obter 9% de margem de erro amostral, e 95% de confiança.

O questionário que foi aplicado pela pesquisa é composto por 5 perguntas de múltipla escolha. Na primeira questão será solicitado que o entrevistado relacione as 5 lojas que mais visita no *shopping* Capim Dourado. As lojas foram apresentadas ao entrevistado em um formato de listas, separadas por categorias. O entrevistado pode então selecionar quaisquer 5 lojas, que puderam ser da mesma categoria, ou de categorias diferentes. A próxima questão solicitou que o entrevistado identificasse qual gênero pertence, em seguida foi solicitado que informasse a renda familiar. Após informar a renda familiar a pesquisa solicitou ao entrevistado que responda se ele prioriza a comodidade que uma loja oferece em vez dos preços. Por fim, foi solicitado que o entrevistado respondesse se tem preferência por lojas que estão mais próximas a saídas do *shopping*.



O objetivo da pesquisa foi alcançar o número proposto de respostas para a amostragem definida, e fornecer as bases para os cálculos que foram realizados na etapa de desenvolvimento deste trabalho.

### 2.3.3. Definição dos estados possíveis de cada variável

Na etapa de definição dos estados possíveis de cada variável, as variáveis selecionadas na etapa 1 foram analisadas e seus possíveis estados foram definidos. O critério utilizado para determinar a quantidade máxima de estados que uma variável possui tem como objetivo controlar o nível de complexidade do trabalho. Para isto deve-se considerar que a quantidade de cálculos para um nó da Rede Bayesiana é dada pela quantidade de estados do nó atual, elevado pelo somatório da quantidade de estados dos pais deste nó. Sendo assim foi definido que cada nó pode ter no máximo dois estados. Desta forma, uma variável com 2 pais em que cada pai possui 2 estados, produz uma tabela de probabilidade condicional com 16 entradas.

Outro critério é que o estado deve representar algum significado do mundo real, por exemplo, devem ser valores que podem ser observados no ambiente de um *shopping*, como estados concretos do tipo “gênero masculino”, que pode ser observado no mundo real, em vez de “gênero 1”, que não apresenta muito significado.

### 2.3.4. Construir a estrutura da rede, conectando as variáveis para formar o grafo

O objetivo da etapa de construir a estrutura da rede é montar a estrutura da rede ligando as variáveis definidas na etapa 1 de uma forma que fique estabelecida entre as variáveis uma relação de causa e efeito, ou seja, que uma variável de efeito preceda uma variável que seja uma causa possível. Também foi utilizado como critério que as ligações estabelecidas entre as variáveis deveriam ter sentido dentro do contexto do *shopping*, de modo a representar uma relação possível no mundo real.

### 2.3.5. Definir a tabela de probabilidade condicional para cada nó da rede.

Com a estrutura da rede montada, na etapa de definição da tabela de probabilidade condicional foram utilizados os dados coletados na etapa 2 além de pesquisas em livros, periódicos científicos, sites com dados estatísticos como IBGE e matérias de revistas. A tabela de probabilidade condicional exibe os estados de uma variável, e também os relacionada com

os estados das variáveis que sejam Pais do nó atual. Isto mostra o sentido de condição, em que para cada valor de um estado do nó pai existe um valor no nó filho. Foi necessário cruzar informações entre as fontes de pesquisa para poder calcular a probabilidade condicional de cada nó. Para o cálculo da probabilidade de cada variável as fontes foram utilizadas segundo a Tabela 4.

Tabela 4: Tabela de fontes

<b>Variável</b>	<b>Fontes</b>
Idade	IBGE
Gênero	Pesquisa <i>online</i> , IBGE
Comodidade	Pesquisa <i>online</i>
Atratividade	Pesquisa <i>online</i>
Poder de Compra	Pesquisa <i>online</i> , IBGE
Marcas	Pesquisa <i>online</i> , (GAZETA ONLINE, 2014)
Variáveis de Lojas	Pesquisa <i>online</i> , (DOURADO, 2015)
Variáveis de Categoria	Pesquisa <i>online</i> , (DOURADO, 2015)

Para calcular as probabilidades condicionais, além das fontes na Tabela 4 buscou-se sempre estabelecer uma relação de causa e efeito entre as variáveis, e quando não foi possível determinar a probabilidade a partir dos dados obtidos, estabeleceu-se um o valor de 50%.

### 2.3.6. Aplicação de Cenários com base nos dados coletados

Com a rede completamente montada na etapa 5, a última etapa do desenvolvimento buscou aplicar cenários para demonstrar o modelo de inferência desenvolvido. Os cenários foram selecionados dentre os dados coletados na etapa 2. O cenário representa um conjunto de dados de resposta, informados por um participante da pesquisa online. Os indivíduos que tiveram suas respostas selecionadas para os cenários foram escolhidos aleatoriamente, e ao todo foram desenvolvidos dois cenários, que foram apresentados da seguinte forma:

- a. Apresentou-se o primeiro cenário e foram feitas considerações a respeito dos dados obtidos.
- b. Apresentou-se uma alteração do primeiro cenário, de modo a comparar as mudanças decorrentes da alteração e considerações a respeito dos dados obtidos;

- c. Apresentou-se o segundo cenário e novamente foram feitas considerações sobre os dados obtidos.
- d. Apresentou-se também uma alteração do segundo cenário e novas considerações foram feitas a respeito dos dados que foram obtidos com a mudança.

O objetivo do primeiro cenário é apresentar as mudanças que ocorrem na rede de acordo com alterações na variável Gênero, enquanto o segundo cenário explora as mudanças que ocorrem na rede quando alteram-se os estados da variável Poder de Compra.

Este capítulo apresentou os procedimentos metodológicos aplicados nesta pesquisa, e é fundamental para definir as direções que foram tomadas e o porquê de serem feitas, de acordo com as explicações apresentadas. A próxima seção apresentará os resultados obtidos na pesquisa.

### 3 RESULTADOS E DISCUSSÃO

Esta seção apresenta o desenvolvimento e os resultados obtidos na realização deste trabalho. Foi desenvolvido um modelo de inferência que, baseado nas características de um cliente, busca dizer o quão provável é o interesse daquele cliente em uma categoria de produtos, utilizando-se para isso uma Rede Bayesiana com variáveis que representem estas características e seus relacionamentos. A rede Bayesiana foi criada respeitando sua definição (LUGER, 2013) que diz:

- Cada nó está associado a uma variável aleatória;
- Um conjunto de vínculos orientados, ou setas, conecta pares de nós e estabelece uma relação de Pai para Filho entre os nós conectados;
- Cada nó  $X_i$  tem uma distribuição de probabilidade condicional  $P(X_i | P(Pais(X_i)))$ .

As próximas seções apresentam o desenvolvimento das etapas que foram elaboradas na seção 2.3 na ordem que foram definidas. A seção 3.1 apresenta as variáveis selecionadas para o modelo, em seguida a seção 3.2 demonstra a pesquisa e tratamento de dados que foram necessários para calcular as probabilidades das variáveis. Após obter os dados da pesquisa online a seção 3.3 exibe o desenvolvimento da rede bayesiana e por fim na seção 3.4 é explicado os cenários que foram aplicados a rede.

#### 3.1. Variáveis

Para o desenvolvimento desta pesquisa foi necessário definir e atribuir valores a um conjunto de variáveis que classificam e caracterizam o cliente e ambiente estudado. Para fazer este processo de analisar clientes e lojas de um *shopping center* e assim, descobrir interesses em categorias de produtos, se fez necessário classificar cada um destes elementos de forma que possam ser trabalhados adequadamente. As próximas subseções têm como objetivo especificar as variáveis que compõem a Rede Bayesiana da pesquisa.

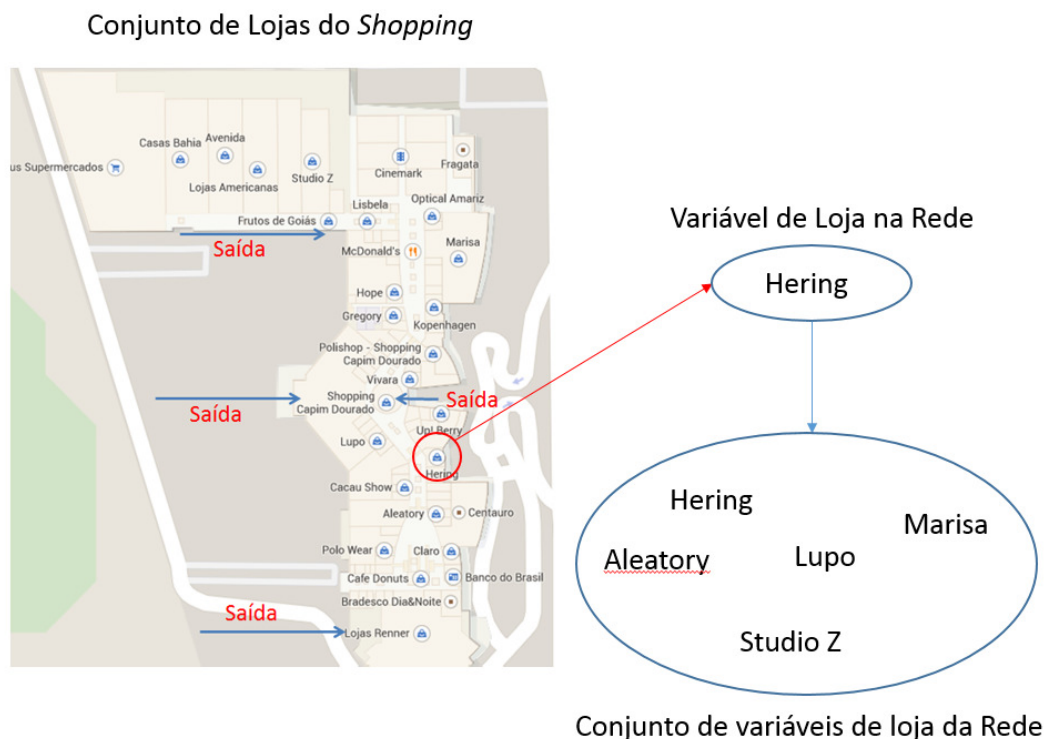
##### 3.1.1. Comodidade

A variável “comodidade” tem a função representar o impacto de ações executadas pela loja no sentido de tentar influenciar no processo de escolha do consumidor, por exemplo, horário de funcionamento diferenciado, estacionamento e atendimento.

### 3.1.2. Lojas

As variáveis do tipo “loja” representam lojas do *shopping* na Rede, ou seja, para cada loja no *shopping*, haverá uma variável que a represente na Rede Bayesiana. Seguindo este princípio, cada variável que representa uma loja na Rede Bayesiana terá uma relação com uma outra variável que será a sua categoria. As lojas que compõe a rede são visualizadas na Figura 21, e o nome da variável na rede é o mesmo nome da loja na imagem, para facilitar a diferenciação entre as lojas. A relação entre lojas do *shopping* e uma variável na Rede Bayesiana pode ser vista na Figura 21.

Figura 21: Relação de lojas e suas variáveis



Fonte: Elaborada pelo Autor

A Figura 21 mostra do lado esquerdo um mapa com todas as lojas do *shopping*, a direita mostra que uma loja é considerada uma variável, que neste caso foi a loja Hering, ainda na imagem é possível observar que Hering faz parte de um conjunto, que possui diversas outras lojas, ou seja, para cada loja do *shopping*, uma variável é criada, e esta variável está em um grupo de variáveis que retratam as lojas do *shopping*.

Nem todas as lojas do *shopping* fazem parte do modelo, para isto foi realizada uma seleção, em que a lista de lojas adicionadas à rede foi definida após a realização da pesquisa de coleta de dados.

### 3.1.3. Categorias

Como mencionado na seção 3.1.2, a variável “categoria” representa o tipo de produto que a loja comercializa, a Tabela 5 mostra a relação de lojas por categorias.

Tabela 5: Lojas por Categorias

<b>Categoria</b>	<b>Loja</b>
Moda Unisex	Hering
	Aleatory
	TNG
Lazer e Entretenimento	Zig & Zag
	V12 Games
	Cinemark
Calçados e Acessórios	Havaianas
	Capodarte
	Studio Z

As categorias de produtos utilizados neste trabalho foram retiradas da relação de categorias divulgadas no site do *shopping* local. Foram selecionadas apenas categorias que possuíam no mínimo duas lojas, para evitar situações em que a probabilidade para tal categoria sempre satisfaz 100%, já que não existem outras lojas comercializando produtos da categoria em questão. Um exemplo do problema de uma categoria ter apenas uma loja, é que dado qualquer probabilidade maior de 0% na loja, automaticamente as chances de interesse naquela categoria seriam de 100%, uma vez que não há outras lojas na distribuição da probabilidade para aquela categoria.

### 3.1.4. Gênero

A variável “gênero” representa o gênero (masculino ou feminino) do indivíduo. Esta variável identifica a medida através da natureza de grupos (2007, Dodd apud Sandall, 1998, p.36), ou seja, os indivíduos são classificados de acordo com alguma característica e assim fazem parte de algum grupo, por exemplo diferenças em grupos de homens e mulheres enquanto consumidores.

### 3.1.5. Poder de Compra

Da mesma forma como a variável “Gênero”, a variável “Poder de Compra” é uma medida através da natureza de grupos, separando os indivíduos em classes sociais. Classes sociais discriminam a renda das famílias, determinando assim o poder de compra que elas possuem em razão dos recursos financeiros que possuem, e foram obtidas no banco de informações do IBGE. Após o levantamento das classes elas foram divididas em dois estados, sendo o estado ClasseAB o conjunto de pessoas pertencentes as classes A e B, e o estado ClasseCDE o conjunto de pessoas que pertencem as classes C, D e E.

### 3.1.6. Atratividade

A variável “Atratividade” é uma entre três etapas subsequentes (atração, conversão e despesa) que altera o comportamento do cliente com a loja (LAM, 2001), e tem como função medir o quão atraente a loja é para o cliente. A atratividade no contexto do trabalho é a razão entre a quantidade de pessoas que selecionaram a loja na pesquisa *online* pela quantidade total de usuários que selecionaram uma loja do mesmo grupo. Por exemplo, caso 1 usuário tenha selecionado Aleatory, para saber a atratividade de Aleatory é necessário dividir 1 pela quantidade total de pessoas que selecionaram uma das 3 lojas da categoria de Moda Unissex, uma vez que Aleatory pertence a este grupo, caso 5 pessoas seja o total de usuários que selecionaram qualquer uma das 3 lojas da categoria Moda Unissex, a atratividade então é representado por:  $\frac{1}{5} = 0,2$ .

### 3.1.7. Marca

A variável “Marca” indica o interesse do cliente por um tipo específico de produto no que diz respeito ao seu tipo de fabricação ou responsável, que neste trabalho é dividido por “popular” e “grife”, além de agir como estímulo ao consumo promovendo também medidas de diferenciação e satisfação para o cliente (Sandall, 2007).

### 3.1.8. Idade

A variável “Idade” representa grandeza de grupos, classificando indivíduos da mesma forma que as variáveis “Poder de Compra” e “Gênero”. Para a classificação foram utilizados dados do IBGE para as faixas etárias de Palmas/Tocantins. Os estados da variável idade são

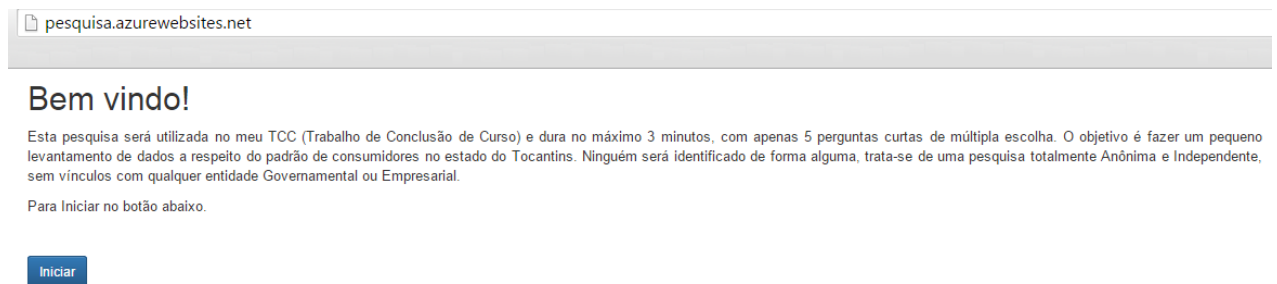
*até34* que representa pessoas que possuem no máximo 34 anos, e *35mais* que representa pessoas que possuem de 35 anos em diante.

### 3.2. Pesquisa e tratamento dos dados

Após selecionar as variáveis do modelo, foi necessário buscar dados que forneceram o embasamento das probabilidades que foram atribuídas aos nós. Em geral as informações utilizadas na construção de Redes Bayesianas são determinadas através de especialistas do domínio (DARWICHE, 2009b) ou utilizando alguns métodos de lógica proposicional, quando lida-se apenas com variáveis determinísticas, ou seja, variáveis em que seus estados variam sempre entre 0 – *false* e 1 – *true*.

Para este trabalho foi definida a utilização de uma pesquisa eletrônica para coleta de dados, que foi apresentada na seção 2.3.2. Para realizar a pesquisa foi utilizado uma aplicação feita especificamente para este trabalho, desenvolvida na plataforma .NET utilizando o ASP.NET para criação do site e o SQL Server 2012 como ferramenta de banco de dados. Ao acessar o site o usuário encontrou a tela que explica o objetivo da pesquisa, e em que ela será utilizada. A Figura 22 exibe a tela inicial da aplicação.

Figura 22: Tela inicial da pesquisa web

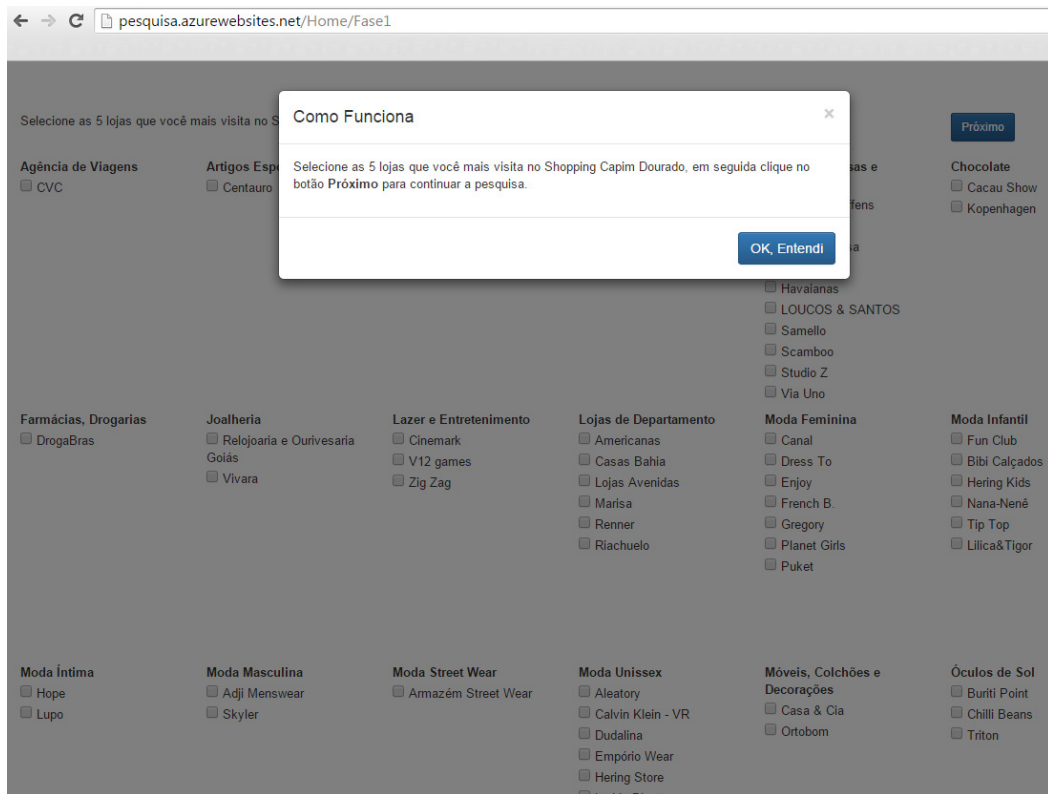


Fonte: <http://pesquisa.azurewebsites.net/>

Após clicar no botão de **Iniciar**, a próxima tela exibida mostrava uma janela em que era explicado o que o usuário poderia responder nesta primeira etapa, e pede a confirmação do usuário. As Figuras 23 e 24 exibem a sequência de confirmação do usuário, e as opções disponíveis para que o usuário selecione.

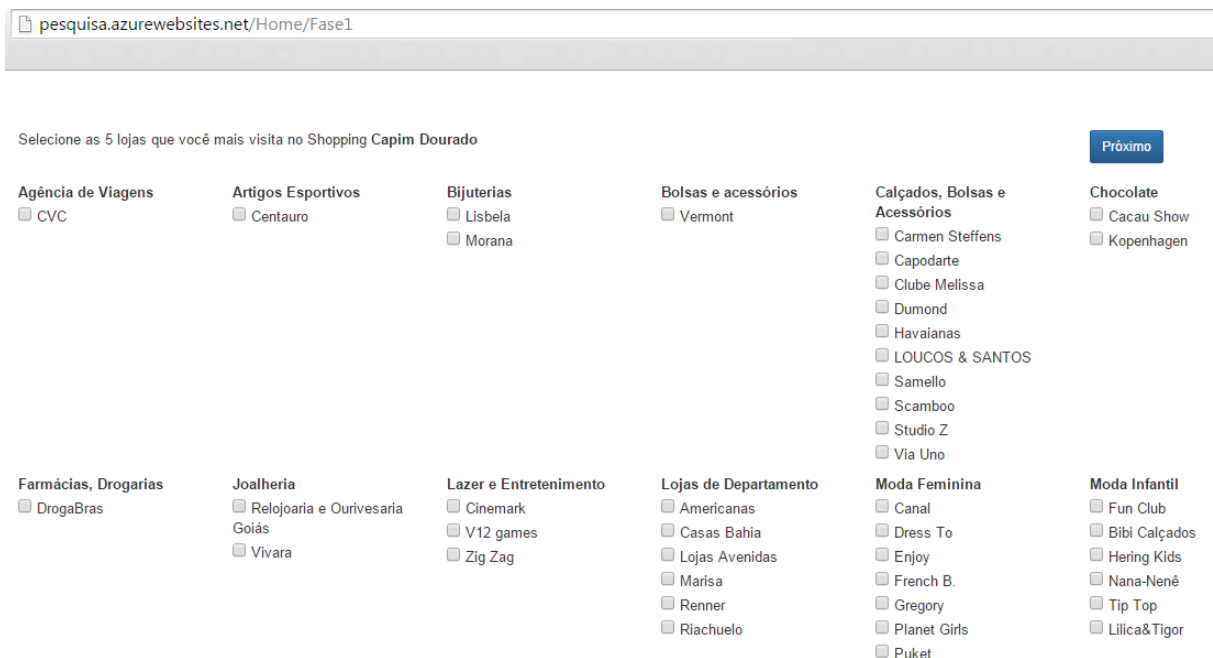


Figura 23: Confirmação da primeira fase da pesquisa



Fonte: <http://pesquisa.azurewebsites.net/>

Figura 24: Tela de seleção de lojas



Fonte: <http://pesquisa.azurewebsites.net/>

O usuário foi orientado a clicar no botão “Próximo” quando já tivesse selecionado todas as 5 lojas para que avançasse para a próxima etapa. Na fase dois da pesquisa eletrônica foi

solicitado mais quatro informações do usuário, que foram: Seu gênero, renda familiar, priorização de comodidade e preferência por lojas mais próximas as saídas. Por fim, o usuário deveria clicar no botão “Concluir” após selecionar as respostas. A Figura 25 mostra a tela da segunda fase da pesquisa eletrônica.

Figura 25: Fase dois da pesquisa eletrônica

pesquisa.azurewebsites.net/Home/Fase2

Nesta ultima etapa, seleccione as opções que melhor se apliquem.

Gênero

- Masculino
- Feminino
- Outro

Renda Familiar (Somatório da renda individual dos moradores do mesmo domicílio)

- Até R\$ 1.449,99
- Até R\$ 2.899,99
- Até R\$ 7.249,99
- Até R\$ 14.499,99
- Acima de 14.500,00
- Desejo não declarar

Prioriza comodidade que a loja oferece em vez dos preços?

- Sim
- Não

Prioriza as lojas mais próximas as saidas do Shopping?

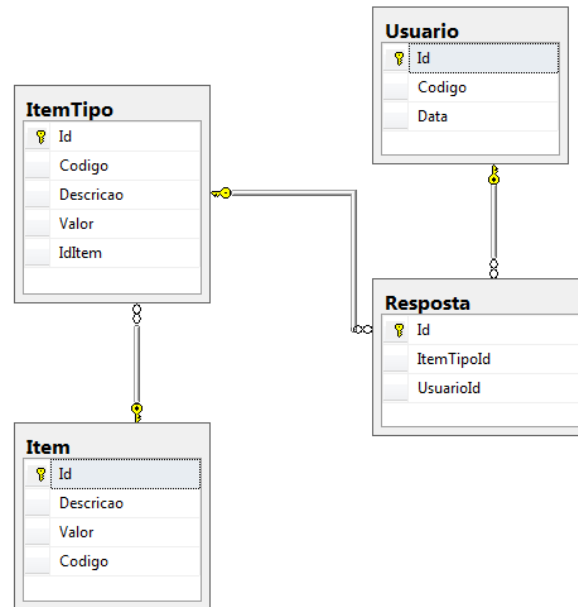
- Sim
- Não

**Concluir**

Fonte: Elaborada pelo autor

A estrutura do banco de dados foi desenvolvida para que fosse registrado um usuário, assim como todas as respostas que foram dadas por ele. Entende-se por resposta cada item que foi selecionado por ele na pesquisa. Por exemplo, quando o usuário seleciona o gênero masculino, esta informação é compreendida como uma resposta, da mesma forma, quando o usuário seleciona uma loja das cinco lojas que ele pode selecionar, é armazenado como uma resposta, e assim selecionar as cinco lojas resulta em cinco registros na tabela de respostas. A estrutura do Banco de Dados pode ser observada na Figura 26.

Figura 26: Diagrama de Banco de Dados



Fonte: Software Microsoft SQL Server 2012

Conforme apresenta a Figura 26, o banco é composto por quatro tabelas. A tabela de respostas armazena as informações que o usuário selecionou na aplicação, e esta tabela referencia outras duas tabelas que são **Usuário** e **ItemTipo**. A tabela **Usuário** armazena os usuários que participaram da pesquisa, os registros inseridos nesta tabela são *Código*, que é do tipo *Guid* e tem como função representar o usuário, e o campo *Data*, que informa a data em que o usuário participou da pesquisa. Os registros da tabela de **Usuário** são criados no instante em que ele seleciona as lojas e clica em continuar na aplicação.

A tabela **Item** representa as categorias de questões que foram feitas na pesquisa. Por exemplo, a questão que pede para o usuário relacionar as cinco lojas mais visitadas no *shopping* é da categoria Lojas, enquanto que a questão que pede para o usuário identificar o Gênero a que pertence é categorizada como Gênero. Ao todo existem cinco categorias de pergunta que são: Gênero, Renda, Lojas, Proximidade e Comodidade, cada uma representando uma questão na pesquisa.

A tabela **ItemTipo** representa as opções de resposta que cada categoria possui. Por exemplo, na tabela Item existe a categoria Gênero, esta categoria deve possuir três opções de resposta que são: Masculino, Feminino ou Outro. Dessa forma, a tabela **ItemTipo** possui três registros que representam cada uma destas opções de resposta, que por sua vez estão relacionadas a alguma categoria cadastrada na tabela Item.

O objetivo da pesquisa era alcançar o número de 110 respostas, para atingir 95% de confiabilidade e 9% de erro amostral para uma amostragem de 1374 visitas no período de uma

hora de acordo com o critério adotado na seção 2.3.2. Ao todo 113 pessoas responderam ao questionário completamente, e algumas informações puderam ser extraídas deste processo, como:

- 54% das pessoas eram do sexo masculino, e 46% do sexo feminino;
- 6% das pessoas tinham renda familiar superior a R\$ 14.500,00, e também 6% ganhavam até R\$ 1.449,00.
- 32% valorizam mais a comodidade do que os preços;
- 19% preferem lojas mais próximas as saídas.
- “Cinemark” foi o estabelecimento mais votado tanto por indivíduos do sexo masculino quanto feminino.

Através da pesquisa foi possível selecionar as lojas e as categorias que fizeram parte da Rede Bayesiana, que foram:

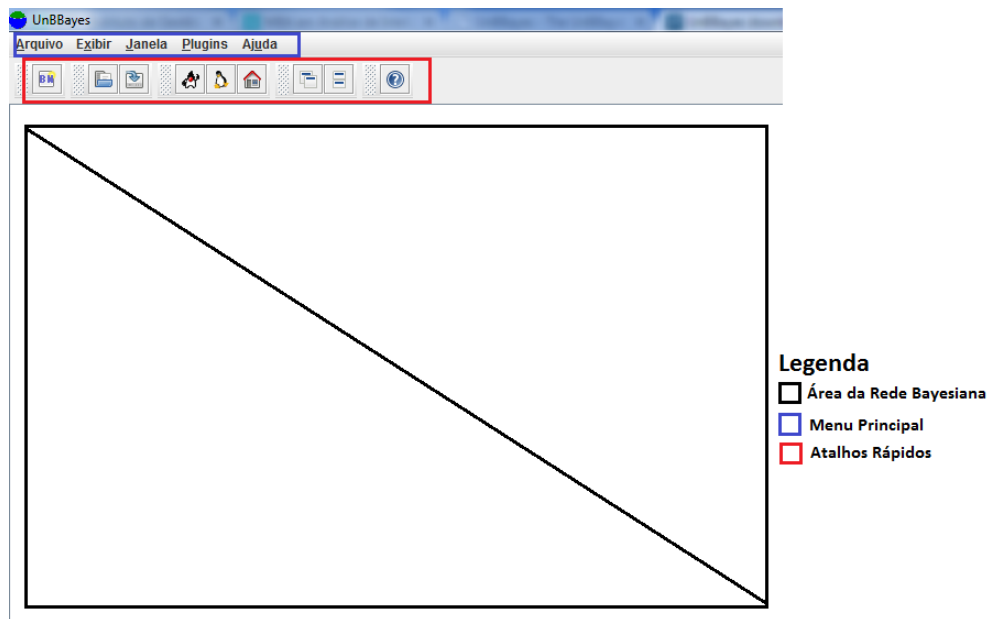
- Categorias: Moda Unisex, Lazer e Entretenimento e por fim Calçados, bolsas e Acessórios.
- Lojas: Hering, Aleatory, Zig Zag, V12 Games, Cinemark, Havaianas, Capodarte e Studio Z.

A pesquisa questionou sobre a renda familiar dos indivíduos entre as 5 classificações de renda utilizadas pelo IBGE, mais a opção de “Desejo não declarar”. As respostas então foram separadas em dois grandes grupos, sendo o grupo AB de pessoas declararam renda familiar entre R\$ 7.300,00 e acima, e de pessoas que fariam parte do grupo CDE, que são todos os usuários que declararam renda familiar de até R\$ 7.299,00.

### **3.3. Construção da Rede Bayesiana com UnbBayes**

Para o desenvolvimento da Rede Bayesiana, o primeiro passo é ter o software para modelagem corretamente instalado. A tela inicial do UnbBayes é apresentada na Figura 27.

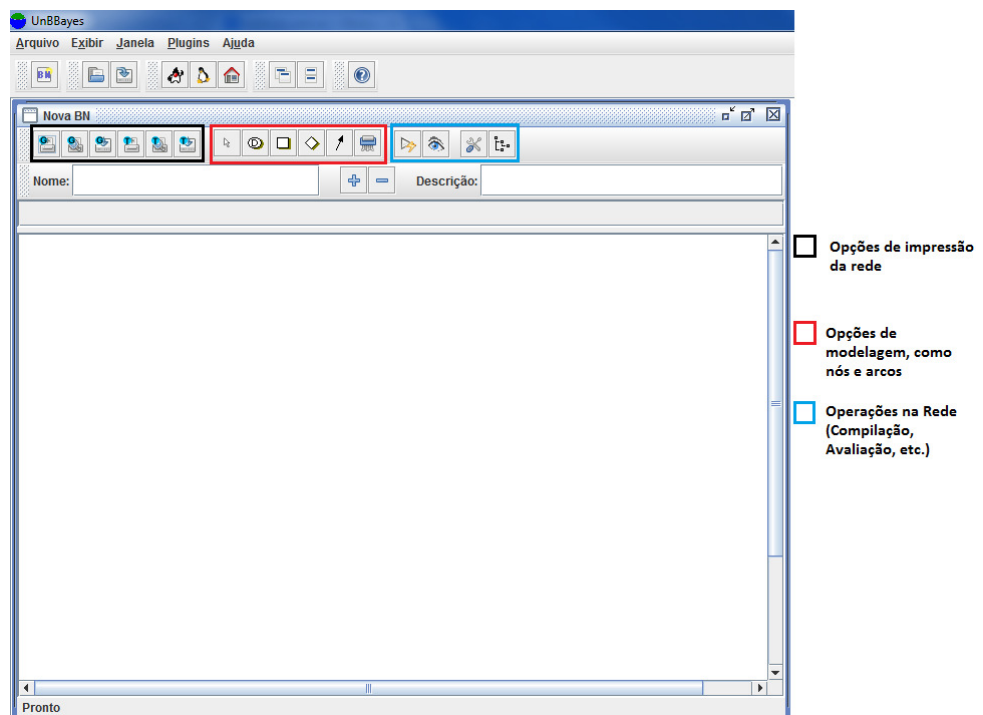
Figura 27: Tela principal do UnBBayes



Fonte: Elaborada pelo autor

Uma nova Rede Bayesiana pode ser criada clicando no atalho “BN” do menu de atalhos rápidos, ou clicando na opção Arquivo, em seguida selecionar a opção novo e clicar no item “BN”. Neste momento uma nova janela será aberta na área destinada a Rede Bayesiana, com as ferramentas necessárias para desenvolver toda a rede, conforme apresenta a Figura 28.

Figura 28: Tela de modelagem da Rede Bayesiana no UnBBayes



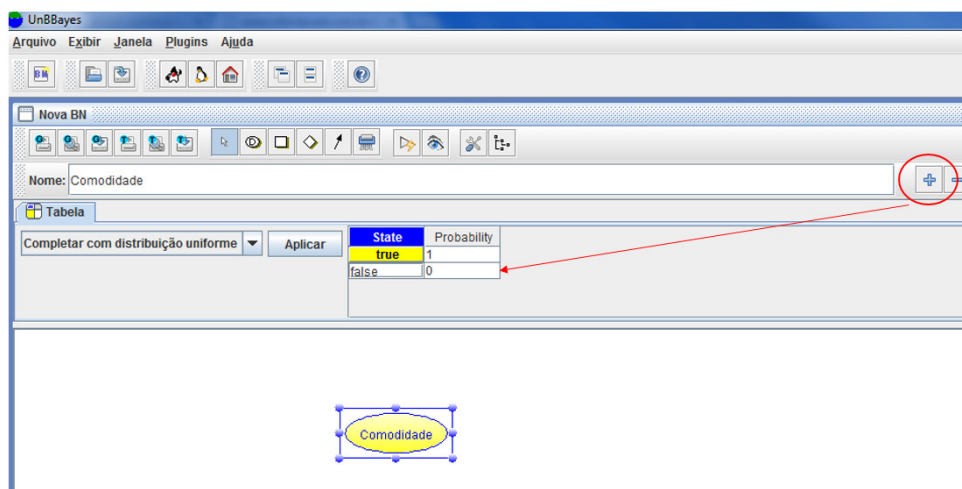
Fonte: Elaborada pelo autor

Primeiramente todas as variáveis foram distribuídas pela rede. As variáveis que compõem a Rede Bayesiana foram definidas na seção 2.1. Para adicionar uma variável a Rede é necessário clicar no item “Adicionar variável de probabilidade” que está localizada na área grifada em vermelho na Figura 28, que representa o conjunto de itens de modelagem. Após clicar na opção, basta clicar novamente em qualquer ponto da área em branco e a variável será inserida. Com a variável na rede pode -se renomeá-la para deixar mais compreensível ao usuário que está fazendo a modelagem.

A primeira variável a ser criada é “Comodidade” – a ordem em que as variáveis serão inseridas não afeta a rede desde que sejam respeitadas as relações entre elas. Cada variável possui um conjunto de estados que refletem as crenças que existam em uma variável. Crenças se referem as probabilidades atribuídas aos estados a partir de algum nível de conhecimento do especialista que está modelando a rede. Para a variável comodidade definiu-se os estados *true* e *false*, em que *true* representa a probabilidade de que comodidade exista, e *false* representa a probabilidade de que comodidade não esteja presente.

Para visualizar os estados de uma variável no UnBBayes no momento em que ela é criada, basta clicar na variável (nó) e a tela é atualizada, apresentando os estados daquela variável. Quando uma variável é criada automaticamente o software atribui um estado a ela, chamado Estado\_0. Para a variável comodidade foi renomeado o estado para *true*, e adicionado o estado *false*, para isso basta clicar no botão com símbolo de adição que fica logo após o campo Nome, em seguida clicar em cima do estado que foi criado e renomeá-lo conforme necessário, como é apresentado na Figura 29.

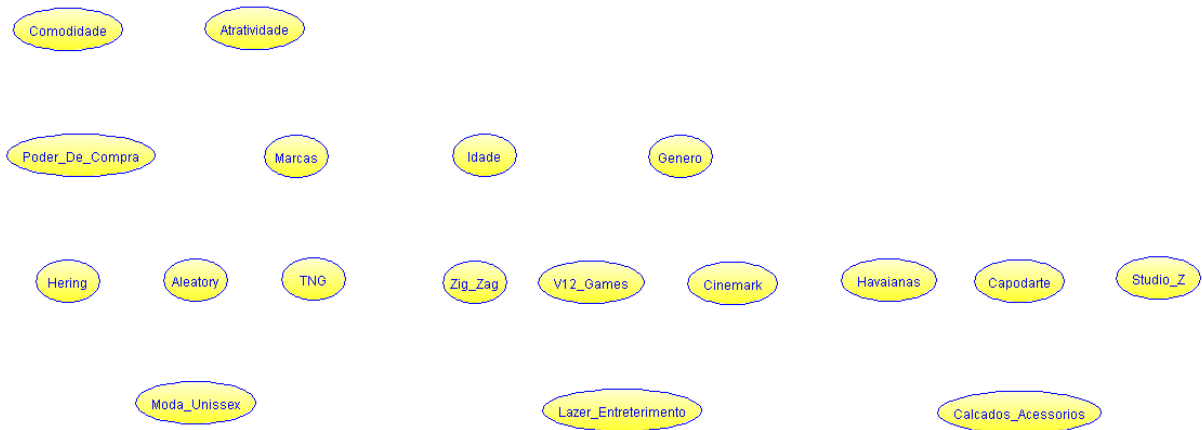
Figura 29: Adicionando e renomeando estados no UnBBayes



Fonte: Elaborada pelo autor

Na Figura 29 é apresentado a variável Comodidade, e seus estados *true* e *false*. Conforme a primeira etapa para o desenvolvimento de uma Rede Bayesiana, apresentada no início da seção 3.2, é necessário definir todo o conjunto de variáveis e seus possíveis estados e valores. Da mesma forma como foi feito para a variável Comodidade, todas as outras variáveis da rede são inseridas, fazendo a rede apresentar o formato da Figura 30.

Figura 30: Variáveis da Rede Bayesiana



Fonte: Elaborada pelo autor

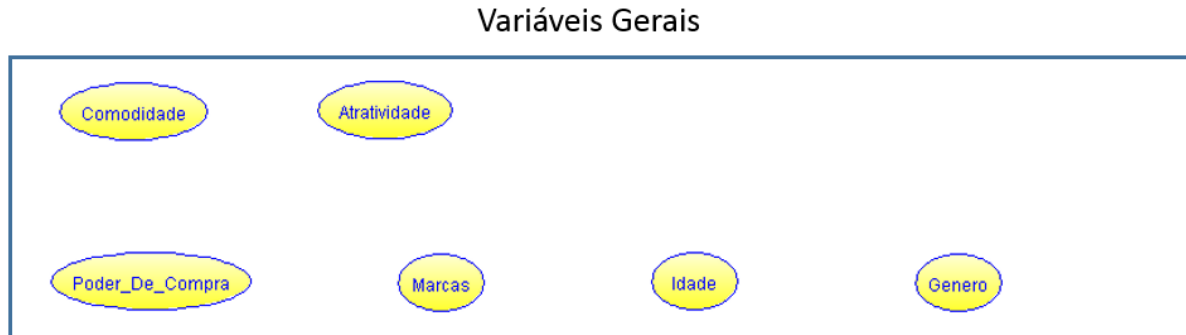
A variável Atratividade junto com as variáveis que representam as categorias (Moda\_Unisex, Lazer\_Entreterimento e Calçados\_Acessorios) e as variáveis que representam as lojas (Hering, Aleatory, TNG, Zig\_Zag, V12\_Games, Cinemark, Havaianas, Capodarte e Studio\_Z) possuem os mesmos estados *true* e *false* que a variável Comodidade. Isto representa a probabilidade de que cada uma destas variáveis seja verdadeiro ou falso. Com as informações obtidas na pesquisa apresentada na seção 3.2 os estados da variável Poder\_de\_Compra foram definidos como ClasseAB e ClasseCDE, de acordo com a distribuição dos usuários que participaram da pesquisa e dividindo-os nestes dois grupos.

A variável Marcas tem os estados grife e popular, que representam a influência que um determinado tipo de bem pode ter em relação a loja que visita. A variável Idade tem os estados Ate34 e 35+, que foram obtidos separando em dois grupos o conjunto de todas as idades relacionadas pelo IBGE, e gênero possui os estados masculino e feminino.

Após definir as variáveis e os estados de cada uma delas, a próxima etapa é construir a estrutura da rede, conectando as variáveis para formar o grafo. Para identificar as relações entre as variáveis, foi utilizado a relação de Causa e Efeito, onde uma variável qualquer X tenha efeito sobre uma segunda variável Y, e esta variável Y é identificada como Causa, enquanto que a variável X é conhecida como efeito.

Para demonstrar esta relação, pode-se dividir as variáveis em três grupos, o primeiro deles é demonstrado na Figura 31.

Figura 31: Grupo de variáveis gerais



Fonte: Elaborada pelo autor

Este grupo de variáveis busca representar o ambiente do *Shopping* e as pessoas que transitam nele, e estabelece um contexto de significado entre elas no modelo de inferência. Analisando a variável comodidade a partir da definição atribuída a ela na seção 3.1.1, Comodidade busca identificar se o aspecto de comodidade afeta na percepção que o usuário pode ter em relação a alguma loja, assim como a variável Atratividade busca complementar o sentido de Comodidade, adicionando o fato de o quão atrativa uma loja pode ser.

Atratividade pode ser calculada baseada na frequência de pessoas que passa em frente com a quantidade de pessoas que entram na loja. Sendo assim, pode-se estabelecer uma relação entre Comodidade e Atratividade, já que a presença ou ausência de comodidade afeta em quão atraente uma loja pode ser para um cliente. A Figura 32 mostra a relação entre Comodidade e Atratividade.

Figura 32: Relação de Comodidade e Atratividade



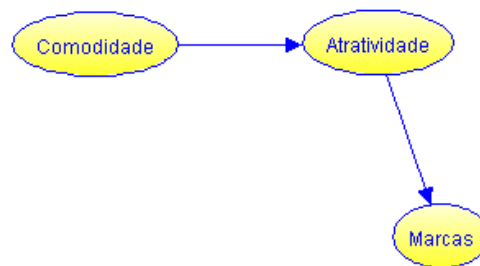
Fonte: Elaborada pelo autor

A Figura 32 mostra a relação de Comodidade e Atratividade, onde Comodidade é um efeito para Atratividade que é uma Causa do efeito anterior. Comodidade também pode ser compreendido como “Pai” de Atratividade, e os valores das probabilidades atribuídas aos estados de Comodidade tem efeito em Atratividade. Além de Atratividade, uma variável que classifica os interesses dos consumidores dentro de um *shopping* é Marcas.



A variável Marcas possui os estados grife e popular, e denota o tipo de item que pode influenciar a escolha de uma loja. Uma vez que a variável Atratividade denota o quão atraente algum lugar pode ser, ela vai influenciar em quais Marcas uma pessoa esteja interessado em consumir, ou seja, a Atratividade de um lugar influencia no desejo de consumir Marcas que sejam de grife ou populares. Esta relação é expressa na Rede Bayesiana atribuindo um arco entre a variável Atratividade, que agora torna-se um efeito de Marcas, que por sua vez é uma Causa da atratividade de algum lugar. A Figura 33 expressa esta relação.

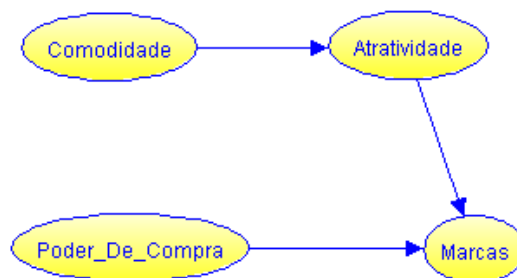
Figura 33:Relação entre as variáveis Atratividade e Marcas



Fonte: Elaborada pelo autor

Da mesma forma como foi a relação de Comodidade e Atratividade, a influência de Atratividade em Marcas é definida pelo arco que sai da variável Atratividade e chega até a variável Marcas. Além de Atratividade, outra variável tem efeito sobre as Marcas que alguém pode estar disposto a adquirir, que é o Poder de Compra. A variável Poder de Compra tem como objetivo refletir o poder econômico dos indivíduos em um *shopping*, e dependendo do volume de recursos disponíveis por alguém, isso influencia em que uma pessoa está disposto a consumir. A relação de Poder de Compra e Marcas é demonstrada na Figura 34.

Figura 34: Relação entre Poder de Compra e Marcas



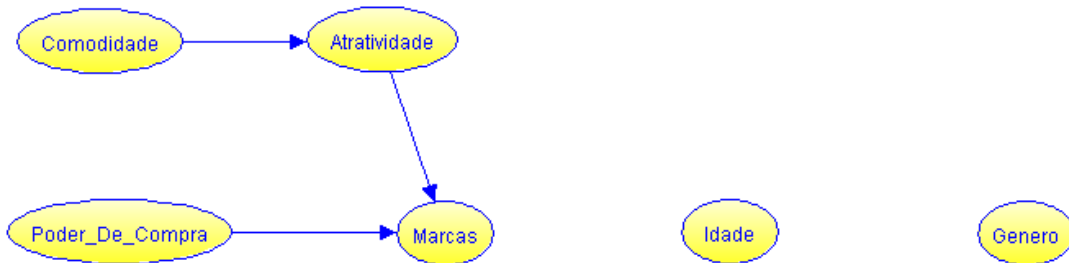
Fonte: Elaborada pelo autor

A Figura 34 mostra a relação entre Poder de Compra e Marcas, além de mostrar que Marcas possui então dois pais, com Atratividade complementando a relação de Marcas e Poder de Compra. Existem mais duas variáveis na rede que possuem a função de representar características das pessoas no modelo de inferência, que são Idade e Gênero. No modelo de

inferência deste trabalho, não há variáveis que tenham algum efeito em Idade e Gênero, dessa forma as duas são representadas sem arcos que liguem alguma variável a elas.

O grupo de variáveis gerais é representado na Figura 35.

Figura 35: Grupo de variáveis gerais com seus relacionamentos



Fonte: Elaborada pelo autor

A Figura 35 mostra todas as variáveis do grupo de variáveis gerais e seus respectivos relacionamentos, onde Marcas é influenciada por Poder de Compra e Atratividade, Atratividade por sua vez influenciada por Comodidade, e por fim as variáveis Comodidade, Poder de Compra, Idade e Gênero que não possuem variáveis que a influenciam, e assim são conhecidas como variáveis raízes.

O próximo grupo de variáveis a ser analisado é o grupo Lojas, que é representado pela Figura 36.

Figura 36: Grupo de variáveis de Lojas

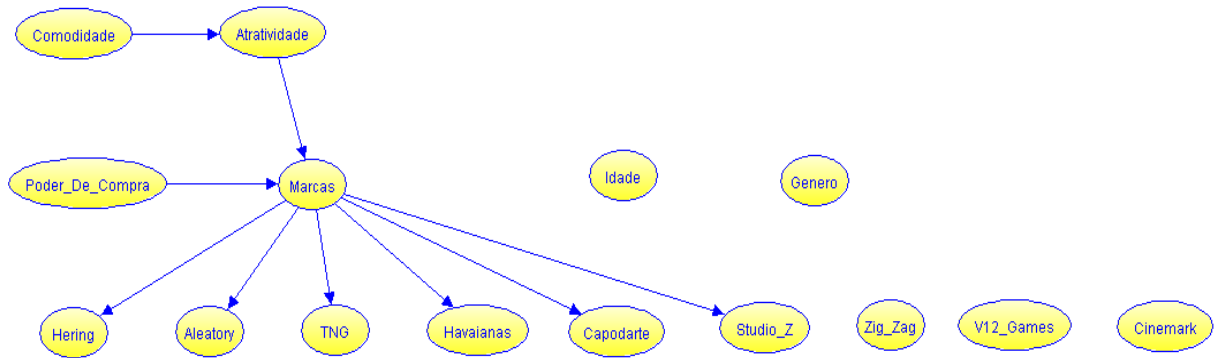
#### Variáveis Lojas



Fonte: Elaborada pelo autor

As variáveis do grupo Lojas representam as lojas que foram selecionadas com base na pesquisa realizada na seção 3.2, e elas sofrem influência direta das variáveis que foram definidas no grupo de variáveis geral. As variáveis Hering, Aleatory, Havaianas, Capodarte e Studio Z lidam com artigos de moda, e acessórios, e por causa destas características pode-se concluir que a variável Marcas tem efeito nestas variáveis de loja, pois é comum que lojas de artigos de moda e acessórios vendam produtos que são considerados "de grife" ou "populares", e esta diferenciação pode provocar ou não o desejo do consumidor, fazendo sentido que Marcas tenha efeito sobre as lojas citadas anteriormente. A Figura 37 mostra a relação de Marcas com as variáveis do grupo de Lojas.

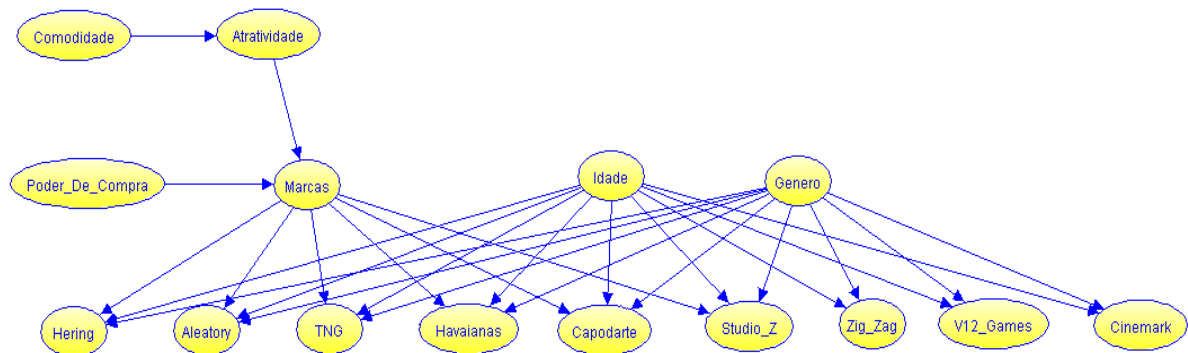
Figura 37: Relacionamento da variável Marcas com as Demais lojas



Fonte: Elaborada pelo autor

Conforme pode ser observado na Figura 37, apenas as variáveis Zig\_Zag e V12\_Games e Cinemark não são afetadas pela variável Marcas. As variáveis Idade e Gênero afetam todas as variáveis de Lojas, ou seja, elas são efeito para todas as lojas da Rede Bayesiana. A Figura 38 mostra a relação de Idade e Gênero com as demais variáveis.

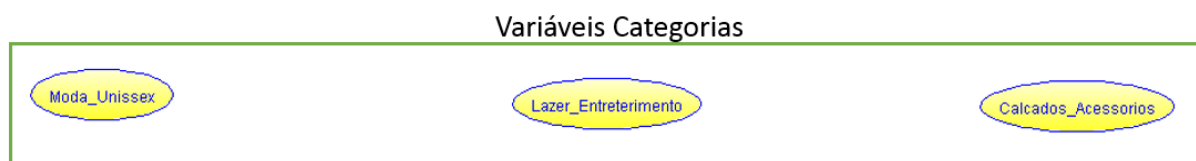
Figura 38: Relação de Idade e Gênero com as demais variáveis



Fonte: UnBBayes

A Figura 38 mostra a relação completa dos dois primeiros grupos de variáveis, o grupo Geral e o grupo Lojas. O último grupo de variáveis da Rede Bayesiana representa as categorias de cada Loja, e pode ser observado na Figura 39.

Figura 39: Grupo de variáveis de categorias

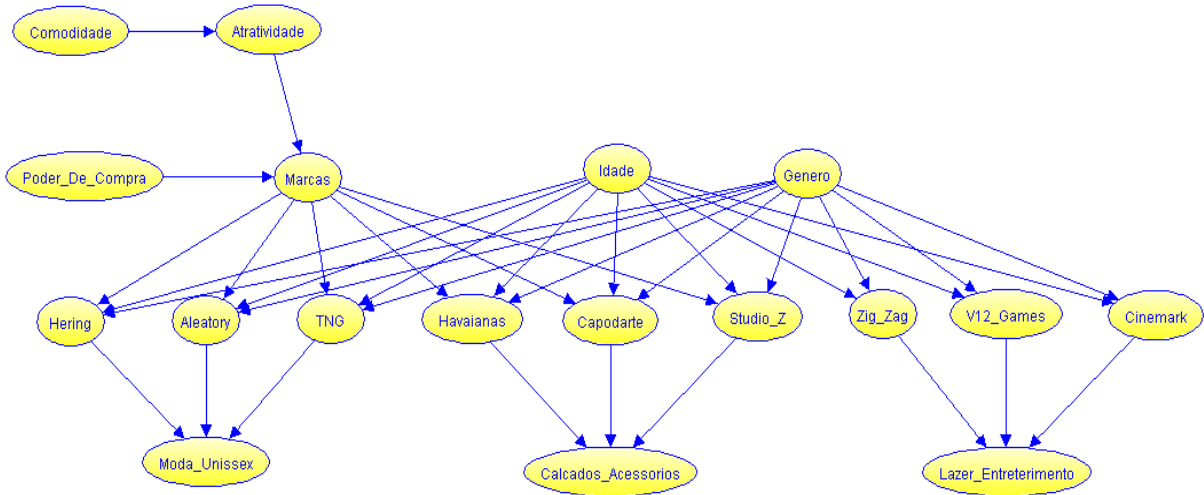


Fonte: UnBBayes

A Figura 39 mostra as variáveis que representam as categorias no modelo de inferência deste trabalho. Elas foram selecionadas com base nas lojas escolhidas na pesquisa da seção 3.2,

e sendo assim, são afetadas diretamente pelas lojas que comercializam produtos da categoria que a variável representa. A Figura 40 mostra a relação entre as categorias e as lojas.

Figura 40: Rede Bayesiana Completa



Fonte: UnBBayes

A Figura 40 apresenta a Rede Bayesiana completa para o modelo de inferência deste trabalho, e assim conclui a 2ª etapa do processo de construção de uma Rede Bayesiana, que trata dos relacionamentos entre as variáveis. O próximo passo é definir as tabelas de probabilidades condicionais de cada variável.

Conforme as relações entre as variáveis foram sendo definidas, o software UnBBayes vai adequando a tabela de probabilidade condicional (TPC) para cada nó, a Figura 41 mostra a TPC da variável Moda\_Unisex.

Figura 41: TPC da variável Moda\_Unisex

Nome: Moda_Unisex		Descrição							
Tabela		true				false			
Completar com distribuição uniforme		Aleatory		Hering		TNG		Havaianas	
Aplicar		true	false	true	false	true	false	true	false
Moda_Unisex	true	1	1	1	1	1	1	1	1
Moda_Unisex	false	0	0	0	0	0	0	0	0

Fonte: UnBBayes

A Figura 41 mostra a TPC da variável Moda\_Unisex, que consiste na relação de todos os estados dos nós Pais para cada um dos estados da variável Moda\_Unisex. Para saber a quantidade de entradas necessária para se preencher a TPC basta multiplicar todos os estados das variáveis envolvidas. Para a TPC da Figura 41 são necessários  $2 \times 2 \times 2 \times 2 = 16$  entradas na TPC. A soma de cada coluna deve resultar em 1, conforme visto na seção 1.1.2.

Para o preenchimento da TPC de todas as variáveis da Rede foram utilizados dados de probabilidade extraídos do site do IBGE, como a proporção de pessoas de até 34 anos e com 35 anos ou mais, além de dados que puderam ser obtidos através da pesquisa realizada na seção 3.2 uma vez que ela possui informações diretamente de pessoas que também são consumidoras em *shoppings centers*. Para definir a TPC basta clicar em cima do nó e em seguida editar a probabilidade para um estado desejado. A Figura 42 mostra a TPC com as probabilidades inseridas para a variável Comodidade.

Figura 42: TPC da variável Comodidade com os valores

State	Probability
true	0,33
false	0,67

Fonte: UnBBayes

A Figura 42 mostra que a probabilidade *a priori* de se ter Comodidade é de  $0,33 * 100\% = 33\%$ .

Na variável Atratividade a TPC foi preenchida com 0.5 para cada estado, devido ao fato de que não se sabe qual o nível de atratividade de uma loja no *shopping* até o desenvolvimento deste trabalho, e assim atribui-se uma distribuição igual de probabilidades entre os estados de Atratividade. Exceto as variáveis Poder de Compra, Comodidade, Idade e Gênero, que tiveram suas probabilidades *a priori* definidas sem a necessidade de cálculos condicionais, todas as demais variáveis tiveram seus estados definidos por meio da fórmula da probabilidade condicional, que foi vista na seção 1.1.4 e tem a seguinte forma:

$$P(A \wedge B) = P(A) \times P(B | A)$$

Por exemplo, para identificar a probabilidade do estado “grife” para a variável “Marcas”, é necessário calcular a probabilidade condicional dos eventos dependentes para os estados de condicionamento necessários  $P(\text{Poder\_de\_Compra} = \text{ClasseAB})$ ,  $P(\text{Atratividade} = \text{true})$  e por fim  $P(\text{Marcas} = \text{grife})$ . Este cálculo pode ser observado da seguinte forma:

$$\text{População} = 113$$

[4]

$$P(\text{Poder de Compra} = \text{ClasseAB}) = 0,78 * 113 = 88,14$$

$$P(\text{Atratividade} = \text{true}) = 0,50 \rightarrow 0,50 * 112 = 56$$

$$P(\text{ClasseAB} \cap \text{true} \cap \text{grife}) = \frac{88,14}{113} * \frac{56}{112} = 0,39$$

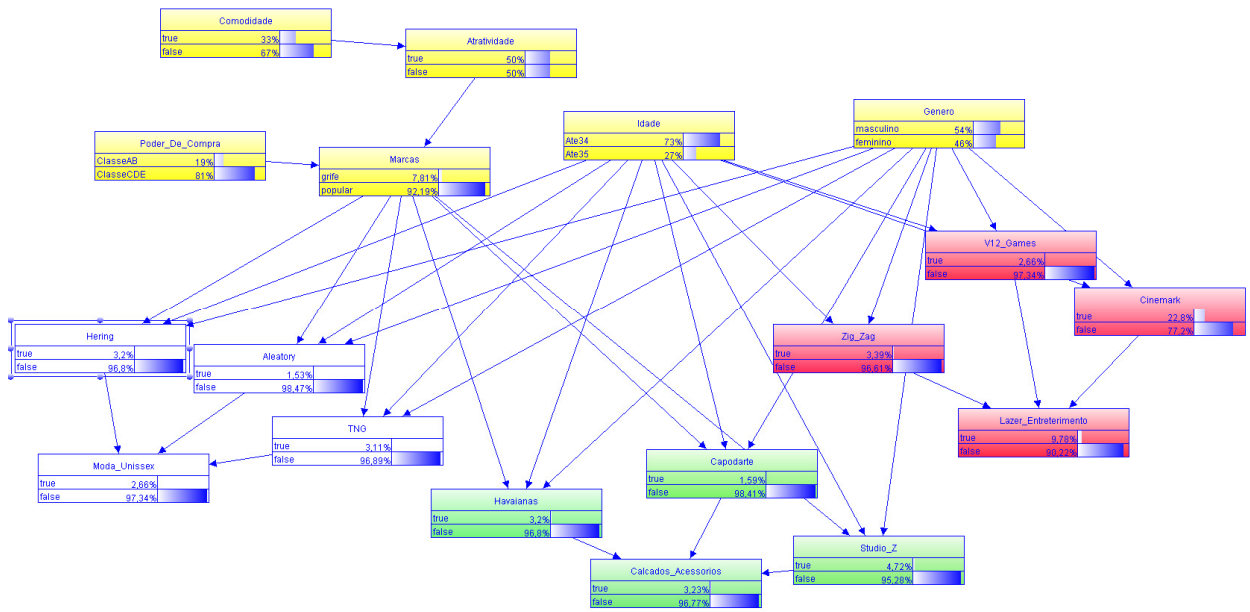
Para realizar o cálculo [4] primeiro foi necessário definir a população em que seria aplicada a análise, neste caso corresponde a 113 pessoas, que foi o volume de participantes da pesquisa. Sabendo que a probabilidade de grife está condicionada a Poder de compra e Atratividade, nos estados (ClasseAB, true), é necessário conhecer quantas pessoas cada um destes estados representa dentro da população de 113 pessoas. Segundo pesquisa realizada pelo site Gazeta Online (2014, online), 78% das vendas de grife estão distribuídas entre as Classes A e B, e o restante, 22%, para as classes C, D e E. Com esta informação, sabe-se que da população de 113 pessoas sabendo que ClasseAB é verdadeiro, então  $\frac{88,14}{113}$  representa a quantidade de pessoas que comprariam grife sabendo que ClasseAB é verdade, enquanto que  $\frac{56}{112}$  representa quantas destas pessoas que são ClasseAB, estão inclinadas a atratividade da loja.

O cálculo [4] foi repetido para todos os estados de todas as variáveis para completar a CPT de todas elas, e isto finaliza as 3 etapas necessárias para a criação de uma Rede Bayesiana.

### 3.4. Aplicação de cenários

Nesta seção serão apresentados alguns cenários que foram aplicadas na Rede Bayesiana, que teve como objetivo observar o comportamento da Rede de acordo com mudanças nos valores das variáveis. Quando a Rede Bayesiana está completa, ou seja, com as variáveis, estados, e probabilidades iniciais definidas, o software UnBBayes permite que o usuário acesse o modo de Compilação, que permite alterar as variáveis e observar a propagação das evidências na rede. Para entrar em modo de compilação basta clicar no botão “Compilar Árvore de Junção”. Quando compilada, a rede apresenta gráficos de barra que indica os valores dos estados das variáveis, estes valores podem ser alterados e assim propagar a alteração por todos os nós, a Figura 43 mostra a Rede criada na seção 3.3 em modo de Compilação.

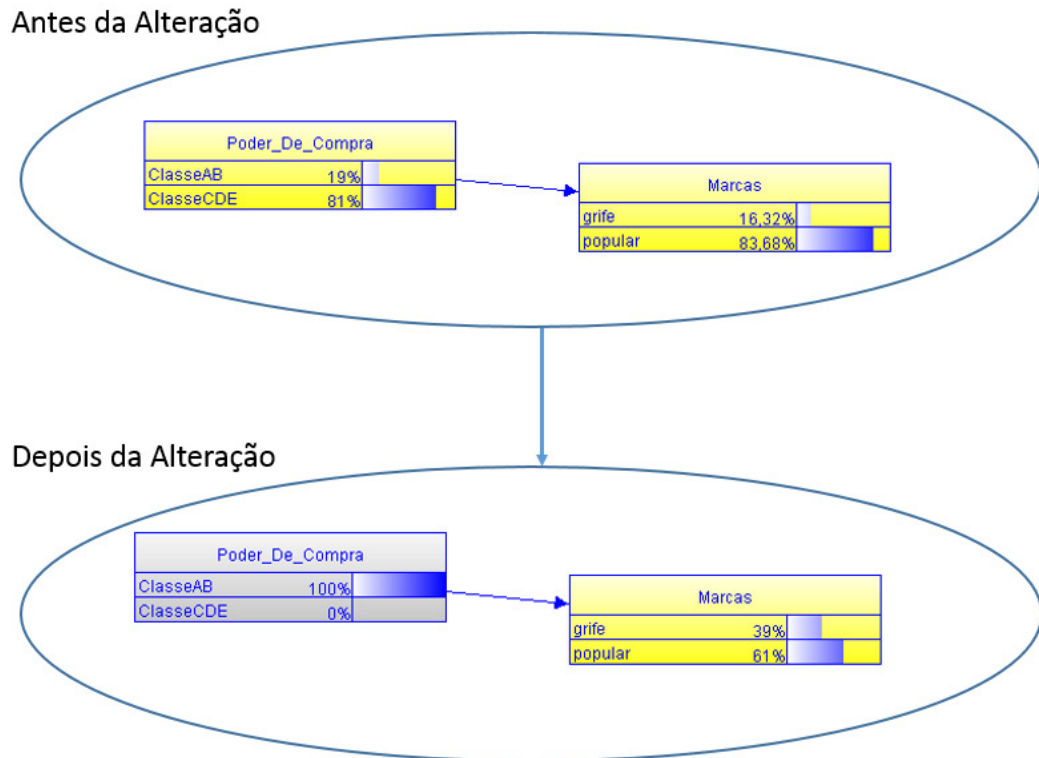
Figura 43: Rede da Seção 3.2 Compilada



Fonte: UnBBayes

As variáveis na cor amarela representam as variáveis do ambiente, as variáveis em branco são relacionadas a categoria Moda Unisex, em verde a categoria Calçados e Acessórios, e em vermelho estão associadas a Lazer e Entretenimento. As barras com as probabilidades representam o quão possível é cada estado, em razão as probabilidades iniciais. Para alterar o estado de uma variável, basta clicar sobre a barra do estado em que se deseja alterar e incrementar ou decrementar o valor, em seguida clicar no botão “Propagar as Evidências”. A Figura 44 mostra o que acontece quando altera-se a probabilidade do estado ClasseAB da variável Poder de Compra.

Figura 44: Alterando a probabilidade do Estado ClasseAB



Fonte: UnBBayes

A Figura 44 mostra que antes da alteração, a probabilidade para o estado ClasseAB era 19%, e a probabilidade para grife era 16,32%. Após alterar a probabilidade de ClasseAB para 100% e clicar no botão de “Propagar evidências”, o estado da rede muda fazendo com que o estado grife assuma 39% de probabilidades, baseado na informação de que ClasseAB é 100%, ou seja, mais que dobraram as chances de interesse em grife condicionado ao fato de que ClasseAB é 100%. Como a probabilidade da variável Marcas aumentou, todas as variáveis que são afetadas por Marcas também tiveram alterações em seus estados.

Baseado nas informações coletadas na pesquisa da seção 3.1, dois cenários foram montados para observar o comportamento da Rede. Para cada cenário, também foi elaborado uma variação do mesmo, utilizando as variáveis abordadas nele mas com valores diferentes para analisar a mudança que isto acarreta em relação ao cenário proposto. A Tabela 6 mostra os cenários e apresenta as informações consideradas em cada um deles.

Tabela 6: Cenários aplicados na Rede Bayesiana

Nome do Cenário	Informações
Cenário 1	<ol style="list-style-type: none"> <li>1. Prefere Comodidade: Sim;</li> <li>2. Renda: Até R\$ 2.899,99; e</li> <li>3. Gênero: Masculino</li> </ol>



Variação do Cenário 1	1. Gênero: Feminino
Cenário 2	1. Gênero: Feminino; 2. Renda: R\$ 14,499,99; e 3. Idade: Maior que 35 anos.
Variação do Cenário 2	1. Idade: Menos que 35 anos.

A Tabela 6 apresenta os dois cenários e suas variações, sendo que os cenários foram elaborados considerando dados encontrados na Pesquisa *Online*, ou seja, houve registros de usuários que informaram os dados apresentados em cada um dos cenários. Os cenários são apresentados as seções a seguir.

#### 3.4.1. Cenário 1

Na pesquisa online, um dos participantes selecionou o seguinte conjunto de respostas:

2. Prefere Comodidade: Sim;
3. Renda: Até R\$ 2.899,99; e
4. Gênero: Masculino

Com base nestas informações, pergunta-se: “Como se comporta a Rede baseada nas características acima?”. Para responder a pergunta, primeiro é necessário informar os valores para as variáveis Comodidade, Renda e Gênero, e em seguida acionar o botão de propagação de evidências. A Tabela 7 mostra as variáveis com seus estados alterados de acordo com as informações anteriores, e depois da propagação.

Tabela 7: Probabilidades do cenário 1

Nome da Variável	Estados Antes	Estados Depois
<b>Comodidade</b>	<i>True</i> – 33%	<i>True</i> – 100%
	<i>False</i> - 67%	<i>False</i> - 0%
<b>Atratividade</b>	<i>True</i> - 50%	<i>True</i> - 50%
	<i>False</i> - 50%	<i>False</i> - 50%
<b>Poder de Compra</b>	ClasseAB - 19%	ClasseAB - 0%
	ClasseCDE - 81%	ClasseCDE – 100%
<b>Marcas</b>	Grife – 16,32%	Grife - 11%

	Popular – 83,66%	Popular - 89%
<b>Gênero</b>	Masculino - 54%	Masculino - 100%
	Feminino - 46%	Feminino - 0%
<b>Hering</b>	<i>True</i> – 9,07%	<i>True</i> – 4,54%
	<i>False</i> – 90,93%	<i>False</i> – 95,45%
<b>Aleatory</b>	<i>True</i> – 8,93%	<i>True</i> – 4,11%
	<i>False</i> – 91,07%	<i>False</i> – 95,89%
<b>TNG</b>	<i>True</i> – 8,67%	<i>True</i> - 0%
	<i>False</i> – 91,33%	<i>False</i> - 100%
<b>Havaianas</b>	<i>True</i> – 8,95%	<i>True</i> - 3,19%
	<i>False</i> – 91,05%	<i>False</i> – 96,81%
<b>Capodarte</b>	<i>True</i> – 8,67%	<i>True</i> - 0%
	<i>False</i> – 91,33%	<i>False</i> - 100%
<b>Studio_Z</b>	<i>True</i> – 8,93%	<i>True</i> – 4,11%
	<i>False</i> – 91,07%	<i>False</i> – 95,89%
<b>Zig_Zag</b>	<i>True</i> – 30,66%	<i>True</i> – 38,17%
	<i>False</i> – 69,42%	<i>False</i> – 61,83%
<b>V12_Games</b>	<i>True</i> – 32,71%	<i>True</i> – 60,58%
	<i>False</i> – 67,29%	<i>False</i> – 39,42%
<b>Cinemark</b>	<i>True</i> – 27,93%	<i>True</i> – 33,25%
	<i>False</i> – 72,07%	<i>False</i> – 66,75%
<b>Moda_Unisex</b>	<i>True</i> – 8,98%	<i>True</i> – 2,94%
	<i>False</i> – 91,02%	<i>False</i> – 97,06%
<b>Calçados e Acessórios</b>	<i>True</i> – 8,94%	<i>True</i> – 2,48%
	<i>False</i> – 91,06%	<i>False</i> – 97,52%
<b>Lazer_Entretenimento</b>	<i>True</i> – 30,51%	<i>True</i> – 44,03%
	<i>False</i> – 69,49%	<i>False</i> – 55,97%

Conforme mostra a Tabela 7, as principais mudanças foram nas probabilidades referentes às categorias de interesse, onde Lazer e Entretenimento tiveram um aumento significativo, passando de 30,51% para 44,03%, enquanto que Calçados e Moda unisex tiveram uma redução, passando de 8,94% e 8,98% para 2,48% e 2,94% respectivamente. Todas as lojas que estão relacionadas com a categoria de Lazer e Entretenimento também sofreram

mudanças em suas probabilidades, o que é coerente com o fato da probabilidade de Lazer e Entretenimento ter aumentado, já que a categoria sofre ação direta de seus pais, que no caso são Zig e Zag, Cinemark e V12 games.

Algumas probabilidades não sofreram nenhum efeito, que foi o caso de Atratividade e Idade. A probabilidade da variável Idade é independente da probabilidade da variável Comodidade, e sem haver qualquer tipo de ligação entre Atratividade e Idade, a regra de que variáveis independentes não se afetam é apresentada na Rede. Já a variável Atratividade tem como pai a variável Comodidade, e mesmo assim permaneceu inalterada. Isto ocorreu por que as probabilidades para Atratividade foram definidas como 50% em todos os estados, ou seja, Comodidade afeta nas probabilidades de Atratividade, mas a definição de 50% na tabela de probabilidade condicional para todos os estados de Atratividade faz com que a alteração não seja visível.

Com algumas pequenas alterações neste cenário é possível visualizar uma saída completamente diferente nas probabilidades para as variáveis. Por exemplo, em vez de trabalhar com gênero masculino, pode-se alterar para feminino e observar como esta informação é propagada na Rede, mas mantendo-se os estados de Renda até R\$ 2.899,99 e Comodidade. A Tabela 8 mostra como ficam as probabilidades após a propagação da evidencia de gênero feminino.

Tabela 8: Probabilidades do cenário 1 com gênero feminino

Nome da Variável	Estados Antes	Estados Depois
<b>Comodidade</b>	<i>True</i> – 33%	<i>True</i> – 100%
	<i>False</i> - 67%	<i>False</i> - 0%
<b>Atratividade</b>	<i>True</i> - 50%	<i>True</i> - 50%
	<i>False</i> - 50%	<i>False</i> - 50%
<b>Poder de Compra</b>	ClasseAB - 19%	ClasseAB - 0%
	ClasseCDE - 81%	ClasseCDE – 100%
<b>Marcas</b>	Grife – 16,32%	Grife - 11%
	Popular – 83,66%	Popular - 89%
<b>Gênero</b>	Masculino - 54%	Masculino - 0%
	Feminino - 46%	Feminino - 10%
<b>Hering</b>	<i>True</i> – 9,07%	<i>True</i> – 12,5%
	<i>False</i> – 90,93%	<i>False</i> – 87,5%
<b>Aleatory</b>	<i>True</i> – 8,93%	<i>True</i> – 12,69%

	<i>False – 91,07%</i>	<i>False – 87,31%</i>
<b>TNG</b>	<i>True – 8,67%</i>	<i>True – 17,04%</i>
	<i>False – 91,33%</i>	<i>False – 82,96%</i>
<b>Havaianas</b>	<i>True – 8,95%</i>	<i>True – 13,85%</i>
	<i>False – 91,05%</i>	<i>False – 86,15%</i>
<b>Capodarte</b>	<i>True – 8,67%</i>	<i>True – 17,04%</i>
	<i>False – 91,33%</i>	<i>False – 82,96%</i>
<b>Studio_Z</b>	<i>True – 8,93%</i>	<i>True – 12,69%</i>
	<i>False – 91,07%</i>	<i>False – 87,31%</i>
<b>Zig_Zag</b>	<i>True – 30,66%</i>	<i>True – 21,68%</i>
	<i>False – 69,42%</i>	<i>False – 78,32%</i>
<b>V12_Games</b>	<i>True – 32,71%</i>	<i>True – 0%</i>
	<i>False – 67,29%</i>	<i>False – 100%</i>
<b>Cinemark</b>	<i>True – 27,93%</i>	<i>True – 21,68%</i>
	<i>False – 72,07%</i>	<i>False – 78,32%</i>
<b>Moda_Unisex</b>	<i>True – 8,98%</i>	<i>True – 14,21%</i>
	<i>False – 91,02%</i>	<i>False – 85,79%</i>
<b>Calçados e Acessórios</b>	<i>True – 8,94%</i>	<i>True – 14,66%</i>
	<i>False – 91,06%</i>	<i>False – 85,34%</i>
<b>Lazer_Entretenimento</b>	<i>True – 30,51%</i>	<i>True – 14,64%</i>
	<i>False – 69,49%</i>	<i>False – 85,36%</i>

Diferente de quanto o gênero é masculino, com gênero feminino as probabilidades para Lazer e Entretenimento reduziram substancialmente, enquanto que Moda Unisex e Calçados e Acessórios cresceram. A probabilidade para Lazer e Entretenimento que para o gênero masculino eram de 44,03%, caíram para 14,64%. Esta mudança está condicionada ao fato de que as lojas Zig Zag, e V12 Games reduziram drasticamente suas probabilidades, passando de 38,17% e 60,58% respectivamente quando o gênero é masculino, para 21,68% e 0% respectivamente quando o gênero é feminino. Neste caso a variável Cinemark continua em ambos os gêneros mantendo uma grande participação na probabilidade de interesse pela categoria de Lazer em ambos os gêneros.

A variável de loja TNG também sofre um grande aumento, passando de 0% quando o gênero é masculino, para 17,04% quando o estado é feminino. Esta baixa probabilidade quando

o status de Gênero é masculino se deve ao fato de que não foram registradas respostas que levassem à loja TNG quando os usuários também definiram ser do sexo masculino, enquanto que grande parte das mulheres relacionaram TNG como uma de suas 5 lojas.

Outra possível alteração deste cenário é mudar o poder de compra o usuário, de ClasseCDE, que corresponde a faixa de R\$ 2.899,99, para ClasseAB. A Tabela 9 mostra os efeitos da alteração do estado de Poder de Compra, mas mantendo os valores originais de Comodidade e Sexo Masculino.

Tabela 9: Alteração de Poder de Compra para ClasseAB no cenário 1

Nome da Variável	Estados Antes	Estados Depois
<b>Comodidade</b>	<i>True</i> – 33%	<i>True</i> – 100%
	<i>False</i> - 67%	<i>False</i> - 0%
<b>Atratividade</b>	<i>True</i> - 50%	<i>True</i> - 50%
	<i>False</i> - 50%	<i>False</i> - 50%
<b>Poder de Compra</b>	ClasseAB - 19%	ClasseAB - 0%
	ClasseCDE - 81%	ClasseCDE – 100%
<b>Marcas</b>	Grife – 16,32%	Grife - 39%
	Popular – 83,66%	Popular - 61%
<b>Gênero</b>	Masculino - 54%	Masculino - 100%
	Feminino - 46%	Feminino - 0%
<b>Hering</b>	<i>True</i> – 9,07%	<i>True</i> – 7,3%
	<i>False</i> – 90,93%	<i>False</i> – 92,7%
<b>Aleatory</b>	<i>True</i> – 8,93%	<i>True</i> – 6,46%
	<i>False</i> – 91,07%	<i>False</i> – 93,54%
<b>TNG</b>	<i>True</i> – 8,67%	<i>True</i> - 0%
	<i>False</i> – 91,33%	<i>False</i> - 100%
<b>Havaianas</b>	<i>True</i> – 8,95%	<i>True</i> – 5,05%
	<i>False</i> – 91,05%	<i>False</i> – 94,95%
<b>Capodarte</b>	<i>True</i> – 8,67%	<i>True</i> - 0%
	<i>False</i> – 91,33%	<i>False</i> - 100%
<b>Studio_Z</b>	<i>True</i> – 8,93%	<i>True</i> – 6,46%
	<i>False</i> – 91,07%	<i>False</i> – 93,54%
<b>Zig_Zag</b>	<i>True</i> – 30,66%	<i>True</i> – 38,17%
	<i>False</i> – 69,42%	<i>False</i> – 61,83%

<b>V12_Games</b>	<i>True – 32,71%</i>	<i>True – 60,58%</i>
	<i>False – 67,29%</i>	<i>False – 39,42%</i>
<b>Cinemark</b>	<i>True – 27,93%</i>	<i>True – 33,25%</i>
	<i>False – 72,07%</i>	<i>False – 66,75%</i>
<b>Moda_Unisex</b>	<i>True – 8,98%</i>	<i>True – 4,66%</i>
	<i>False – 91,02%</i>	<i>False – 95,34%</i>
<b>Calçados e Acessórios</b>	<i>True – 8,94%</i>	<i>True – 3,9%</i>
	<i>False – 91,06%</i>	<i>False – 96,1%</i>
<b>Lazer_Entretenimento</b>	<i>True – 30,51%</i>	<i>True – 44,03%</i>
	<i>False – 69,49%</i>	<i>False – 55,97%</i>

Como pode ser observado na Tabela 9, a principal mudança foi na relação de Poder de Compra e Marcas, em Marcas passou de 16,32% no estado grife, para 39%, e esta alteração impactou todas as variáveis que tem como pai Marcas. A probabilidade de um indivíduo do sexo masculino, que ganha acima de R\$ 2.899,99 que seja afetado por Comodidade tenha interesse na loja Hering é de 7,3%, esta probabilidade é maior que a encontrada anteriormente para quando o estado do Poder de Compra era ClasseCDE, que naquele cenário era de 4,54%, conforme mostra a Tabela 7.

### 3.4.2. Cenário 2

O próximo cenário da rede envolve as seguintes escolhas:

2. Gênero: Feminino;
3. Renda: R\$ 14,499,99; e
4. Idade: Maior que 35 anos.

A Tabela 10 mostra como ficam as probabilidades para a rede de acordo com cenário

2.

Tabela 10: Probabilidades do cenário 2

Nome da Variável	Estados Antes	Estados Depois
<b>Comodidade</b>	<i>True – 33%</i>	<i>True – 100%</i>
	<i>False - 67%</i>	<i>False - 0%</i>
<b>Atratividade</b>	<i>True - 50%</i>	<i>True - 50%</i>
	<i>False - 50%</i>	<i>False - 50%</i>
<b>Poder de Compra</b>	ClasseAB - 19%	ClasseAB - 0%

	ClasseCDE - 81%	ClasseCDE – 100%
<b>Marcas</b>	Grife – 16,32%	Grife - 39%
	Popular – 83,66%	Popular - 61%
<b>Gênero</b>	Masculino - 54%	Masculino - 100%
	Feminino - 46%	Feminino - 0%
<b>Hering</b>	<i>True</i> – 9,07%	<i>True</i> – 7,3%
	<i>False</i> – 90,93%	<i>False</i> – 92,7%
<b>Aleatory</b>	<i>True</i> – 8,93%	<i>True</i> – 6,46%
	<i>False</i> – 91,07%	<i>False</i> – 93,54%
<b>TNG</b>	<i>True</i> – 8,67%	<i>True</i> - 0%
	<i>False</i> – 91,33%	<i>False</i> - 100%
<b>Havaianas</b>	<i>True</i> – 8,95%	<i>True</i> – 5,05%
	<i>False</i> – 91,05%	<i>False</i> – 94,95%
<b>Capodarte</b>	<i>True</i> – 8,67%	<i>True</i> - 0%
	<i>False</i> – 91,33%	<i>False</i> - 100%
<b>Studio_Z</b>	<i>True</i> – 8,93%	<i>True</i> – 6,46%
	<i>False</i> – 91,07%	<i>False</i> – 93,54%
<b>Zig_Zag</b>	<i>True</i> – 30,66%	<i>True</i> – 38,17%
	<i>False</i> – 69,42%	<i>False</i> – 61,83%
<b>V12_Games</b>	<i>True</i> – 32,71%	<i>True</i> – 60,58%
	<i>False</i> – 67,29%	<i>False</i> – 39,42%
<b>Cinemark</b>	<i>True</i> – 27,93%	<i>True</i> – 33,25%
	<i>False</i> – 72,07%	<i>False</i> – 66,75%
<b>Moda_Unisex</b>	<i>True</i> – 8,98%	<i>True</i> – 4,66%
	<i>False</i> – 91,02%	<i>False</i> – 95,34%
<b>Calçados e Acessórios</b>	<i>True</i> – 8,94%	<i>True</i> – 3,9%
	<i>False</i> – 91,06%	<i>False</i> – 96,1%
<b>Lazer_Entretenimento</b>	<i>True</i> – 30,51%	<i>True</i> – 44,03%
	<i>False</i> – 69,49%	<i>False</i> – 55,97%

As probabilidades para Moda Unisex, Calçados e Acessórios e Lazer e Entretenimento na Tabela 10 apresentam algumas semelhanças mesmo se comparadas com a Tabela 8, que também associava o gênero feminino, com diferença de que na Tabela 10 o estado da variável

Poder de Compra é ClasseAB. Na Tabela 10 as probabilidades de Hering, Aleatory e Havaianas são de 8,29%, 8,68% e 9,68% respectivamente, enquanto que na Tabela 8 elas são de 12,5%, 12,69% e 13,85% respectivamente. Esta diferença entre as duas tabelas representa o impacto que Idade tem nas variáveis da rede, em que o estado Ate34 afeta muito mais as variáveis de lojas do que o estado 35+.

Caso se altere o estado da variável Idade para Ate34, os valores são bem diferentes, neste caso as probabilidades para um indivíduo de até 34 anos, do gênero feminino e com Poder de compra de R\$ 14,499,99 faz com que exista uma probabilidade de 26,58% de estar interessado em Moda Unisex, 27,43% em Calçados e Acessórios, e 17,54% de interesse em Lazer e Entretenimento, contra 9,73%, 10,20% e 6,78% respectivamente quando o estado da variável Idade é Mais35.

Este capítulo apresentou os resultados da pesquisa a partir dos critérios de desenvolvimento elaborados na seção 2.3. A importância do desenvolvimento esteve relacionada em conseguir aplicar os conceitos de Redes Bayesianas para estimar a probabilidade de interesse do indivíduo, e as etapas desenvolvidas nas seções anteriores tiveram o objetivo de ir evoluindo o modelo até chegar a etapa de aplicação de cenários, que mostrou como os dados se propagam de acordo com alterações nas variáveis.



## 5. CONSIDERAÇÕES FINAIS

O presente trabalho buscou desenvolver um modelo de inferência para identificar o quanto interessado o cliente pode estar em uma categoria de produto a partir da modelagem de uma Rede Bayesiana com as características do consumidor. Para alcançar o objetivo uma série de etapas foram necessárias ao longo da pesquisa como descrito na seção 2.3, e ao fim da execução de cada uma destas etapas algumas decisões foram sendo tomadas a medida que novos resultados surgiram.

O desenvolvimento da pesquisa começa com um levantamento sobre o comportamento do consumidor em *shoppings*, em busca de características do consumidor observadas nestes ambientes onde vale destacar alguns trabalhos relacionados na seção 1.3, que contribuíram para a pesquisa no sentido de dar uma visão mais clara sobre os pontos que atualmente já são observados na interação do consumidor com o *shopping*. Conhecer estas características foi fundamental para a tese central da pesquisa, pois limitou o espaço sobre quais pontos a Rede Bayesiana deveria contemplar para chegar ao resultado desejado, uma vez que analisar um ambiente e o relacionamento de um indivíduo com ele é uma tarefa bastante complexa que demandaria diversos outros recursos para se realizar.

Uma das principais decisões da pesquisa foi onde conseguir os dados para calcular a Rede Bayesiana. Algumas abordagens foram consideradas como por exemplo utilizar softwares que informassem o percurso de uma pessoa em um *shopping*, sendo que este software deveria estar instalado no dispositivo móvel do usuário, entretanto nenhuma solução neste sentido foi encontrada de uma forma que pudesse ser aplicada ao trabalho. Outras abordagens como captura de imagens por câmeras de vídeo e o posterior tratamento delas já foi empregado em outros trabalhos, mas nesta pesquisa não era possível realizar esta atividade pois demandava diversos recursos e etapas adicionais que não estavam disponíveis ao pesquisador, sendo assim utilizar uma pesquisa *online*, e definir as perguntas que seriam aplicadas de forma que as respostas pudessem ser utilizadas como fonte de dados, foi essencial.

A principal relevância da pesquisa foi demonstrar como Redes Bayesianas podem ser utilizadas para ajudar as empresas a compreender melhor seus clientes, direcionando o investimento da loja para clientes em potencial de acordo com suas características. Neste sentido, o ponto crítico da pesquisa foi a seleção das variáveis e a definição das tabelas de probabilidade condicional, que concentraram a maior parte do esforço no sentido de manter coerentes os dados informados.

Para isso foram necessárias consultas a diversas fontes de dados como periódicos científicos, revistas e pesquisas online, e principalmente, fazer a ligação entre todas estas informações em um modelo computacional que buscou demonstrar um ambiente real de *shopping center*.

Ainda neste processo de definição de variáveis boa parte do tempo foi dedicado em uma dúvida fundamental, que foi “como distribuir as variáveis de forma que esta organização expresse o papel daquela característica no Modelo?”. Neste ponto um critério fundamental foi focar sempre em causa e efeito, pois são estas as bases do relacionamento de variáveis em uma Rede Bayesiana quaisquer, mas que ganham ainda mais sentido quando se busca estabelecer sentido entre uma característica do consumidor, e como esta característica pode levar ou não o consumidor até uma loja.

As mudanças no modelo foram então ocorrendo na medida em que o processo foi evoluindo até se chegar ao modelo final proposto pela pesquisa, pois conforme novas informações foram surgindo, algumas relações deixaram de fazer sentido pois possuíam um baixo nível de influência na rede, e outras foram sendo mais destacadas de acordo com o nível o quanto a variável tem a capacidade de alterar as probabilidades quando alterada.

Não fez parte do escopo do trabalho fazer a validação do modelo proposto, esta validação pode ser empregada em um trabalho futuro, para identificar o nível de assertividade deste modelo em ambientes controlados ou não. Um meio de validar o modelo proposto pelo trabalho é aplica-lo em um cenário real, onde os dados são obtidos através de análise dos consumidores sem que seja necessário alguma interação com eles, utilizando um sistema de câmeras de vídeo, e assim inserir na Rede Bayesiana os dados obtidos e por fim comparar com dados cedidos pelas lojas e pelo *shopping*, e assim comparar os resultados e identificar taxas de variação. O centro da pesquisa são as variáveis que expressão as características dos clientes, mas na pesquisa estas características são aplicadas a todas as lojas, além de serem características gerais de um consumidor que podem ser observadas não apenas em *shoppings*, mas em outros locais de consumo, sendo assim, um trabalho futuro pode ser identificar características que estejam relacionadas não apenas com um consumidor em geral, mas com um tipo específico de consumidor, por exemplo, para uma determinado segmento de produtos como calçados, identificar características específicas do consumidor de calçados, e utilizar eles para a influência das lojas neste segmento.



## 6. REFERÊNCIAS BIBLIOGRÁFICAS

ABF – Associação Brasileira de *Franchising*. *Evolução do Setor de Franchising*. São Paulo(SP); 2013

ABRASCE – Associação Brasileira de *Shopping Centers*. *Números do setor*. São Paulo(SP); 2014 [acesso em 24 de abril de 2014]. Disponível em <http://www.portaldoshopping.com.br/numeros-do-setor/grandes-numeros>

ABRASCE – Associação Brasileira de *Shopping Centers*. *Merchandising em Shopping Centers*. 2008 [acesso em 24 de abril de 2014]. Disponível em <http://www.ibope.com.br/pt-br/conhecimento/relatoriospesquisas/Lists/RelatoriosPesquisaEleitoral/Job%20081043%20-%20Abrasce%20-%20Merchandising%20em%20shopping%20centers.pdf>

CARVALHO, R. **UnBBayes Framework**. [s.l: s.n.]. Disponível em: <<http://pt.slideshare.net/rommelnc/unbbayes-plugin-framework>>.

CHEN, H. *Business and Market Intelligence 2.0*. p. 2–15, 2010.

DARWICHE, A. **Modeling and Reasonin with Bayesian Networks**. [s.l: s.n.]. p. 548

DARWICHE, A. **Modeling and Reasoning with Bauesian Networks**. [s.l.] Cambridge University Press, 2009b. p. 548

DOURADO, C. **Capim Dourado: O Shopping**. Disponível em: <<http://www.capimdouradosc.com.br/portal/pagina-o-shopping.html>>. Acesso em: 3 mar. 2015.

GNERI, M. A. A evolução histórica do conceito de probabilidade 1. In: **A EVOLUÇÃO HISTÓRICA DO CONCEITO DE PROBABILIDADE**. [s.l: s.n.]. p. 1–5.

GRINSTEAD, C. M.; SNELL, J. L. *Introduction to Probability*. [s.d.].

LUGER, G. F. **Inteligência Artificial**. 6° ed ed. São Paulo: Pearson, 2014. p. 613

LUNA, E. O. *Algoritmos EM para Aprendizagem de Redes Bayesianas a partir de Dados Incompletos*. 2004.

MARQUES, R. L.; DUTRA, I. *Redes Bayesianas : o que são , para que servem , algoritmos e exemplos de aplicações* . p. 1–22, 2000.

MATSUMOTO, S. Universidade de Brasília Instituto de Ciências Exatas Departamento de Ciência da Computaç ão Framework para Redes Bayesianas Multi-Entidades e Ontologias Probabilísticas. p. 137, 2008.

MICHAEL, K.; CLARKE, R. Location and tracking of mobile devices: Überveillance stalks the streets. **Computer Law & Security Review**, v. 29, n. 3, p. 216–228, jun. 2013.

MICROSOFT. **SQL Server**. Disponível em: <<https://www.microsoft.com/en-us/server-cloud/products/sql-server/>>. Acesso em: 14 abr. 2015a.

MICROSOFT. **Microsoft**. Disponível em: <<https://www.microsoft.com>>. Acesso em: 14 maio. 2015b.

NEGÓCIOS., F. &. **Manual de operação em shoppings**.

ONLINE, G. **Classe média: Ela quer e compra luxo**. Disponível em: <[http://gazetaonline.globo.com/\\_conteudo/2014/03/noticias/cidades/1480795-classe-media-ela-quer-e-compra-luxo.html](http://gazetaonline.globo.com/_conteudo/2014/03/noticias/cidades/1480795-classe-media-ela-quer-e-compra-luxo.html)>. Acesso em: 15 maio. 2015.

RUSSEL, J. S.; NORVIG, P. **Artificial Intelligence: A modern Approach**. [s.l: s.n.]. p. Prentice Hall 415–429, 436–457

SÁ, I. P. DE. **As três definições de probabilidades**, 2012.

SANDALL, H. L. P. Comportamentos Precorrentes de Aproximação de Lojas : Efeitos do nível de diferenciação de marca e localização sobre taxas de conversão em um shopping center Comportamentos Precorrentes de Aproximação de Lojas : 2007.

TEIXEIRA, D. R. **Redes de Valor e de Cooperação - evoluindo com a Inteligência Empresarial**. Disponível em: <<http://revie.com.br/site-blog?id=29&ano=2009>>.

VIALI, L. Algumas Considerações Sobre a Origem da Teoria da Probabilidade. **Revista Brasileira de História da Algumas Matemática - Vol. 8 n**, v. 8, p. 143–153, 2009.

YAN, P. **SPATIAL-TEMPORAL DATA ANALYTICS AND CONSUMER SHOPPING MODELING**. [s.l: s.n.].

YIM YIU, C.; CHEONG NG, H. Buyers-to-shoppers ratio of shopping malls: A probit study in Hong Kong. **Journal of Retailing and Consumer Services**, v. 17, n. 5, p. 349–354, set. 2010.

