

Universidade Federal de Campina Grande  
Centro de Engenharia Elétrica e Informática  
Programa de Pós-Graduação em Ciência da Computação

Um Estudo Comparativo Entre Técnicas de Geração  
Semiautomática de Tabelas de Probabilidades de  
Nós em Redes Bayesianas

João Batista Nunes Bezerra

Dissertação submetida à Coordenação do Curso de Pós-Graduação em  
Ciência da Computação da Universidade Federal de Campina Grande -  
Campus I como parte dos requisitos necessários para obtenção do grau  
de Mestre em Ciência da Computação.

Área de Concentração: Ciência da Computação

Linha de Pesquisa: Redes Bayesianas

Kyller Costa Gorgônio e Angelo Perkusich

Campina Grande, Paraíba, Brasil

©João Batista Nunes Bezerra, 28/01/2019

B574e

Bezerra, João Batista Nunes.

Um estudo comparativo entre técnicas de geração semiautomática de tabelas de probabilidades de nós em redes bayesianas / João Batista Nunes Bezerra. – Campina Grande, 2019.

121 f. : il. color.

Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Campina Grande, Centro Engenharia Elétrica e Informática, 2019.

"Orientação: Prof. Dr. Kyller Costa Gorgônio, Prof. Dr. Angelo Perkusich".

Referências.

1. Redes Bayesianas. 2. Probabilidade. 3. Tabela de Probabilidade de Nós. I. Gorgônio, Kyller Costa. II. Perkusich, Angelo. III. Título.

CDU 519.226(043)

**“UM ESTUDO COMPARATIVO ENTRE TÉCNICAS DE GERAÇÃO  
SEMIAUTOMÁTICA DE TABELAS DE PROBABILIDADES DE  
NÓS EM REDES BAYESIANAS”**

**JOÃO BATISTA NUNES BEZERRA**

**DISSERTAÇÃO APROVADA EM 28/02/2019**

**KYLLER COSTA GORGÔNIO, Dr., UFCG  
Orientador(a)**

**ANGELO PERKUSICH, Dr., UFCG  
Orientador(a)**

**DANILO FREIRE DE SOUZA SANTOS, Dr., UFCG  
Examinador(a)**

**MIRKO BARBOSA PERKUSICH, Dr., IFPB  
Examinador(a)**

**CAMPINA GRANDE - PB**

## Resumo

Um dos desafios para o uso de redes Bayesianas refere-se à construção das Tabelas de Probabilidade dos Nós (TPN). Dado que a complexidade para definição de TPN cresce exponencialmente, para redes Bayesianas de larga escala torna-se impraticável a definição manual das TPN. Existem diversas técnicas que abordam este problema, dentre elas, métodos de construção semiautomática de TPN. O objetivo deste trabalho é mensurar e comparar a capacidade de modelagem de TPN e acurácia de alguns destes métodos: *Weighted Sum Algorithm* (WSA), *Ranked Nodes Method* (RNM), uma adaptação do *Analytic Hierarchy Process* (AHP) para o contexto de redes Bayesianas, e uma combinação do WSA com o AHP. Dessa forma, um estudo de caso com duas unidades de análise e um experimento controlado foram realizados. No estudo de caso participaram quatro especialistas de domínio e foram construídas quatro redes Bayesianas para estimativa de valor com os métodos WSA e RNM. A acurácia dos métodos foi avaliada com dados históricos e cenários simulados. No que se refere ao experimento controlado, foi adotado um Delineamento em Blocos Casualizados Completo (DBC). Participaram do experimento dez desenvolvedores de software que aplicaram os métodos para quantificar incertezas em uma rede Bayesiana. Estes mesmos desenvolvedores definiram manualmente TPN que serviram como referência para medir a capacidade de modelagem e acurácia dos métodos semiautomáticos. Foram utilizadas as seguintes medidas de similaridade para medir a capacidade de modelagem dos métodos: *Brier Score* (BS), Distância Euclidiana (ED) e Erro Médio Absoluto (EM). Para medir a acurácia foi utilizado a Proporção de Acertos (PA) dos métodos. A principal limitação deste estudo foi o escopo reduzido para quatro métodos, decorrente dos altos custos envolvidos na aplicação dos mesmos. A partir dos resultados obtidos, foi possível concluir que há diferenças estatisticamente significativas entre os métodos e o método de melhor desempenho é o RNM, seguido do WSA, WSA-AHP e AHP.

## **Abstract**

One of the challenges for the use of Bayesian networks is the construction of the Node Probability Tables (NPT). The complexity for defining NPT is exponential, for large-scale Bayesian networks, it is not feasible to manually define NPT. There are several techniques that address this problem, among them, semiautomatic methods for the construction of NPT. The objective of this work is to measure and compare the modelling capability and accuracy of some of these methods: Weighted Sum Algorithm (WSA), Ranked Nodes Method (RNM), an adaptation of the Analytic Hierarchy Process (AHP) to the context of Bayesian networks, and a combination of WSA and AHP. Therefore, a case study with two units of analysis and a controlled experiment was performed. The accuracy of the methods was evaluated using historical data and simulated scenarios. Regarding the controlled experiment, a Randomized Complete Block Design (RCBD) was adopted. Ten software developers participated in the experiment and applied the methods to quantify uncertainties in a Bayesian network. These same developers manually defined NPT that served as a reference to measure the modelling capabilities and accuracy of the methods. The following measures of similarity were used to measure the modelling capacity of the methods: Brier Score (BS), Euclidean Distance (ED) and Mean Absolute Error (MAE). In order to measure the accuracy, the Accuracy Ratio (AR) of the methods was used. The main limitation of this study is its reduced scope of four methods due to the high costs involved in their application. According to the results obtained, it is possible to conclude that there are statistically significant differences between the methods. The method that presented the best result was the RNM, followed by the WSA, WSA-AHP and AHP.

## **Agradecimentos**

Primeiro de tudo, gostaria de agradecer a Deus por me guiar, iluminar e me dar tranquilidade para seguir em frente com os meus objetivos.

Agradeço à minha família pelo apoio irrestrito.

Agradeço aos meus orientadores Kyller e Ângelo, pelos ensinamentos, apoio e incentivo.

Agradeço aos pesquisadores Mirko e Emilia pelo suporte dado.

Agradeço aos amigos do grupo de estudos ISE, pelo aprendizado compartilhado e companhia.

Agradeço à Universidade Federal de Campina Grande, ao Centro de Engenharia Elétrica e Informática, ao Departamento de Sistemas e Computação, ao Programa de Pós-Graduação em Ciência da Computação da UFCG e seus respectivos professores e funcionários que direta ou indiretamente contribuíram na minha formação.

Por fim, agradeço ao apoio financeiro sem o qual não seria possível a dedicação exclusiva ao presente trabalho, o qual foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES).

# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Problemática . . . . .	4
1.2	Objetivo . . . . .	6
1.3	Metodologia . . . . .	7
1.4	Resultados e Contribuições . . . . .	9
1.5	Estrutura da Dissertação . . . . .	10
<b>2</b>	<b>Fundamentação Teórica</b>	<b>11</b>
2.1	Rede Bayesiana . . . . .	11
2.1.1	Teorema de Bayes . . . . .	12
2.1.2	Custo da definição de um TPN . . . . .	13
2.1.3	Propriedades de separação direcional . . . . .	14
2.1.4	Redes Bayesianas em VBSE . . . . .	15
2.2	Métodos de Geração de TPN . . . . .	17
2.2.1	Método WSA . . . . .	17
2.2.2	Método AHP . . . . .	19
2.2.3	Método RNM . . . . .	23
<b>3</b>	<b>Estudo de Caso</b>	<b>30</b>
3.1	Visão Geral . . . . .	30
3.2	Motivação . . . . .	32
3.3	Metodologia . . . . .	32
3.3.1	Framework VALUE . . . . .	33
3.3.2	Contexto . . . . .	34

3.3.3	Pergunta de Pesquisa . . . . .	35
3.3.4	Seleção de Casos . . . . .	35
3.3.5	Seleção de Participantes . . . . .	36
3.3.6	Treinamento . . . . .	40
3.4	Coleta de Dados . . . . .	40
3.4.1	Caso A . . . . .	42
3.4.2	Caso B . . . . .	52
3.5	Resultados e Discussão . . . . .	59
3.5.1	Caso A . . . . .	59
3.5.2	Caso B . . . . .	60
3.6	Ameaças à Validade . . . . .	62
<b>4</b>	<b>Experimento</b>	<b>63</b>
4.1	Definição . . . . .	63
4.1.1	Objetivos específicos . . . . .	63
4.2	Visão Geral . . . . .	64
4.3	Contexto . . . . .	65
4.4	Planejamento . . . . .	65
4.4.1	Questões de pesquisa e hipóteses . . . . .	65
4.4.2	Definição Formal das Hipóteses . . . . .	67
4.4.3	Seleção do Modelo . . . . .	67
4.4.4	Estudo Piloto . . . . .	67
4.4.5	Seleção dos Participantes . . . . .	71
4.4.6	Delineamento do Experimento . . . . .	71
4.4.7	Instrumentação . . . . .	75
4.4.8	Procedimento de Coleta de Dados . . . . .	78
4.4.9	Procedimento de Análise . . . . .	79
4.5	Execução . . . . .	80
4.6	Análise dos Dados . . . . .	84
4.6.1	Análise Descritiva . . . . .	84
4.6.2	Validação do Modelo . . . . .	90



---

4.6.3	Análise da Variância . . . . .	92
4.7	Resultados e Discussão . . . . .	102
4.7.1	Avaliação dos Resultados e Implicações . . . . .	102
4.7.2	Ameaças à Validade . . . . .	103
<b>5</b>	<b>Discussão</b>	<b>105</b>
5.1	Acurácia . . . . .	105
5.2	Capacidade de modelagem . . . . .	107
<b>6</b>	<b>Considerações Finais</b>	<b>109</b>
<b>A</b>	<b>Artefatos Utilizados no Experimento</b>	<b>118</b>

# Lista de Símbolos

TPN - *Tabela de Probabilidade dos Nós*

RNM - *Ranked Nodes Method*

WSA - *Weighted Sum Algorithm*

AHP - *Analytic Hierarchy Process*

BS - *Brier Score*

ED - *Distância Euclidiana*

EM - *Erro Médio Absoluto*

PA - *Proporção de Acertos*

VALUE - *Improving Decision-Making Relating to Software-Intensive Products and Services*

VBSE - *Value-Based Software Engineering*

EKEBN - *Expert-based Knowledge Engineering of Bayesian Network*

# Lista de Figuras

1.1	Exemplo de rede Bayesiana . . . . .	2
1.2	Visão geral da metodologia aplicada . . . . .	8
2.1	Probabilidades calculadas usando o teorema de Bayes . . . . .	13
2.2	Exemplo de rede Bayesiana com três pais . . . . .	14
2.3	Exemplo de rede Bayesiana para estimativa de valor (adaptado de [33]) . . . . .	15
2.4	TPN dos fatores apresentados na Figura 2.3 . . . . .	16
2.5	Exemplos de TNormal . . . . .	24
2.6	Algoritmo para definição de TPN . . . . .	25
2.7	Conversão de escala ordinal para contínua . . . . .	26
3.1	Visão geral do estudo de caso . . . . .	31
3.2	Exemplo de rede Bayesiana para estimativa de valor . . . . .	40
3.3	Rede Bayesiana construída no Caso A . . . . .	42
3.4	Nó pai com frequência contabilizada . . . . .	43
3.5	Exemplo de dados de entrada para o método WSA (Caso A) . . . . .	44
3.6	Rede Bayesiana alimentada com evidências (Caso A) . . . . .	49
3.7	Rede Bayesiana construída no Caso B . . . . .	53
3.8	Exemplo de nó com probabilidades a priori . . . . .	54
3.9	Exemplo de dados de entrada para o método WSA (Caso B) . . . . .	55
4.1	Visão geral das atividades realizadas no experimento . . . . .	64
4.2	Rede Bayesiana utilizada no experimento . . . . .	68
4.3	Planilhas utilizadas na coleta de dados no estudo piloto . . . . .	69
4.4	Rede Bayesiana utilizada no estudo piloto . . . . .	69

---

4.5	Protótipo para calibragem do método RNM . . . . .	76
4.6	Plugin para aplicação do método WSA . . . . .	78
4.7	Plugin para aplicação do método AHP . . . . .	79
4.8	Boxplot das medidas de similaridade e proporção . . . . .	89
4.9	Histograma dos resíduos da análise de variância (TPN Adaptabilidade) . .	93
4.10	Quantil-quantil normal dos resíduos da análise de variância (TPN Adaptabilidade) . . . . .	94
4.11	Gráfico dos resíduos pelos valores preditivos (TPN Adaptabilidade) . . . .	95
4.12	Teste de Tukey (diferença de médias entre métodos) . . . . .	102

# Lista de Tabelas

2.1	Distribuição probabilísticas para $X$ para configurações de pais compatíveis $\{Comp(Y_1 = s)\}$ . . . . .	19
2.2	Distribuição probabilísticas de $X$ para configurações de pais compatíveis $\{Comp(Y_i = s)\}$ , para $i \in \{1, 2, 3\}$ . . . . .	20
2.3	Escala para realizar comparações pareadas (idioma original) . . . . .	21
2.4	Matriz de comparação para elicitación de probabilidade a priore de uma variável $X$ , adaptado de Chin et al. [5] . . . . .	21
2.5	Índice de consistência aleatória onde $n$ é o número de estados, adaptada de Chin et al. [5] . . . . .	22
2.6	Matriz de comparação de um nó $X$ condicionado a um nó pai $Y$ no estado $y^{s_j}$ , adaptada de Chin et al. [5] . . . . .	22
2.7	TPN de um nó com um pai gerada com o método AHP, adaptado de Chin et al. [5] . . . . .	23
2.8	Tabela verdade utilizada para descobrir a expressão ponderada . . . . .	28
3.1	Fatores de valor do Caso A (adaptado de [17]) . . . . .	37
3.2	Fatores de valor do Caso B (adaptado de [17]) . . . . .	38
3.3	Dados demográficos dos participantes do estudo de caso A . . . . .	39
3.4	Dados demográficos dos participantes do estudo de caso B . . . . .	39
3.5	Tempo necessário para aplicar o método WSA (Caso A) . . . . .	45
3.6	Exemplo de tabela utilizada na coleta de dados dos especialistas para o método RNM (Caso A) . . . . .	46
3.7	TPN aproximada obtida com o “Especialista Simulado” (Caso A) . . . . .	48
3.8	Parâmetros utilizados com o método RNM (Caso A) . . . . .	48

3.9	Cenários simulados para validação do modelo construído no Caso A . . . . .	50
3.10	Matriz utilizada para criar cenários a partir de dados históricos . . . . .	51
3.11	Dados de decisões tomadas para a Sprint 10, dispostos em conformidade com a Tabela 3.10 . . . . .	52
3.12	Tempo necessário para aplicar o método WSA (Caso B) . . . . .	55
3.13	Exemplo de tabela utilizada na coleta de dados dos especialistas do Caso B	56
3.14	TPN aproximada obtida com o especialista simulado (Caso B) . . . . .	57
3.15	Parâmetros utilizados com o método RNM (Caso B) . . . . .	57
3.16	Cenários simulados para validação do modelo construído no Caso B . . . . .	58
3.17	Dados de decisões tomadas para a Sprint 10, dispostos em conformidade com a Tabela 3.10 . . . . .	58
3.18	Resultado da validação com cenários simulados (Caso A) . . . . .	59
3.19	Resultado da validação com dados históricos da sprint 10 (Caso A) . . . . .	60
3.20	Resultado da validação com cenários simulados (Caso B) . . . . .	61
3.21	Resultado da validação com dados históricos (Caso B) . . . . .	61
4.1	Hipóteses do experimento . . . . .	67
4.2	Cenários simulados (Estudo Piloto) . . . . .	70
4.3	Resultado dos Cenários Simulados (Estudo Piloto) . . . . .	71
4.4	Dados demográficos dos participantes do experimento . . . . .	72
4.5	Blocagem dos desenvolvedores com randomização dos níveis de $F$ . . . . .	75
4.6	Registro de tempo da execução do experimento (em minutos) . . . . .	83
4.7	Sumário dos dados (BS). Legenda: azul > verde > preto > vermelho . . . . .	84
4.8	Dados brutos coletados (BS). Legenda: azul > verde > preto > vermelho . . . . .	85
4.9	Sumário dos dados (ED). Legenda: azul > verde > preto > vermelho . . . . .	86
4.10	Dados brutos coletados (ED). Legenda: azul > verde > preto > vermelho . . . . .	86
4.11	Sumário dos dados (EM). Legenda: azul > verde > preto > vermelho . . . . .	87
4.12	Dados brutos coletados (EM). Legenda: azul > verde > preto > vermelho . . . . .	87
4.13	Sumário dos dados (PA). Legenda: azul > verde > preto > vermelho . . . . .	88
4.14	Dados brutos coletados (PA). Legenda: azul > verde > preto > vermelho . . . . .	88
4.15	Teste de normalidade dos resíduos (Shapiro-Wilk) . . . . .	92

---

4.16	Alocação de variação do experimento . . . . .	96
4.17	Teste F com nível de significância 0.05 ( <i>Brier Score</i> ) . . . . .	97
4.18	Teste F com nível de significância 0.05 (Distância Euclidiana) . . . . .	97
4.19	Teste F com nível de significância 0.05 (Erro Médio Absoluto) . . . . .	97
4.20	Teste F com nível de significância 0.05 (Acurácia) . . . . .	98
4.21	Testes simultâneos de Tukey para as diferenças de médias (BS) . . . . .	100
4.22	Testes simultâneos de Tukey para as diferenças de médias (ED) . . . . .	100
4.23	Testes simultâneos de Tukey para as diferenças de médias (EM) . . . . .	101
4.24	Testes simultâneos de Tukey para as diferenças de médias (PA) . . . . .	101

# Capítulo 1

## Introdução

Redes Bayesianas são modelos matemáticos que representam as relações probabilísticas de variáveis aleatórias de forma gráfica e numérica por meio do teorema de Bayes. Atualmente, devido à evolução da capacidade computacional que possibilita o cálculo de redes Bayesianas complexas, esta técnica está se tornando popular para auxílio na tomada de decisões [12]. Alguns exemplos de áreas de aplicação de redes Bayesianas são: gerenciamento de projetos de desenvolvimento de software [41,42], projetos de engenharia de larga escala [27], análise de novos produtos [5], planejamento estratégico de marketing [6] e estimativa de sucesso de projetos de inovação [9].

Redes Bayesianas são usadas para representar conhecimento em domínios com incerteza [12]. Este conhecimento é representado em duas partes, uma qualitativa e outra quantitativa. A parte qualitativa refere-se à representação gráfica das variáveis e suas relações (i.e., relações de dependência e independência entre variáveis) por meio de um Grafo Acíclico Dirigido (DAG) que constitui a estrutura da rede Bayesiana. A parte quantitativa, por sua vez, quantifica probabilisticamente as relações causais identificadas na parte qualitativa, por meio de Tabelas de Probabilidade de Nós (TPN).

Um exemplo de rede Bayesiana é apresentado na Figura 1.1, no qual elipses representam nós e setas representam arcos. As funções de probabilidade são representadas pelas TPN. Embora os arcos representem a direção da conexão causal entre as variáveis, a informação pode se propagar em qualquer direção [39] de acordo com as regras de d-separação. No exemplo da Figura 1.1, a probabilidade de uma pessoa contrair câncer é calculada de acordo com duas variáveis: “Câncer na Família” e “Fumante”. A direção das setas dos arcos



indica a relação de dependência para definir as funções de probabilidade. Neste exemplo, assume-se que todas as variáveis são booleanas. Como o nó “Câncer de Pulmão” é apontado por “Câncer na Família” e “Fumante”, a TPN é composta de probabilidades para todas as combinações de estados de “Câncer na Família” e “Fumante”.

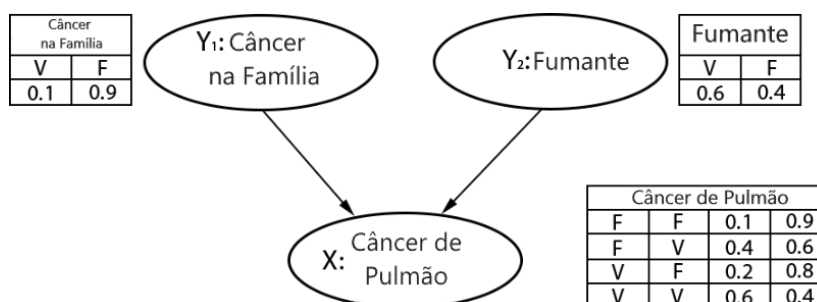


Figura 1.1: Exemplo de rede Bayesiana

Apesar dos avanços, a construção de redes Bayesianas ainda apresenta desafios práticos, estes podem ser subdivididos em dois subproblemas: (i) construir o DAG; (ii) definir as TPN. O foco de interesse deste trabalho é o subproblema (ii). Nos casos em que há um banco de dados com informações suficientes acerca do domínio a ser modelado, é possível automatizar o processo de construção das TPN através de *batch learning* [21] ou usando o algoritmo *Expectation Maximization* que é capaz de lidar com a falta de dados [20]. Infelizmente, na prática, na maioria dos casos, não há dados suficientes [14]. Ou seja, é necessário coletar dados de especialistas e, manualmente, definir as TPN.

Há na literatura propostas de processos para construção de redes Bayesianas com conhecimento de especialistas. Uma destas propostas é o processo de Engenharia de Conhecimento de Rede Bayesiana (KEBN) apresentado em [30]. Este processo é dividido em três etapas: construção da estrutura, quantificação das incertezas e validação do modelo. Estas etapas são executadas em um processo iterativo até que uma rede Bayesiana completa seja construída e validada com o conhecimento de especialistas de domínio.

Por outro lado, a complexidade de definir TPN cresce exponencialmente com o número de nós pais [3, 7, 52], o que torna o processo de definição manual caro e propenso a erros. O KEBN pode reduzir a incidência destes erros, mas não eliminá-los por completo. No que se refere às redes Bayesianas de larga escala, torna-se impraticável a definição manual de todas as combinações possíveis das TPN [12]. Além disso, especialistas geralmente possuem res-

---

trições de tempo e raramente apresentam interesse em definir manualmente uma TPN, visto que consiste em trabalhar com uma grande quantidade de distribuições de probabilidade, independente do tempo e esforço que cada distribuição possa exigir [8].

Ademais, uma série de outros fatores podem comprometer o processo de extração de probabilidades e dificilmente o especialista consegue manter as distribuições mutuamente consistentes durante a definição manual de uma TPN. Em um estudo de caso realizado por Fenton, Neil e Caballero [12], descobriu-se que, para manter a consistência entre as distribuições, os especialistas revisitam partes da TPN já elicitadas, alterando-as. Não obstante, ao invés de melhorar, mais frequentemente este tipo de alteração degrada a TPN, gerando retrabalho. Além disso, fatores como tédio e fadiga são suficientes para fazer com que os critérios usados para descobrir as distribuições não sejam aplicados de forma uniforme durante o processo [8].

Dessa forma, um impasse surge na construção de redes Bayesianas. De um lado, para a definição automática de TPN em redes de larga escala são necessários dados suficientes para aplicar os algoritmos de aprendizagem, e geralmente, este não é o caso para a maioria dos domínios. Do outro, a definição manual de TPN torna-se inviável à medida em que a complexidade das redes Bayesianas aumenta. Uma solução para este impasse é o uso de métodos semiautomáticos, que combinam o conhecimento de especialistas e algoritmos para criar as TPN, e podem inclusive, serem incorporados ao KEBN.

Dois métodos semiautomáticos populares são o Noisy-OR e Noisy-MAX. No entanto, o Noisy-OR tem seu uso limitado a nós booleanos e o Noisy-MAX, por sua vez, não é capaz de modelar a extensão dos relacionamentos necessários para uma rede Bayesiana de larga escala [14]. Contudo, há métodos semiautomáticos aplicáveis a uma faixa maior de redes Bayesianas.

Em Das e Balaram [8], foi apresentado o método *Weighted Sum Algorithm* (WSA). Este método é baseado na heurística da disponibilidade [48] e heurística da simulação [23]. Seu foco principal é montar um subconjunto da TPN a partir do conhecimento de especialistas, com perguntas que englobam casos que são de fácil visualização e simulação por parte dos especialistas e, por conseguinte, estão relacionados com probabilidades mais realistas. Uma vez em posse do subconjunto da TPN e pesos relativos dos nós pais, a TPN completa é gerada por meio de uma soma ponderada.

Em Fenton et al. [12], é apresentado o método *Ranked Nodes Method* (RNM). Neste trabalho, apresenta-se o conceito de nós ranqueados, variáveis aleatórias ordinais representadas em uma escala contínua ordenada monotonamente. Uma característica fundamental deste método é que a TPN dos nós filhos é gerada a partir de expressões matemáticas. Estas expressões definem a tendência central do nó filho para cada combinação de estados dos nós pais e têm como parâmetros de entrada o peso de cada nó pai, que denota a influência exercida sobre o nó filho, e o nível de confiança esperado nos resultados.

Em Chin et al. [5], é apresentada uma abordagem sistemática para gerar probabilidades condicionais de nós com múltiplos pais. Trata-se de uma adaptação do método *Analytic Hierarchy Process* (AHP) para a tarefa de elicitação de probabilidade e geração semiautomática de TPN. Neste método, os especialistas fornecem probabilidades de forma indireta, condicionadas a apenas um pai por vez. Nesta abordagem, as avaliações de probabilidades são derivadas da comparação pareada de estados do nó filho. A TPN é gerada através do produto das probabilidades do nó filho condicionado a cada um dos nós pais.

É neste contexto de construção de redes Bayesianas com métodos híbridos, que combinam conhecimento de especialistas e algoritmos, que se insere este trabalho. Mais especificamente, na avaliação dos métodos semiautomáticos no que se refere à capacidade de modelagem e acurácia dos mesmo em relação à definição de TPN.

## 1.1 Problemática

A definição manual de TPN consiste em um dos maiores obstáculos para a aplicação de redes Bayesianas em diversos domínios. Além do tempo necessário para construí-las, é questionado até que ponto pode-se esperar que especialistas forneçam um grande número de probabilidades de forma consistente durante o processo de construção de uma rede Bayesiana, haja vista as limitações humanas [49]. Os métodos semiautomáticos são uma solução viável para este problema. A combinação do conhecimento de especialistas com algoritmos para definição de TPN completas viabiliza o uso de redes Bayesianas em variados domínios.

Alguns destes métodos já foram avaliados na literatura. Em Mkrtchyan *et al.* [34], cinco métodos de geração de TPN foram aplicados em uma base de conhecimento comum, hipoteticamente extraída de especialistas de domínio. Os critérios utilizados para avaliar os

métodos foram os seguintes: influências e interações de fatores; representação de incertezas; e carga de elicitación. A maior diferença do trabalho de Mkrtychyan *et al.* [34] e esta pesquisa, diz respeito à forma de aplicação dos métodos e os critérios utilizados para avaliá-los.

Nesta pesquisa, os métodos são aplicados diretamente por especialistas de domínio e medidas de similaridade entre TPN são utilizadas para avaliar os métodos. Por outro lado, aqui é também avaliado um dos métodos testados em Mkrtychyan *et al.* [34], o método RNM. Acerca desta característica em comum, as TPN criadas nesta pesquisa permitiram explorar um pouco mais o método RNM (e.g., foi possível a aplicação de mais de um tipo de expressão ponderada).

Em Baker e Mendes [3], dois estudos empíricos foram realizados para avaliar o método WSA em termos de eficiência e acurácia de estimativas, usando como referência TPN definidas manualmente por especialistas de domínio. A acurácia foi analisada a partir do cálculo do erro absoluto entre as células das TPN geradas e as células das TPN de referência. A eficiência foi avaliada em termos de redução do esforço do especialista. A realização da análise de eficiência se deu através da comparação do número de parâmetros de entrada necessários ao algoritmo para construir as TPN e o tamanho das mesmas (i.e., o que seria necessário para preencher as TPN de forma manual).

Embora o trabalho de Baker e Mendes [3] tenha fornecido evidências empíricas que comprovam a eficiência e acurácia do WSA - também objeto de interesse desta pesquisa -, o escopo do trabalho foi limitado à avaliação de um único método. Além disso, uma das ameaças elencadas pelos autores foi o uso de apenas um especialista de domínio, o que pode ter influenciado na acurácia do WSA (o resultado poderia ser melhor, na visão dos mesmos). Assim como em Baker e Mendes [3], neste trabalho pretende-se usar o erro absoluto como medida de similaridade para avaliar as TPN geradas pelos métodos.

Em Zagorecki e Druzdel [51], o método Noisy-OR é avaliado por meio de um estudo empírico em que TPN geradas com o método Noisy-OR são comparadas com TPN elicitadas manualmente por especialistas de domínio. Nesse estudo, a distância euclidiana é utilizada como medida para comparar as TPN. Embora Zagorecki e Druzdel [51] tenham considerado a divergência de Kullback-leibler, esta não foi utilizada devido às especificidades do contexto em que o método foi aplicado. Nesta pesquisa, a divergência de Kullback-leibler foi também considerada, contudo, foi rejeitada por causa de uma limitação: a divergência de

Kullback-leibler não pode ser aplicada em casos em que a probabilidade estimada é zero e a probabilidade real é diferente de zero [52].

Ao contrário do trabalho de Baker e Mendes [2] e Mkrtchyan *et al.* [34], que avaliaram métodos de interesse desta pesquisa, WSA e RNM, respectivamente, o método avaliado em Zagorecki e Druzdzal [51] não foi considerado nesta pesquisa porque é aplicável somente a nós binários. No mais, o uso da distância euclidiana como medida de similaridade e a comparação de TPN geradas de forma semiautomática com TPN de referência definidas manualmente por especialistas de domínio é um ponto em comum entre o trabalho de Zagorecki e Druzdzal [51] e esta pesquisa. A distância euclidiana é uma medida de similaridade comumente utilizada para comparar TPN [19, 51, 52].

Em relação às comparações entre métodos de geração de TPN, a literatura é limitada. Até onde se sabe (a partir do estudo exploratório da literatura que foi realizado), nunca foi realizado nenhum estudo empírico comparativo entre múltiplos métodos semiautomáticos de geração de TPN em que os resultados dos métodos são comparados com TPN definidas de forma manual por especialistas de domínio. O único estudo encontrado que aborda este problema é o trabalho de Mkrtchyan *et al.* [34], todavia, nesse estudo os métodos semiautomáticos não são utilizados diretamente por especialistas de domínio.

Desta forma, o problema em questão é a inexistência de parâmetros que possam ser utilizados por especialistas em redes Bayesianas para avaliar e decidir qual o método semiautomático deve ser aplicado quando o domínio a ser modelado não dispõe de dados históricos para a aplicação de *batch learning* e a definição manual de TPN é inviável. De forma mais específica, o problema em questão é a inexistência de evidências empíricas da aplicação destes métodos em um mesmo contexto, que possibilita a comparação direta e justa dos mesmos.

## 1.2 Objetivo

O objetivo geral deste trabalho consiste em analisar a capacidade de métodos semiautomáticos no que se refere à construção de TPN.

Os objetivos específicos são:

1. Medir e comparar a acurácia dos métodos;
2. Medir e comparar a capacidade de modelagem dos métodos.

Os métodos aqui considerados foram: WSA, RNM, AHP, e um método híbrido denominado WSA-AHP, que combina o método WSA com o método AHP (adaptado para o contexto de redes Bayesianas). Estes métodos foram identificados em uma revisão exploratória da literatura. Os métodos WSA e RNM foram selecionados por terem sido aplicados em diversos domínios e já terem sido amplamente explorados na literatura.

O método AHP, por sua vez, foi selecionado devido a uma de suas características mais proeminentes: a capacidade de extração de probabilidades de forma indireta por meio de comparações entre pares de estados. Esta característica viabilizou a criação de um método híbrido que combina os métodos AHP e WSA. Este método híbrido possibilitou a investigação dos efeitos do uso de uma técnica indireta de extração de probabilidades em conjunto com o conceito de configurações de pais compatíveis e o algoritmo do método WSA.

Outros métodos foram considerados nesta pesquisa, como aqueles explorados em Mkrtchyan *et al.* [34], mas por uma questão prática, foi necessário limitar o escopo desta pesquisa para no máximo quatro métodos, haja vista o custo demasiadamente alto para incluir mais métodos na pesquisa.

## 1.3 Metodologia

De acordo com o objetivo desta pesquisa, a mesma classifica-se como um estudo empírico exploratório. Para responder às questões de pesquisa foram realizados um estudo de caso e um experimento controlado em um contexto de desenvolvimento de software e gerenciamento de projetos.

O estudo de caso foi realizado por meio de uma parceria com pesquisadores da Universidade de Oulu, vinculados ao projeto Value<sup>1</sup>, e pesquisadores da Universidade Federal de Campina Grande. Participaram do estudo de caso: gerentes de projetos e desenvolvedores. Foram construídos dois modelos baseados em redes Bayesianas para estimativa de valor em um contexto de Engenharia de Software Baseada em Valor (VBSE). A construção destes modelos esteve em alinhamento com os passos estabelecidos no processo KEBN.

Para alcançar o objetivo de pesquisa 1, estabelecido na Seção 1.2, foram utilizados dados referentes à validação dos modelos. Estes dados foram utilizados para medir a acurácia dos

---

<sup>1</sup><http://valueproject.fi/>

métodos.

Para validar os modelos, considerou-se a acurácia das estimativas, derivada da comparação das predições dos modelos para com dados históricos de projetos finalizados, e o passo a passo cognitivo do modelo, que utiliza cenários hipotéticos para avaliar se os modelos fornecem saídas que fazem sentido a partir da ótica dos especialistas de domínio. Estes dois métodos são comumente utilizados para validar este tipo de modelo [30, 32].

Uma visão geral da metodologia aplicada para alcançar os objetivos desta pesquisa é apresentado na Figura 1.2.

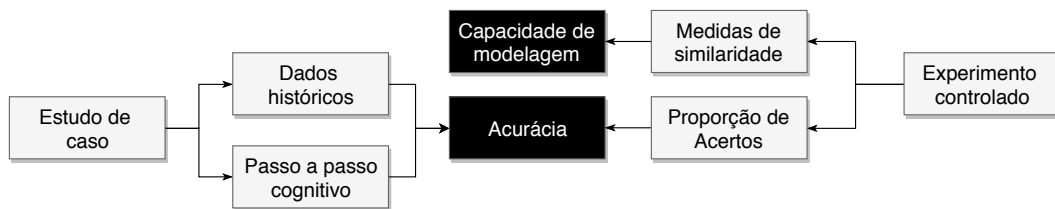


Figura 1.2: Visão geral da metodologia aplicada

Com relação ao experimento controlado, seu intuito foi realizar os objetivos 1 e 2 desta pesquisa. Para realizar o objetivo 2, especialistas de domínio aplicaram os métodos e definiram TPN de forma manual, viabilizando a comparação das TPN geradas pelos métodos com TPN de referência, que teoricamente, representam o conhecimento dos especialistas de domínio.

A comparação entre as TPN geradas de forma semiautomática e as TPN de referência se deu através do uso de medidas de similaridade comumente utilizadas na literatura para este fim. Dessa forma, para realizar a comparação entre as TPN geradas com os métodos semiautomáticos e as TPN de referência, foram utilizadas as seguintes medidas de similaridade: Erro Médio Absoluto (EM); *Brier Score* (BS); e Distância Euclidiana (ED).

Para alcançar o objetivo de pesquisa 1, foram coletados dados acerca do número de acertos obtidos com cada TPN gerada pelos métodos semiautomáticos no que diz respeito à moda dos vetores de probabilidade ou distribuições de probabilidade que compõem as TPN. Este método de avaliação se assemelha ao passo a passo cognitivo do modelo, porém, os valores que seriam esperados pelos especialistas de domínio em cada caso hipotético são extraídos diretamente das TPN de referência e os cenários são todas as combinações possíveis de estados dos nós pais.

Dessa forma, é definida a métrica: Proporção de Acertos (PA), que representa a acurácia dos métodos em relação à capacidade de estimar corretamente os estados de maior probabilidade dos nós em uma rede Bayesiana.

## 1.4 Resultados e Contribuições

Neste trabalho, foi realizado uma análise comparativa entre métodos de geração semiautomática de TPN através de dados coletados em um estudo de caso e um experimento controlado, ambos realizados em um contexto de desenvolvimento de software.

A partir dos resultados, foi possível obter evidências empíricas de que os métodos apresentam diferenças significativas em relação à acurácia e capacidade de modelagem de TPN. Os métodos de melhor desempenho em relação à acurácia foram RNM e AHP. Em relação à capacidade de modelagem, os métodos RNM, WSA e WSA-AHP obtiveram resultados superiores aos do método AHP.

As contribuições deste trabalho são as seguintes:

- Evidências empíricas sobre a acurácia e capacidade de modelagem de conhecimento de especialistas de domínio, dos métodos WSA, RNM, WSA-AHP e AHP;
- Uma extensão ao método WSA para elicitação de probabilidades de forma indireta;
- Uma nova abordagem para aplicação do método RNM que viabiliza seu uso sem que seja necessário a presença de um especialista em RNM;
- Disponibilização de uma ferramenta para construção de redes Bayesianas com suporte ao método RNM (protótipo que antecedeu o desenvolvimento da aplicação utilizada no experimento controlado)<sup>2</sup>;
- Disponibilização de um serviço web que calcula TPN completas a partir de dados de entrada do método WSA<sup>3</sup>;
- Disponibilização dos algoritmos necessários para aplicar os métodos RNM<sup>4</sup> e WSA<sup>5</sup>.

<sup>2</sup><https://github.com/joaonunes-copin/kaizen-prototype>

<sup>3</sup><https://wsa-ws.herokuapp.com/>

<sup>4</sup><https://github.com/joaonunes-copin/KaizenRNM>

<sup>5</sup><https://github.com/joaonunes-copin/wsa-algorithm>



## **1.5 Estrutura da Dissertação**

O conteúdo desta dissertação está organizado da seguinte forma:

- **Capítulo 2** - Neste capítulo, os principais conceitos relevantes para este trabalho são apresentados, incluindo uma descrição detalhada dos métodos de geração de TPN;
- **Capítulo 3** - Neste capítulo, o estudo de caso que foi utilizado para comparar os métodos RNM e WSA é apresentado;
- **Capítulo 4** - Neste capítulo, o experimento controlado para comparação dos métodos RNM, WSA, WSA-AHP e AHP é apresentado;
- **Capítulo 5** - Neste capítulo, é apresentada uma discussão dos resultados obtidos no estudo de caso e no experimento controlado;
- **Capítulo 6** - Neste capítulo, é apresentada a conclusão deste estudo e trabalhos futuros.

# Capítulo 2

## Fundamentação Teórica

Neste capítulo, são abordados os principais conceitos relacionados a redes Bayesianas no contexto deste trabalho.

### 2.1 Rede Bayesiana

Formalmente, uma rede Bayesiana  $B$  é um gráfico acíclico direcionado que representa uma distribuição conjunta de probabilidade sobre o conjunto de variáveis aleatórias  $V$  [18]. Dessa forma, uma rede Bayesiana é definida pelo par  $B = \{G, \Theta\}$ , onde  $G = (V, E)$  é um grafo acíclico direcionado em que os nós  $V_1, \dots, V_n$  representam variáveis aleatórias e os arcos  $E$  representam as relações de dependência entre essas variáveis.  $\Theta$  representa o conjunto de funções de probabilidade. Este conjunto contém os parâmetros  $\theta_{v_i|\pi_i} = P_B(v_i|\pi_i)$  para cada  $v_i$  em  $V_i$  condicionado por  $\pi_i$ , o conjunto parâmetros de  $V_i$  em  $G$ . A equação 2.1 apresenta a distribuição conjunta definida por  $B$  sobre  $V$ .

$$P_B(V_1, \dots, V_n) = \prod_{i=1}^n P_B(v_i|\pi_i) = \prod_{i=1}^n \theta_{v_i|\pi_i} \quad (2.1)$$

Esta técnica constitui-se em uma ferramenta que pode lidar com problemas que podem ser modelados em uma estrutura de rede, representando o conhecimento de especialistas em domínios em que esse conhecimento é probabilístico. A maior vantagem desta técnica é a possibilidade de atualizar os dados a medida que novas observações se tornam disponíveis. Ou seja, com os algoritmos e softwares existentes, é possível atualizar convenientemente

os julgamentos anteriores quando novas evidências se tornam disponíveis. Isso é possível devido ao teorema de Bayes.

### 2.1.1 Teorema de Bayes

O teorema de Bayes é a base para o uso de redes Bayesianas. A equação 2.2 apresentada a seguir representa o teorema de Bayes [30]:

$$p(X|Y) = \frac{p(Y|X)p(X)}{p(Y)} \quad (2.2)$$

onde:

- $p(X)$  é a distribuição de probabilidade a priori de  $X$  ou distribuição marginal de probabilidade. Ela representa a distribuição de probabilidade de  $X$  sem levar em consideração o que se conhece de  $Y$ .
- $p(X|Y)$  é a distribuição de probabilidade a posteriori de  $X$  quando seu valor depende da evidência de  $Y$ . Representa a probabilidade condicional de  $X$  dado  $Y$ .
- $p(Y|X)$  é a distribuição de probabilidade de  $Y$  dado  $X$ .
- $p(Y)$  é a distribuição de probabilidade a priori da evidência de  $Y$ , e é utilizada como uma constante normalizadora.

Para ilustrar o funcionamento do teorema de Bayes na prática, considere a rede Bayesiana apresentada na Figura 2.1a e suas respectivas TPN.

Este modelo apresenta a relação de dependência entre o nó “Câncer de Pulmão” e o nó “Fumante”. A TPN do nó “Câncer de Pulmão” mostra que a probabilidade de um dado indivíduo contrair câncer é de 25% quando o mesmo é fumante. Por outro lado, se o indivíduo não for fumante a probabilidade do mesmo contrair câncer é de 10%.

Digamos que o indivíduo tenha contraído câncer (i.e., há evidência para o nó “Câncer de Pulmão”), para calcular a probabilidade dele ser fumante dado que ele tem câncer é utilizado o teorema de Bayes:

$$p(X|Y) = \frac{25 \cdot 60}{19} \quad (2.3)$$

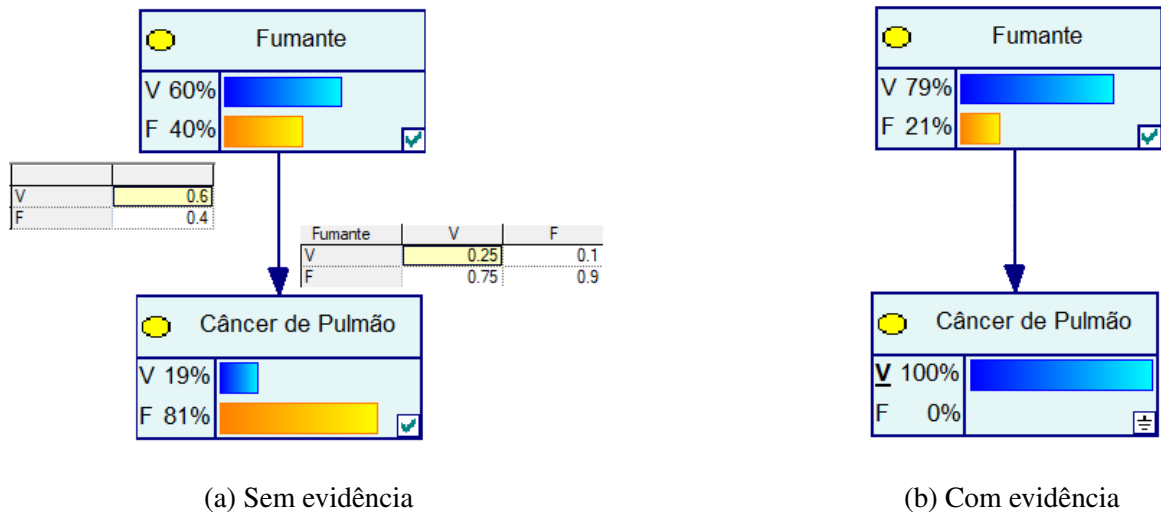


Figura 2.1: Probabilidades calculadas usando o teorema de Bayes

Portanto, a probabilidade do indivíduo ser fumante é de 78,94% conforme apresentado com valores arredondados na Figura 2.1b.

### 2.1.2 Custo da definição de um TPN

Existe um consenso geral de que a parte mais delicada da construção de uma rede Bayesiana é a definição das TPN [10]. As TPN podem ser aprendidas a partir de dados ou construídas manualmente com o uso de especialistas. Na prática, é comum não ter dados suficientes para o aprendizado e a definição manual torna-se a única opção.

Para ilustrar o custo de definir manualmente uma TPN, considere o exemplo apresentado na Figura 2.2. Nessa rede Bayesiana, o objetivo é avaliar a eficiência de uma equipe ágil de desenvolvimento de software. O trabalho em equipe é diretamente influenciado pela autonomia (i.e., capacidade de autogestão e liderança compartilhada); Coesão (i.e., a capacidade de estar em estreita concordância e trabalhar bem em conjunto); e Colaboração (i.e., a capacidade de comunicar e coordenar as atividades). Este exemplo será usado ao longo deste capítulo.

Para eliciar todas as probabilidades necessárias para construir a TPN do nó filho “Trabalho em Equipe”, um facilitador (e.g., um especialista em redes Bayesianas) deve fazer  $5^3$  perguntas ao especialista de domínio, uma pergunta para cada  $P(v_i|\pi_i)$ . Generalizando este cálculo nós temos que a quantidade de probabilidades a serem elicitadas é igual a  $s^k$ , onde

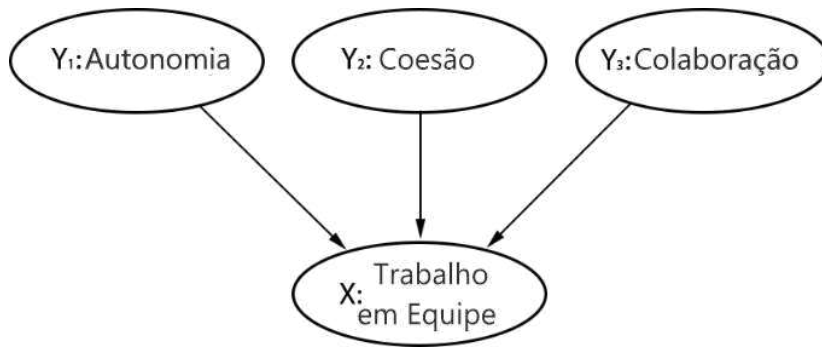


Figura 2.2: Exemplo de rede Bayesiana com três pais

Exemplo adaptado de [15] onde um nó filho “Trabalho em Equipe” é influenciado por três nós pais: “Autonomia” ( $Y_1$ ), “Coesão” ( $Y_2$ ) e “Colaboração” ( $Y_3$ ). Cada nó tem cinco estados ordinais: muito baixo (MB), baixo (B), mediano (M), alto (A), muito alto (MA)

$s$  é o número de estados (supondo que todos os nós possuem a mesma quantidade) e  $k$  é o número de pais. Ou seja, a complexidade da TPN aumenta em função do número de pais de um nó, de forma exponencial, tornando o processo de definição manual bastante custoso e propenso a erros.

### 2.1.3 Propriedades de separação direcional

Para definir o DAG é necessário definir as relações de dependência entre as variáveis, aplicando as propriedades de separação direcional (d-separação) [147]. Dado um grafo direcionado, um caminho entre dois conjuntos de nós  $X$  e  $Y$  é qualquer sequência de nós entre um membro de  $X$  e um membro de  $Y$ , tal que todos os pares adjacentes de nós são conectados por um arco e cada nó apenas aparece uma vez na sequência. Um caminho é bloqueado, dado um conjunto de nós  $E$ , se existir um nó  $Z$  no caminho para a qual uma das condições a seguir é verdadeira:

1.  $Z$  pertence a  $E$  e  $Z$  tem um arco no caminho apontando para  $Z$  e um arco apontando de  $Z$  (i.e., corrente);
2.  $Z$  pertence a  $E$  e  $Z$  tem ambos os arcos apontando para fora (i.e., causa comum);
3. Nem  $Z$  e nenhum descendente de  $Z$  pertence a  $E$ , e ambos os arcos apontam para  $Z$  (i.e., efeito comum).

Para exemplificar a propriedade 1, considere a estrutura  $A \rightarrow B \rightarrow C$ . Não se pode afirmar que  $A \perp\!\!\!\perp C$ . Por outro lado, pode-se afirmar que  $A \perp\!\!\!\perp C|B$  e  $C \perp\!\!\!\perp A|B$ . Para a propriedade 2, considere a configuração  $A \rightarrow B \leftarrow C$ . Pode-se afirmar que  $A \perp\!\!\!\perp B$ , se não houver evidências nos nós. Por outro lado, caso haja evidência em  $B$  e um dos nós-pai (e.g.,  $A$ ), o outro pai (e.g.,  $B$ ) é dependente. Para a propriedade 3, considere a configuração  $A \leftarrow B \rightarrow C$ . Essa configuração é também conhecida como Classificador Nãive Bayes, onde os nós  $A$  e  $C$  são considerados nós indicadores [66]. Nesse caso, não se pode afirmar que  $A \perp\!\!\!\perp C$ . Por outro lado, pode-se afirmar que  $A \perp\!\!\!\perp C|B$  e  $C \perp\!\!\!\perp A|B$ .

### 2.1.4 Redes Bayesianas em VBSE

Em um contexto de Engenharia de Software Baseada em Valor, uma rede Bayesiana pode ser utilizada para representar o conhecimento de especialistas em termos de fatores considerados importantes à tomada de decisões baseadas em valor. Estas decisões estão relacionados ao gerenciamento e desenvolvimento de produtos de software. A Figura 2.3 apresenta um exemplo simples de rede Bayesiana para estimativa de valor.

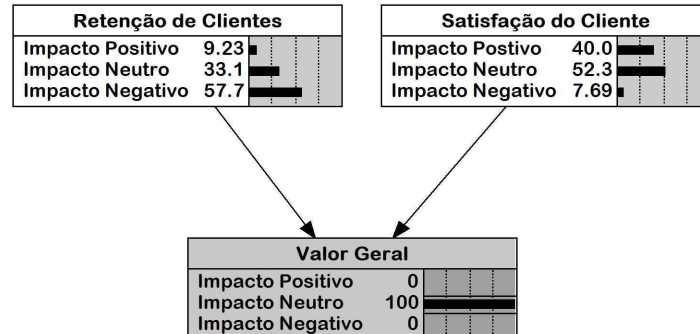


Figura 2.3: Exemplo de rede Bayesiana para estimativa de valor (adaptado de [33])

O fragmento de rede Bayesiana apresentado na Figura 2.3 representa um cenário em que diferentes *stakeholders* decidem sobre um conjunto de funcionalidades a serem selecionados para implementação em determinada *release* de um produto de software [33], com base em três fatores: (i) “Valor Geral”, que representa o impacto geral associado à implementação de uma funcionalidade específica em um produto de software; (ii) “Retenção de Clientes”, que representa o potencial de retenção do cliente; e (iii) “Satisfação do Cliente”, que representa o potencial de satisfazer o cliente. Estes fatores ajudam os *stakeholders* a decidir quais

funcionalidades serão selecionadas.

Podemos ainda observar duas setas na Figura 2.3, apontando dos fatores “Retenção de Clientes” e “Satisfação de Cliente” em direção ao fator “Valor Geral”. As setas representam uma relação de causa e efeito entre os fatores de origem e de destino, respectivamente. Isso significa que os valores dos fatores “Retenção de Cliente” e “Satisfação do Cliente” exercem influência sobre o fator “Valor Geral”.

Conforme já abordado, cada nó de uma rede Bayesiana, aqui também chamado de fator, está associado a uma TPN que quantifica probabilisticamente todos os cenários possíveis de tomada de decisões (exemplos são apresentados na Figura 2.4). As TPN relacionadas aos fatores “Retenção de Clientes” e “Satisfação do Cliente” representam as probabilidades a priori, ou seja, as probabilidades mais prováveis para cada estado na ausência de evidências.

TPN do nó Retenção do Cliente			TPN do nó Satisfação do Cliente		
Impacto Positivo	Impacto Neutro	Impacto Negativo	Impacto Positivo	Impacto Neutro	Impacto Negativo
60	10	30	65	10	25

TPN do nó Valor Geral				
Retencao_de_Clientes	Satisfacao_do_Cliente	Impacto Positivo	Impacto Neutro	Impacto Negativo
Impacto Positivo	Impacto Positivo	100	0	0
Impacto Positivo	Impacto Neutro	95	5	0
Impacto Positivo	Impacto Negativo	60	0	40
Impacto Neutro	Impacto Positivo	95	5	0
Impacto Neutro	Impacto Neutro	50	50	0
Impacto Neutro	Impacto Negativo	0	10	90
Impacto Negativo	Impacto Positivo	60	5	35
Impacto Negativo	Impacto Neutro	0	30	70
Impacto Negativo	Impacto Negativo	0	0	100

Figura 2.4: TPN dos fatores apresentados na Figura 2.3

De acordo com a Figura 2.3, podemos observar que 60% das funcionalidades discutidas em reuniões passadas foram consideradas de impacto positivo sobre a retenção de clientes, 65% foram consideradas de impacto positivo sobre a satisfação do cliente, e assim por diante. No que se refere ao fator “Valor Geral”, dado que este fator é influenciado pelos outros dois, sua TPN está condicionada a todas as combinações de estados possíveis dos outros fatores (i.e., nós pais).

A Figura 2.4, por exemplo, mostra que o impacto geral (i.e., “Valor Geral”) que uma determinada funcionalidade trará para um produto existente será 100% positivo se o impacto desta funcionalidade sobre os fatores “Retenção de Clientes” e “Satisfação do Cliente” forem avaliados como positivo, e assim por diante. Essas probabilidades são os meios utilizados em redes Bayesianas para a quantificação probabilística das incertezas relacionadas à tomada de decisões em um domínio a ser modelado.

## 2.2 Métodos de Geração de TPN

Para a realização dos estudos de caso e experimentos contidos neste trabalho, três métodos de geração semiautomática de TPN foram utilizados. Nesta seção, estes métodos são apresentados em detalhes, tal como foram utilizados neste trabalho.

### 2.2.1 Método WSA

O método WSA foi proposto em Das [7]. Neste trabalho, é introduzido o conceito de configuração de pais compatíveis, baseado em duas heurísticas: disponibilidade e simulação. A heurística da disponibilidade opera sob a suposição de que quanto mais cognitivamente acessível é um evento, mais provável é que ele ocorra. O que significa dizer que, uma pessoa relaciona a probabilidade de um evento com a facilidade com que este evento é lembrado ou retratado na sua mente. A heurística de simulação, por sua vez, é utilizada quando não é possível usar a heurística da disponibilidade (i.e., quando não há instâncias do evento na mente da pessoa). De acordo com esta heurística, as pessoas determinam a probabilidade de um evento com base em quão fácil é simulá-lo mentalmente.

Para definir formalmente o conceito de configuração de pais compatíveis, é tomado como base o trabalho de Mendes et al. [32]. Portanto, sobrescrito é usado para representar os estados de um nó, e subscrito, para diferenciar os nós pais. Dessa forma, considere que para  $Y_i$  é atribuído um estado arbitrário  $y_i^v$ , ou seja,  $Y_i = y_i^v$ , já que  $Y_j$  é outro nó pai, tal que  $Y_j$  é considerado compatível com  $Y_i = y_i^v$  somente quando  $Y_j$  está no estado  $y_j^w$  que é mais provável, de acordo com o conhecimento do especialista, coexistir com  $Y_i = y_i^v$ . Portanto, é utilizada a notação  $Comp[Y_i = y_i^v]$  para representar o conjunto de estados que são compatíveis com  $Y_i = y_i^v$  para todos os nós pais.

$$Comp[Y_i = y_i^v] = \{y_j^w, \forall j \neq i \mid \max_{w=1 \dots |Y_j|} P(y_j^w | y_i^v)\} \quad (2.4)$$

Para obter as configurações de pais compatíveis, se faz necessária escolher uma combinação plausível de estados para cada  $Comp[Y_i = y_i^v]$ , que, teoricamente, são mais fáceis de simular e, portanto, propenso a probabilidades mais realistas. Por isso, é extraído de especialistas as distribuições de probabilidade para todas as configurações de pais compatíveis e



os pesos relativos dos nós pais. A TPN é calculada usando um algoritmo de soma ponderada [7], que usa essas distribuições de probabilidade e pesos para gerar a TPN completa. Portanto, os dados necessários para aplicar o método WSA são:

- Peso relativo (entre zero e um) de cada nó pai, denotando seu grau de influência sobre o nó filho  $w_1, \dots, w_n$ ;
- Distribuições de probabilidade  $k_1 + \dots + k_n$  de  $X$  para configurações de pais compatíveis.

$$p(x^l | y_1^{v_1}, y_2^{v_2}, \dots, y_n^{v_n}) = \sum_{i=1}^n w_i p(x^l | \text{Comp}[Y_j = y_j^{v_j}]) \quad (2.5)$$

Onde  $w$  é o peso relativo do nó pai,  $l = 0, 1, \dots, m$  e  $v_j = 1, 2, \dots, k_j$ . Uma restrição deve ser observada: a soma de todos os pesos relativos (i.e., de todos os nós pais) deve ser igual a 1. Um peso igual a 0 indica que o nó pai não tem influência sobre o nó filho e, portanto, pode ser omitido da relação. Por outro lado, um peso relativo igual a 1 indica que o nó pai é o único determinante das probabilidades condicionais no nó filho.

Por exemplo, considerando a rede Bayesiana mostrada na Figura 2.2, onde o objetivo é avaliar o trabalho em equipe. Por uma questão de simplicidade, digamos que todos os pais possuem apenas os estados “Baixo”, “Médio” e “Alto” em vez dos cinco estados do exemplo original. Com o método WSA,  $3 * 3$  distribuições probabilísticas são necessárias para construir uma TPN completa ao invés de  $3^3$  no caso de definição manual. Começando com o pai  $Y_1$ , digamos que o especialista interprete subjetivamente as configurações de pais compatíveis como uma relação de equivalência da seguinte forma:

$$\{\text{Comp}(Y_1 = s)\} \equiv \{\text{Comp}(Y_2 = s)\} \equiv \{\text{Comp}(Y_3 = s)\}, s \in \{b, m, a\} \quad (2.6)$$

Quando o especialista fornece 3 distribuições de probabilidade para o nó  $Y_1$ , todas as distribuições das configurações de pais compatíveis são obtidas, sendo necessárias 3 perguntas apenas, ao invés de 9. Para gerar a TPN, o especialista deve atribuir pesos relativos dos pais. Digamos que o especialista interprete “Autonomia” e “Coesão” como tendo a mesma força de influência sobre o nó filho, e “Colaboração” como três vezes mais importante do que “Coesão” ou “Autonomia”. Assim, atribuindo os seguintes pesos:  $w_1 = 2, w_2 = 2, w_3 = 6$ .

Configurações de pais compatíveis	Distribuições probabilísticas		
	$s = b$	$s = m$	$s = a$
$p(X = b \{Comp(Y_1 = s)\})$	0.8	0.2	0.1
$p(X = m \{Comp(Y_1 = s)\})$	0.1	0.6	0.1
$p(X = a \{Comp(Y_1 = s)\})$	0.1	0.2	0.8

Tabela 2.1: Distribuição probabilísticas para  $X$  para configurações de pais compatíveis  $\{Comp(Y_1 = s)\}$

Com os pesos e 3 distribuições de probabilidade sobre o nó  $Y_1$  como entradas, o algoritmo de soma ponderada calcula todas as  $3^3$  distribuições necessárias para preencher a TPN. Por outro lado, digamos que a Eq. 2.7 não seja satisfeita, todas as distribuições de probabilidade (i.e.,  $3 * 3$ ) devem ser elicítadas, como mostrado na Tabela 2.2:

Neste caso, a probabilidade para “Trabalho em Equipe” = “baixo” condicionado a “Autonomia” = “baixa”, “Coesão” = “média” e “Colaboração” = “alta” seria dada por:

$$\begin{aligned}
 p(X = b|Y_1 = b, Y_2 = m, Y_3 = a) = & w_1p(X = b|\{Comp(Y_1 = b)\}) + \\
 & w_2p(X = b|\{Comp(Y_2 = m)\}) + \\
 & w_3p(X = b|\{Comp(Y_3 = a)\})
 \end{aligned} \tag{2.7}$$

Isto resume o método WSA, para uma descrição detalhada do método e sua validação matemática consultar o trabalho de Das [7]. Infelizmente, em Das [7] não é descrito como lidar com situações em que o especialista encontra mais de uma configuração de pais compatíveis e não sabe qual escolher. Assim, uma extensão para este método é proposta por Baker e Mendes [2] para lidar com tais situações, calculando a média das probabilidades de configurações de pais compatíveis válidas que os especialistas possam selecionar.

### 2.2.2 Método AHP

Embora a avaliação direta de probabilidades durante a definição manual de TPN seja viável para redes Bayesianas simples, para redes de médio a grande porte, a complexidade e o ônus para os especialistas crescem substancialmente. À medida em que o número de pais e

Configurações de pais compatíveis	Distribuições probabilísticas		
	$s = b$	$s = m$	$s = a$
$p(X = b \{Comp(Y_1 = s)\})$	0.8	0.2	0.1
$p(X = m \{Comp(Y_1 = s)\})$	0.1	0.6	0.1
$p(X = a \{Comp(Y_1 = s)\})$	0.1	0.2	0.8
$p(X = b \{Comp(Y_2 = s)\})$	0.6	0.4	0.3
$p(X = m \{Comp(Y_2 = s)\})$	0.4	0.5	0.4
$p(X = a \{Comp(Y_2 = s)\})$	0.0	0.1	0.3
$p(X = b \{Comp(Y_3 = s)\})$	0.4	0	0
$p(X = m \{Comp(Y_3 = s)\})$	0.6	0.8	0.1
$p(X = a \{Comp(Y_3 = s)\})$	0	0.2	0.9

Tabela 2.2: Distribuição probabilísticas de  $X$  para configurações de pais compatíveis  $\{Comp(Y_i = s)\}$ , para  $i \in \{1, 2, 3\}$

estados aumenta, mais difícil se torna para os especialistas raciocinarem sobre probabilidades condicionadas a múltiplos pais e múltiplas combinações de estados de uma vez, e mais suscetível se torna a vieses e imprecisões [5].

Em Chin et al. [5] uma abordagem sistemática para gerar probabilidades condicionais de nós com múltiplos pais é proposta. Trata-se de uma adaptação do método AHP para a tarefa de elicitação de probabilidade e geração semiautomática de TPN. Neste método, os especialistas fornecem probabilidades de forma indireta condicionadas a apenas um pai por vez. As avaliações de probabilidades são derivadas da comparação de pares de estados. A TPN é gerada através do produto das probabilidades do nó filho condicionado a cada um dos nós pais.

Antes de usar este método, no entanto, é necessário definir uma escala para realizar comparação dos estados do nó. A escala de Saaty [45] pode ser usada para este propósito ou uma escala pode ser criada. Um bom exemplo de como obter uma escala é apresentado no trabalho de Renooij e Witteman [43], no qual quatro experimentos sucessivos foram realizados para gerar uma escala com números e palavras. Na Tabela 2.3 é apresentada a escala de Saaty com nove valores, adaptada para um contexto de probabilidades.

Escala	Definição	Explicação
1	Equal likely	Event A and event B are equal likely
2	Weak or slight	
3	Moderate more likely	Event A is moderate more likely than event B
4	Moderate plus	
5	Strong more likely	Event A is Strong more likely than event B
6	Strong plus	
7	Very strong more likely	Event A is very strong more likely than event B
8	Very, very Strong	

Tabela 2.3: Escala para realizar comparações pareadas (idioma original)

	$x^{s1}$	$x^{s2}$	$x^{sn}$	$\omega$
$x^{s1}$	$a_{11}$	$a_{12}$	$a_{1n}$	$\omega_1$
$x^{s2}$	$a_{21}$	$a_{22}$	$a_{2n}$	$\omega_2$
$x^{sn}$	$a_{n1}$	$a_{n2}$	$a_{nn}$	$\omega_n$

Tabela 2.4: Matriz de comparação para elicitación de probabilidade a priore de uma variável  $X$ , adaptado de Chin et al. [5]

Para um melhor entendimento do método, a terminologia original usada no AHP é substituída por termos mais apropriados ao contexto de probabilidades. Assim, o termo *atributo* é substituído por *evento* e o termo *importância* é substituído por *probabilidade*. Para obter probabilidades a priore, comparações de pares de todos os estados do nó são realizadas. Como todos os estados são comparadas entre si, é possível montar uma matriz de comparação. Na Tabela 2.4 é apresentado um exemplo de matriz de comparação que pode ser usada para definir as probabilidades a priore de um nó.

Na matriz da Tabela 2.4,  $a_{ij}(i = 1, 2, \dots, n; j = 1, 2, \dots, n)$  é especificado pela pergunta “comparando o estado  $x^{si}$  com  $x^{sj}$ , qual é mais provável e quão mais provável?”. Depois de preencher os valores de  $a_{ij}$ , é possível encontrar os valores de  $a_{ji}$  calculando a inversa de  $a_{ij}$ , ou seja,  $1/a_{ij}$ . O resultado final é uma matriz recíproca com todos os elementos na diagonal igual a 1, isto é,  $a_{ii} = 1$  para todo  $i$ .

$n$	1	2	3	4	5	6	7	8	9
$RI$	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45

Tabela 2.5: Índice de consistência aleatória onde  $n$  é o número de estados, adaptada de Chin et al. [5]

	$x^{s1}$	$x^{s2}$	$x^{sn}$	$\omega_x$
$x^{s1}$	$a_{11}$	$a_{12}$	$a_{1n}$	$\omega_{1j}$
$x^{s2}$	$a_{21}$	$a_{22}$	$a_{2n}$	$\omega_{2j}$
$x^{sn}$	$a_{n1}$	$a_{n2}$	$a_{nn}$	$\omega_{nj}$

Tabela 2.6: Matriz de comparação de um nó  $X$  condicionado a um nó pai  $Y$  no estado  $y^{sj}$ , adaptada de Chin et al. [5]

A prioridade relativa de  $x^{si}$  é obtida a partir do autovetor máximo  $\omega = (\omega_1, \omega_2, \dots, \omega_n)^T$  da matriz  $(a_{ij})_{n \times n}$  e a consistência da matriz é a razão de consistência  $CR = CI/RI$ , onde  $CI$  é o índice de consistência, definido por  $(\lambda_{\max} - n)/(n - 1)$  onde  $\lambda_{\max}$  é o valor máximo correspondente a  $\omega$ , e  $RI$  é um índice aleatório dado pela tabela 2.5. Uma matriz de comparação com  $CR$  inferior a 0.10 é considerada aceitável [5], embora Monti e Carenini [35] tenham observado que esse limiar pode ser inadequado para o propósito de avaliar as probabilidades. Como a soma de todos os elementos em  $\omega$  é 1 e o  $i$ -ésimo elemento  $\omega_i$  representa a importância relativa do estado  $x^{si}$ ,  $\omega_i$  é agora interpretado como a probabilidade a priori do estado  $x^{si}$ , isto é,  $P(x^{si}) = \omega_i$ .

De forma similar, para obter as probabilidades de um nó  $X$  com um único pai  $Y$ , estimamos  $P(x^{si}|y^{sj})$ . Na Tabela 2.6 é possível observar a matriz resultante quando o nó  $Y = y^{sj}$ , na qual  $a_{pq}(p = 1, 2, \dots, n; q = 1, 2, \dots, n)$  é especificada por questões como “se o nó  $Y$  está no estado  $y^{sj}$ , comparando os estados  $x^{si}$  e  $x^{sj}$  do nó filho  $X$ , qual é mais provável e quão mais provável?”. Depois de obter  $\omega_{ij}(i = 1, \dots, n)$  temos  $P(X = x^{si}|Y = y^{sj}) = \omega_{ij}$ . O número de matrizes necessárias para obter  $\omega_{ij}(i = 1, 2, \dots, n; j = 1, 2, \dots, m)$  é igual ao número de estados de  $Y$ . Os resultados obtidos compõem a TPN de um nó filho  $X$  condicionado aos estados de um nó pai  $Y$ , como mostrado na Tabela 2.7.

A abordagem para gerar as probabilidades condicionais para nós com múltiplos pais é ba-

		<i>Estados de Y</i>			
		$x^{s1}$	$x^{s2}$	$x^{sn}$	$\omega_x$
<i>Estados de X</i>	$x^{s1}$	$\omega_{11}$	$\omega_{12}$	$\omega_{1n}$	$\omega_{1j}$
	$x^{s2}$	$\omega_{21}$	$\omega_{22}$	$\omega_{2n}$	$\omega_{2j}$
	$x^{sn}$	$\omega_{n1}$	$\omega_{n2}$	$\omega_{nn}$	$\omega_{nj}$

Tabela 2.7: TPN de um nó com um pai gerada com o método AHP, adaptado de Chin et al. [5]

seada em Kim e Pearl [24], que afirma que quando um nó  $A$  possui dois pais  $B$  e  $C$ , sua probabilidade condicionada a  $B$  e  $C$  pode ser aproximada por  $P(A|B, C) = \alpha P(A|B)P(A|C)$  onde  $\alpha$  é um fator de normalização que assegura que  $\alpha \sum_{a \in A} P(a|B, C) = 1$ . Portanto, para gerar a TPN, a equação 2.8 é aplicada:

$$P(X = x^{s_i} | Y_1 = y_1^{s_i}, Y_2 = y_2^{s_i}, \dots, Y_k = y_k^{s_i}) = \alpha \prod_{j=1}^k P(X = x^{s_i} | Y_j = y_j^{s_i}) \quad (2.8)$$

O foco desta abordagem é a redução do esforço dos especialistas, gerando automaticamente distribuições probabilísticas de nós com múltiplos pais e, conseqüentemente, a TPN completa através do cálculo do produto das probabilidades condicionadas a cada pai. Assim, o especialista avalia as probabilidades de um determinado nó filho condicionado a cada um de seus pais, de forma individual, considerando um nó pai de cada vez, e essas probabilidades são combinadas para obter as probabilidades condicionadas a todos os seus pais.

### 2.2.3 Método RNM

Em Fenton, Neil e Caballero [14] é apresentado o método RNM. Neste trabalho, é introduzido o conceito de nós ranqueados, variáveis aleatórias ordinais representadas em uma escala contínua ordenada monotonicamente no intervalo  $[0, 1]$ . Por exemplo, para a escala ordinal [“Baixo”, “Médio”, “Alto”], “Baixo” é representado pelo intervalo  $[0, 1/3]$ , “Médio”, pelo intervalo  $[1/3, 2/3]$ , e “Alto”, pelo intervalo  $[2/3, 1]$ . Este conceito é baseado na distribuição Normal (TNormal) duplamente truncada.

Uma distribuição normal é feita de quatro parâmetros:  $\mu$ , média (i.e., tendência central);  $\sigma^2$ , variância (i.e., incerteza sobre a tendência central);  $a$ , o limite inferior (i.e., 0);  $e$ ,  $b$ ,

limite superior (i.e., 1). Com uma distribuição normal é possível modelar uma variedade de curvas (i.e., relacionamentos) como uma distribuição uniforme, obtida quando  $\sigma^2 = \infty$ , e distribuições altamente distorcidas, obtidas quando  $\sigma^2 = 0$ . A Figura 2.5 apresenta um exemplo de TNormal com  $\mu$  igual e  $\sigma^2$  diferente.

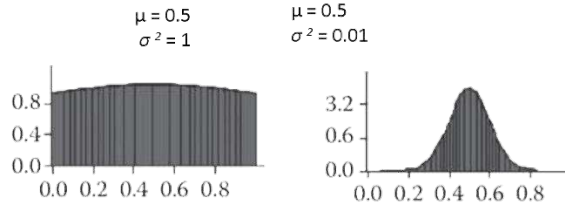


Figura 2.5: Exemplos de TNormal

Nesse método,  $\mu$  é definido por uma função ponderada dos nós-pais. Existem quatro tipos de funções: média ponderada (*MEAN*) Eq. (2.9), mínimo ponderado (*WMIN*) Eq. (2.10), máximo ponderado (*WMAX*) Eq. (2.11) e uma mistura das funções MIN e MAX (*MIXMINMAX*) Eq. (2.12). Na prática, essas funções definem a tendência central do nó filho para a combinação de estados de nós pais. O peso de cada nó pai, que quantifica sua influência sobre o nó filho, deve ser definido por uma constante  $w$  em que  $w \in \mathbb{N}$ .

$$WMEAN(z_1, k, \dots, z_n, k, w_1, \dots, w_n) = \frac{\sum_{i=1}^n w_i * z_i, k}{\sum_{i=1}^n w_i} \quad (2.9)$$

$$WMIN(z_1, k, \dots, z_n, k, w_1, \dots, w_n) = \min_{i=1, \dots, n} \left\{ \frac{w_i * z_i, k + \sum_{j \neq i} z_j, k}{w_i + n - 1} \right\} \quad (2.10)$$

$$WMAX(z_1, k, \dots, z_n, k, w_1, \dots, w_n) = \max_{i=1, \dots, n} \left\{ \frac{w_i * z_i, k + \sum_{j \neq i} z_j, k}{w_i + n - 1} \right\} \quad (2.11)$$

$$MIXMINMAX(z_1, k, \dots, z_n, k, WMIN, WMAX) = \frac{WMIN * \min_{i=1, \dots, n} z_i, k + WMAX * \max_{i=1, \dots, n} z_i, k}{WMIN + WMAX} \quad (2.12)$$

Em Fenton, Neil e Caballero [14] não são apresentados os detalhes para, na prática, implementar a solução. Apesar de apresentar as funções de mistura, não há informações

sobre os algoritmos usados para gerar e misturar a TNormal, definir o tamanho das amostras e definir uma TPN convencional, considerando as TNormals calculadas. A definição de um TPN convencional permite a integração de nós ranqueados com outros tipos de nós, como booleanos e contínuos, o que traz mais flexibilidade de modelagem.

Em Nunes et al. [38], é apresentado um algoritmo probabilístico para este fim, composto de duas etapas principais: (i) gerar amostras para os nós pais e (ii) construir o TPN. Na etapa (ii), para cada possível combinação de valores dos nós pais (i.e., cada coluna do TPN), as amostras definidas na etapa anterior são misturadas usando a função selecionada pelo especialista em RNM e uma TNormal é gerada a partir do resultado da mistura e uma variância, também definida pelo especialista em RNM. O algoritmo é limitado as funções *WMEAN*, *WMIN* e *WMAX*. Uma visão geral do algoritmo é apresentado na Figura 2.6.

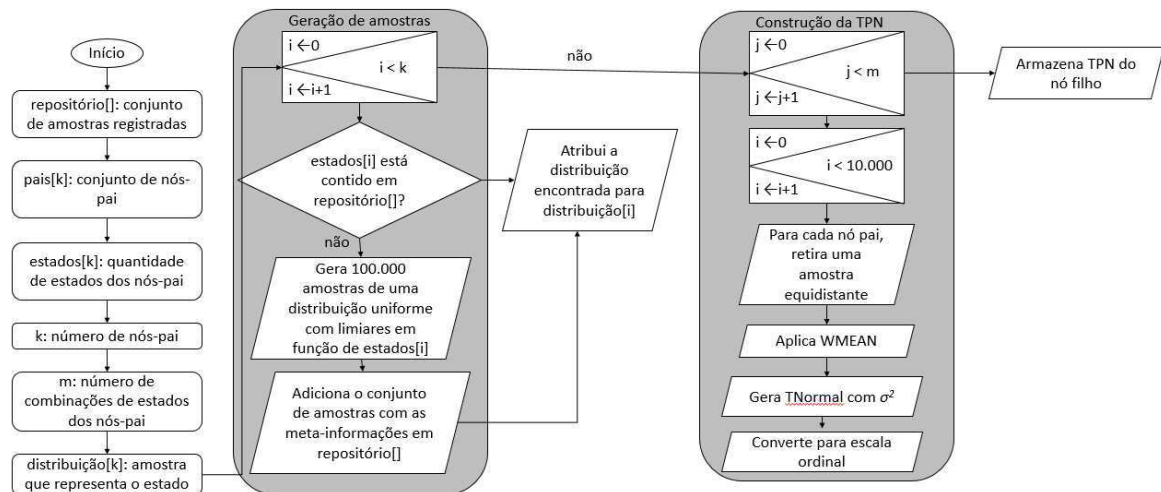


Figura 2.6: Algoritmo para definição de TPN

Como já mencionado, um nó ranqueado é representado conceitualmente por uma escala ordinal, que é mapeada para o intervalo contínuo  $[0, 1]$ . Assim, é representado como um conjunto de distribuições uniformes. Para uma escala ordinal com três valores (e.g., “Ruim”, “Moderado” e “Bom”):  $\cup (0,1) = p_{\text{bad}} * \cup (0,1 / 3) \cup p_{\text{moderado}} * \cup (1 / 3, 2 / 3) \cup p_{\text{good}} * \cup (2 / 3, 1)$ , onde  $p$  é a densidade da distribuição.

Para o exemplo apresentado na Figura 2.7 o conjunto de distribuições uniformes é composto da união de três distribuições uniformes:  $\cup (0,1) = 54,7 * \cup (0,1 / 3) \cup 36,5 * \cup (1 / 3, 2 / 3) \cup 8,80 * \cup (2 / 3, 1)$ . Numericamente, essa união é calculada usando amostras. Considerando um tamanho de amostra de 10000, para representar a TPN do exemplo apresentado



na Figura 2.7, é necessário coletar 5700 amostras aleatórias de  $\cup (0,1 / 3)$ , 3650 amostras aleatórias de  $\cup (1 / 3,2 / 3)$  e 880 amostras aleatórias de  $\cup (2 / 3,1)$ .

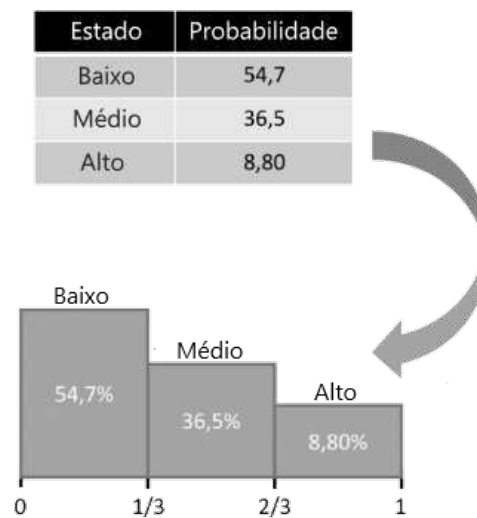


Figura 2.7: Conversão de escala ordinal para contínua

O algoritmo é composto por quatro coleções, conforme podemos observar na Figura 2.6: repositório[], um vetor para armazenar as amostras dos estados base dos nós pais; pais[k], um vetor para armazenar referências dos nós pais; estados[k], um vetor para armazenar os estados do nó filho, dado as combinações de estados dos seus pais; e distribuição[k], um vetor para armazenar as distribuições de probabilidade.

O algoritmo aplica uma estratégia de *cache* com o intuito de otimizar sua execução. Primeiro, são registradas na memória (i.e., no repositório[]) distribuições que representam os estados base, que são estados com evidência (i.e., um nó tem 100% de chance para um dado estado). Por exemplo, para um nó composto pelos estados [“Ruim”, “Moderado”, “Bom”], são registradas amostras para: 100% “Ruim”, 100% “Moderado” e 100% “Bom”, que tem respectivamente  $\mu = 1/6$ ,  $\mu = 1/2$  e  $\mu = 5/6$ . Para este propósito, amostras de uma distribuição uniforme com os limites definidos de acordo com os limiares da escala são coletadas.

Assim, para 100% “Bom”, são coletadas amostras de uma distribuição uniforme limitada no intervalo  $[2/3, 1]$ . Em Nunes et al. [38] é empiricamente definido que usar um tamanho de amostra de 10000 é suficiente para garantir uma margem de erro menor que 0,1%. Cada amostra é registrada com metadados referentes à sua configuração (i.e., número de estados e  $\mu$ ).

Os dados no repositório[] são usados para gerar amostras para um nó, portanto, as amostras para os estados base são geradas apenas uma vez e reutilizadas posteriormente. A próxima etapa consiste em: para cada combinação dos nós pai, misturar a TNormal usando amostras equidistantes, selecionadas aleatoriamente para cada nó pai. As amostras são misturadas usando uma das funções ponderadas (i.e., *WMEAN*, *WMIN* ou *WMAX*) e a variância definida.

Para misturar as distribuições, um elemento aleatório de cada amostra dos pais é removido e usado para calcular um elemento resultante usando uma das função ponderadas. Neste sentido, considere o nó A com dois pais, B e C, ao calcular as probabilidades de A para a combinação “Ruim” - “Bom” com a função *WMEAN* e pesos iguais (dos nós pais), considerando que os valores coletados em uma iteração são 0,1 e 0,7, o valor resultante será 0,4. Esse passo deve ser repetido até que as coleções de amostras estejam vazias.

Depois, o conjunto de elementos calculados e a  $\sigma^2$  são usados como entrada para gerar uma TNormal. A distribuição resultante é convertida em uma escala ordinal e representa uma coluna na TPN do nó filho (no exemplo dado, a coluna para a combinação “Ruim” - “Bom”). No final deste passo, todas as combinações possíveis de estados dos nós pais são avaliadas e a TPN do nó filho é concluída.

Assim, as entradas para gerar a TPN de um nó filho são: uma expressão ponderada capaz de gerar curvas equivalentes às distribuições esperadas pelos especialistas; um conjunto de pesos dos nós pai; e um valor para  $\sigma^2$ . Para determinar a expressão ponderada, pode-se pedir aos especialistas que avaliem a moda do nó filho para diferentes combinações de estados extremos dos nós pais [26]. Consideremos a rede Bayesiana apresentada na Figura 2.2, juntamente com moda das distribuições esperadas pelos especialistas apresentada na Tabela 2.8.

Primeiro, vamos considerar as linhas 1 e 8, em que todos os nós pais estão nos estados mais alto e mais baixo, respectivamente. Como pode ser visto na Tabela 2.8, quando todos os nós pais estão nos estados mais baixo ou mais alto, a moda do nó filho também está no estado mais baixo ou mais alto, sendo que tal distribuição de probabilidade pode ser obtida por qualquer uma das expressões ponderadas.

Agora, vamos considerar a linha 1 como o estado inicial, as linhas 2, 3 e 5 indicam que, quando o estado de um único nó pai muda do estado mais baixo para o mais alto, a moda do

Casos	País			Trabalho em Equipe				
	Autonomia	Coesão	Colaboração	Muito Baixo	Baixo	Médio	Alto	Muito Alto
1	Muito Baixa	Muito Baixa	Muito Baixa	*				
2	Muito Baixa	Muito Baixa	Muito Alta			*		
3	Muito Baixa	Muito Alta	Muito Baixa				*	
4	Muito Baixa	Muito Alta	Muito Alta				*	
5	Muito Alta	Muito Baixa	Muito Baixa		*			
6	Muito Alta	Muito Baixa	Muito Alta			*		
7	Muito Alta	Muito Alta	Muito Baixa				*	
8	Muito Alta	Muito Alta	Muito Alta					*

Tabela 2.8: Tabela verdade utilizada para descobrir a expressão ponderada

Tabela verdade contendo avaliação da moda da distribuição probabilística esperada pelos especialistas para diferentes combinações de estados de país do nó "Trabalho em Equipe". Estrela marca o estado que deve apresentar a moda da distribuição.

nó filho muda em direção ao estado mais alto. Da mesma forma, considere a linha 8 como o estado inicial, as linhas 4, 6 e 7 indicam que quando o estado de um único nó pai muda do estado mais alto para o mais baixo, a moda do nó filho também muda em direção ao estado mais baixo.

No entanto, é bastante claro que o efeito de deslocamento é mais forte quando ocorre do menor para o maior estado. Portanto, a Tabela 2.8 revela que a moda do nó filho tende a ir mais para os estados mais altos do que os mais baixos, o que torna a função  $WMAX$  mais adequada para expressar a distribuição esperada pelos especialistas.

O processo para determinar os pesos dos nós pais e a variância não é tão simples quanto determinar a expressão ponderada. Não há nenhum método na literatura, até onde se sabe, que ajude nessa tarefa. No entanto, pode-se usar as avaliações da moda (ver Tabela 2.8) como ponto de partida para definir o peso dos nós pais. Considerando que  $WMAX$  é a função mais adequada para expressar a distribuição probabilística, vamos examinar as linhas nas quais os estados mudam do menor para o maior valor na Tabela 2.8.

Dessa forma, considerando a linha 1 como o estado inicial, as linhas 2, 3 e 5 indicam que os nós pais possuem diferentes forças de influência sobre o nó filho, ou seja, quando o nó pai "Autonomia" muda do estado mais baixo para o mais alto, a moda do nó "Trabalho em Equipe" ou "Colaboração" muda ligeiramente para os estados mais altos; no entanto,

a mudança é maior quando o estado muda no nó pai “Coesão”, como pode ser visto, se compararmos as linhas 2, 3 e 5. Um efeito semelhante é observado ao comparar as linhas 4, 6 e 7. Portanto, é derivado da Tabela 2.8 a seguinte restrição em relação ao nível de influência dos nós pais sobre o nó filho: “Autonomia” < “Colaboração” < “Coesão”. No entanto, tentativas e erros ainda são necessários para descobrir valores adequados para os pesos e variância.

Por fim, o método RNM resolve o problema de magnitude da construção de TPN em redes Bayesianas complexas. Por outro lado, uma desvantagem deste, é que o contexto do domínio precisa se encaixar a um padrão que possa ser modelado por uma das expressões ponderadas. Este método já foi validado através de estudos de caso em diferentes domínios do mundo real, tais como gestão de recursos humanos em projetos de software [11], previsão de qualidade de software [13], controle de tráfego aéreo [37] e gerenciamento operacional [36].

# Capítulo 3

## Estudo de Caso

Neste capítulo são apresentados os resultados da comparação entre dois métodos de geração semiautomática de TPN, RNM e WSA.

### 3.1 Visão Geral

Trata-se de um estudo de caso com unidades múltiplas de análise nas quais decisões realizadas por *stakeholders* durante reuniões de tomada de decisão para seleção de funcionalidades (i.e., Caso A) e seleção de *designs* de interfaces de usuário (i.e., Caso B) em um contexto de VBSE, são utilizadas para construir e validar modelos de estimativa de valor baseados em redes Bayesianas.

O objetivo do estudo foi avaliar os métodos RNM e WSA aplicando-os em um cenário real. Os métodos foram utilizados para quantificar as incertezas de dois modelos de estimativa de valor. Para a realização do mesmo foram seguidas as diretrizes propostas por Runeson and Höst [44]. Sendo realizadas as seguintes etapas: (i) preparação do estudo, (ii) seleção de cenários, (iii) seleção de especialistas participantes, (iv) coleta de dados, (v) síntese dos dados, e (vi) apresentação do resultados. As referidas etapas e uma visão geral do estudo de caso são apresentadas na Figura 3.1.

As atividades realizadas foram as seguintes:

1. Instrumentação do estudo de caso e preparação do protocolo de coleta de dados;
2. Seleção das unidades de análise;

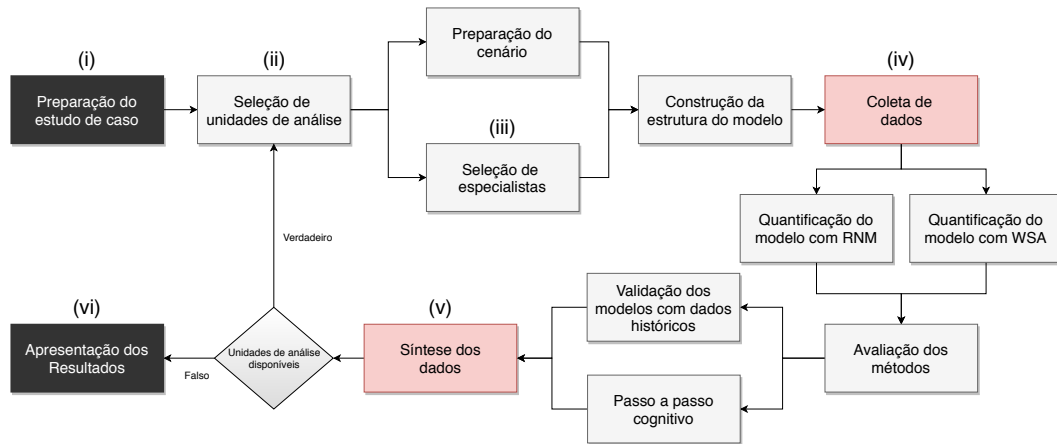


Figura 3.1: Visão geral do estudo de caso

3. Preparação dos cenários (i.e., construção do DAG base de cada unidade de análise);
4. Seleção de especialistas de domínio;
5. Construção das estruturas dos modelos usando o conhecimento dos especialistas de domínio (i.e., incrementação do DAG base);
6. Quantificação das incertezas com os métodos semiautomáticos (i.e., construção das TPN);
7. Validação dos modelos criados;
8. Compilação e análise dos resultados;
9. Apresentação dos resultados.

A atividade 1 está relacionada com a etapa (i); as atividades 2 e 3 estão relacionadas com as etapas (ii); a atividade 4 está relacionada com a etapa (iii); a atividade 5 está relacionada com as etapas (ii) e (iii); as atividades 6 e 7 estão relacionadas com a etapa (iv); a atividade 8 está relacionada com a etapa (v); e a atividade 9 está relacionada com a etapa (vi) que converge na escrita deste Capítulo.

Para construção da estrutura das redes Bayesianas, quantificação e validação (atividades 5, 6 e 7), o processo de Engenharia de Conhecimento de Rede Bayesiana (KEBN) apresentado em Mendes [30] foi adotado.

## 3.2 Motivação

Os processos de tomada de decisão na indústria de software relacionados ao desenvolvimento e gerenciamento de produtos são, em parte, baseadas em um ambiente neutro de valor no qual são considerados apenas fatores de curto prazo (e.g., custo, estimativa de esforço) [4, 31] e seus artefatos (e.g., requisitos, casos de uso) são tratados como igualmente importantes.

Além disso, muitas vezes os fatores representam apenas preocupações técnicas, ou seja, fatores de interesse de outros grupos de *stakeholders* (e.g., consumidores, marketing) são ignorados. Por outro lado, para que empresas de inovação se mantenham competitivas é preciso migrar de processos de tomada de decisão baseados apenas em fatores de curto prazo para processos de tomada de decisão baseados em fatores de curto e longo prazo, em conjunto. Essa necessidade é o cerne do VBSE, idealizada por Barry Boehm [4].

Em Freitas et al. [17] foi apresentado um estudo de caso múltiplo<sup>1</sup>, realizado no Laboratório Embedded - um laboratório de desenvolvimento de software localizado na Universidade Federal de Campina Grande, Brasil - em que os passos A1 e A2 do *framework* VALUE foram executados em três projetos distintos (uma breve descrição deste *framework* é dada a seguir). Neste trabalho investigou-se até que ponto uma ferramenta web pode apoiar *stakeholders* no processo de tomada de decisões em um contexto de VBSE.

O estudo de caso aqui descrito é classificado como um estudo de caso incorporado [46] com múltiplas unidades de análise (i.e., Caso A e Caso B). Os dados coletados em Freitas et al. [17] foram utilizados em conjunto com o conhecimento de especialistas de domínio para construir e validar modelos de estimativa de valor, viabilizando a comparação dos métodos RNM e WSA, utilizados no processo de quantificação de incertezas (i.e., construção das TPN) dos modelos em um contexto de VBSE.

## 3.3 Metodologia

Participaram deste estudo de caso, três especialistas de redes Bayesianas, dois especialistas do método RNM e um especialista do método WSA. Participaram ainda, quatro especialistas de domínio (e.g., gerentes de projetos e desenvolvedores). A criação da estrutura da rede

---

<sup>1</sup>Referenciado ao longo do texto como estudo de caso base.

Bayesiana e sua quantificação foi realizada usando uma combinação de dados históricos e conhecimento dos especialistas de domínio.

A metodologia aplicada para construir e validar as redes Bayesianas de estimativa de valor está em alinhamento com o *framework* VALUE [32]. Esse *framework* inclui uma abordagem de métodos mistos, em que o conhecimento tácito de *stakeholders* em relação a fatores usados durante um processo de tomada de decisão são registrados e aplicados como entrada em uma ferramenta web. Essa ferramenta é usada por esses mesmos *stakeholders* para apoiar seus processos de tomada de decisão. Uma breve descrição do *framework* é dado na próxima subseção.

O estudo de caso aqui descrito realiza os passos A3 e A4, dando continuidade a dois dos três casos (i.e., unidades de análise) do estudo de caso base. Os dados utilizados são referentes aos fatores (ver Tabelas 3.1 e 3.2) e tomadas de decisões extraídos do banco de dados da ferramenta Value [17, 31, 32]. Estes dados são resultantes dos passos A1 e A2 do estudo de caso base. Os passos A3 e A4 (detalhados em seguida) consistem na geração semiautomática de um modelo probabilístico e sua validação, respectivamente.

### 3.3.1 Framework VALUE

O *framework* VALUE foi proposto em Mendes et al. [32]. VALUE significa “improVing decision-mAking reLating to software-intensive prodUcts and sERvices”. Seu objetivo principal é apoiar empresas de software que adotam VBSE a melhorar seus processos de tomada de decisões baseada em valor, em relação ao gerenciamento e desenvolvimento de produtos de software. Uma breve descrição dos passos que compõe o *framework* é dada a seguir:

- A1 - Descobrir fatores de valor específicos da empresa: esse passo consiste na extração de proposições de valor dos *stakeholders* (i.e., fatores de valor). Seu objetivo é tornar explícito a todos os *stakeholders* um conjunto de fatores de valor, para que eles tenham uma visão mais ampla dos fatores considerados importantes pelos diversos perfis de *stakeholders* envolvidos.
- A2 - Empregar A1 (com suporte de uma ferramenta) em reuniões de tomada de decisão: esse passo é caracterizado pelo uso de uma ferramenta Web, denominada Value, para apoiar processos de tomada de decisão, permitindo que diferentes *stakeholders*



empreguem seus fatores de valor ao participarem de decisões, e avaliem o impacto que cada um dos itens que estão sendo decididos (e.g., requisitos, funcionalidades) possa ter sobre cada um dos fatores de valor previamente identificados. Mais detalhes sobre a ferramenta Value são dados em [16, 17, 32].

- A3 - Geração semiautomática de um modelo probabilístico de estimativa de valor: esse passo do *framework* compreende o uso de dados de decisões passadas, armazenados no banco de dados de decisões da ferramenta Value, conhecimento dos *stakeholders* e um algoritmo de geração semiautomática de TPN para construir e validar um modelo (i.e., rede Bayesiana) específico para a empresa. O uso do conhecimento dos *stakeholders* para construir e quantificar o modelo está em alinhamento com o processo KEBN [29, 32].
- A4 - Validação do modelo de estimativa de valor: esse passo valida o modelo de estimativa de valor resultante do passo A3, que determina a necessidade ou não de refazer os passos anteriores. Dois métodos de validação diferentes são geralmente utilizados: *Model Walkthrough* (i.e., passo a passo do modelo) e *Outcome Adequacy* (i.e., análise da acurácia da predição utilizando dados históricos) [32].
- A5 - O Modelo de estimativa de valor é aplicado em reuniões de tomada de decisão: esse passo constitui a última parte do *framework* e corresponde ao uso do modelo como um componente adicional para a ferramenta Value, a ser empregado como parte de um processo de tomada de decisão.

### 3.3.2 Contexto

O estudo de caso foi realizado no Laboratório Embedded. O laboratório possui aproximadamente 60 colaboradores. As equipes de desenvolvimento são compostas por desenvolvedores júnior (que trabalham em meio período) e gerentes de projetos com experiência anterior no setor.

Os projetos são executados em parceria com outras empresas (seus clientes) para desenvolver produtos de software ou protótipos. Os desenvolvedores seguem processos maduros de desenvolvimento de software e os projetos são executados seguindo um processo baseado

em Scrum [47]. As equipes possuem entre três e cinco desenvolvedores e contam com o suporte de uma equipe de garantia da qualidade de software (SQA) e uma equipe de experiência de usuário (UX).

A equipe de SQA é responsável por avaliar a qualidade dos processos e entregas das equipes de desenvolvimento e executar testes independentes (e.g., desempenho, segurança). A equipe de experiência do usuário é responsável por definir o design das interfaces de usuário e fluxos de interação dos produtos. O laboratório produz softwares para plataforma web e para dispositivos móveis em diversos contextos, como embarcados, computação pervasiva e finanças.

### 3.3.3 Pergunta de Pesquisa

O objetivo deste estudo de caso foi comparar os métodos WSA e RNM com o propósito de avaliar a acurácia dos métodos no que se refere à tarefa de construção de TPN de redes Bayesianas, do ponto de vista de especialistas de domínio em um contexto de VBSE. Este objetivo levou às seguintes questões de pesquisa:

1. Qual o método que apresenta maior acurácia na tarefa de construção semiautomática de TPN de redes Bayesianas para estimativa de valor em um contexto de VBSE?

### 3.3.4 Seleção de Casos

Os critérios utilizados na seleção dos casos (i.e., unidades de análise do caso base, que aplicaram os passos A1 e A2 do *framework* VALUE) foram: i) existência de dados suficientes para aplicar os métodos de validação planejados; ii) disponibilidade dos especialistas de domínio que participaram do estudo de caso base. Dois desses casos preencheram esses critérios e foram selecionados. Uma breve descrição do contexto desses casos e seus resultados é dada a seguir.

### 3.3.4.1 Caso A

Um grupo de três *stakeholders* – dois gerentes de projeto e um *product owner*<sup>2</sup> usaram uma ferramenta web, denominada Value<sup>3</sup>, para selecionar funcionalidades para a próxima sprint em dois projetos de Internet das Coisas. Os dados foram coletados durante o desenvolvimento de dois aplicativos, um para controlar remotamente a rega de plantas e outro para monitorar e controlar o nível e fluxo de água em tanques. Os dois projetos foram desenvolvidos para a plataforma Android.

O resultado do passo A1 foi a identificação de 13 proposições de valor (fatores de valor na ferramenta Value), apresentados em tom claro na Tabela 3.1. Três *stakeholders* usaram a ferramenta Value para selecionar funcionalidades para duas sprints, sprint 9 e sprint 10. Eles avaliaram 10 funcionalidades para a sprint 9 e 7 para a *sprint* 10.

### 3.3.4.2 Caso B

Um grupo de cinco *stakeholders* – um *product owner*, três desenvolvedores e um web designer usaram a ferramenta Value para selecionar designs de interface de usuário para implementação. Os dados foram coletados durante o desenvolvimento de uma aplicação web para o gerenciamento de projetos com apoio à melhoria contínua do processo de desenvolvimento de software de equipes ágeis (e.g., *Scrum*).

O resultado do passo A1 para o Caso B foram 10 proposições de valor, apresentados em tom claro na Tabela 3.2. Cinco *stakeholders* usaram a ferramenta Value para selecionar designs de 18 telas para implementação, realizadas em três reuniões de decisão.

## 3.3.5 Seleção de Participantes

Os participantes deste estudo são especialistas de domínio que participaram dos estudos de caso base. O critério utilizado na seleção foi: disponibilidade de horário para participar dos estudos de caso. No que se refere ao Caso A, dos três especialistas de domínio que participaram dos estudos de caso base, dois foram selecionados. Com relação ao Caso B,

---

<sup>2</sup>*Product Owner* é a pessoa que define os itens que compõem o *Product Backlog* e os prioriza nas reuniões de planejamento da *Sprint*.

<sup>3</sup>Ao longo deste documento, referências ao termo VALUE dizem respeito ao *framework*, enquanto referências a Value dizem respeito a uma ferramenta web que compõe o *framework* VALUE.

<b>Grupo</b>	<b>Fator de Valor</b>	<b>Descrição</b>
Aprendizagem	F1	Funcionalidade envolve a implementação de sensores
	F2	Curva de aprendizagem
	F3	Motivação da equipe
Cliente	F4	Funcionalidade faz parte do núcleo do software
	F5	Importância da funcionalidade para o cliente
	F6	Retorno sobre o investimento
	F7	Viabilidade para se tornar um produto
Técnico	F8	Conformidade com o cronograma do sprint
	F9	Parte de uma funcionalidade já iniciada
	F10	Artefatos existentes relacionados à interface do usuário
	F11	Complexidade da funcionalidade
	F12	Resultados da última sprint pendentes
	F13	Funcionalidade faz parte do núcleo
Representação do Modelo	F14	Perspectiva do desenvolvedor
	F15	Sucesso da sprint
	F16	Valor geral

Tabela 3.1: Fatores de valor do Caso A (adaptado de [17])

<b>Grupo</b>	<b>Fator de Valor</b>	<b>Descrição</b>
Negócios	F1	Atende aos requisitos da interface de usuário
Técnico	F2	Complexidade de implementação
	F3	Disponibilidade de bibliotecas de terceiros
	F4	Potencial de reuso
Experiência de Usuário	F5	Cores correspondem à paleta de cores do projeto
	F6	Não possui muitos itens na interface do usuário
	F7	Segue o fluxo de navegação padrão
	F8	Número de cliques para realizar a tarefa
	F9	Interface de usuário significativa
	F10	Simplicidade da interface de usuário
Representação do Modelo	F11	Redução de esforço
	F12	Usabilidade
	F13	Valor geral

Tabela 3.2: Fatores de valor do Caso B (adaptado de [17])

dos cinco especialistas de domínio que participaram dos estudos de caso base, dois foram selecionados.

### 3.3.5.1 Caso A

Os dados demográficos dos participantes são apresentados na Tabela 3.3. A idade média dos participantes é de 28,3 (desvio padrão = 1,53) anos. Eles possuem entre um e três anos de experiência profissional com atividades relacionadas ao desenvolvimento de software.

### 3.3.5.2 Caso B

Os dados demográficos dos participantes são apresentados na Tabela 3.4. A idade média dos participantes é de 28,3 (DP = 1,53) anos. Eles possuem entre um e três anos de experiência profissional com atividades relacionadas ao desenvolvimento de software.

<b>Categoria Demográfica</b>	<b>Resultados</b>	
Idade	Média	29
	Desvio Padrão	2.8
Grau de Escolaridade	Doutorando	2
Experiência Profissional	1-3 anos	1
	4-6 anos	1
Papel no Projeto	Gerente de Projetos	2

Tabela 3.3: Dados demográficos dos participantes do estudo de caso A

<b>Categoria Demográfica</b>	<b>Resultados</b>	
Idade	Média	26
	Desvio Padrão	4.24
Grau de Escolaridade	Doutorando	1
	Graduando	1
Experiência Profissional	1-3 anos	1
	4-6 anos	1
Papel no Projeto	Gerente de Projetos	1
	Desenvolvedor	1

Tabela 3.4: Dados demográficos dos participantes do estudo de caso B

### 3.3.6 Treinamento

Antes de começar a construção da estrutura da rede Bayesiana, os especialistas participaram de um treinamento no qual um exemplo simples foi utilizado para explicar os conceitos básicos sobre redes Bayesianas, o objetivo do estudo de caso, e qual a contribuição necessária dos especialistas. O exemplo utilizado no treinamento foi de um modelo simples de estimativa de valor, que teve como objetivo fornecer uma exposição preliminar aos especialistas de conceitos como “estrutura causal”, cenários de “e se”, e pensamento probabilístico. O exemplo utilizado no treinamento é apresentado na Figura 3.2. A duração do treinamento foi fixada em 20 minutos para os participantes das duas unidades de análise.

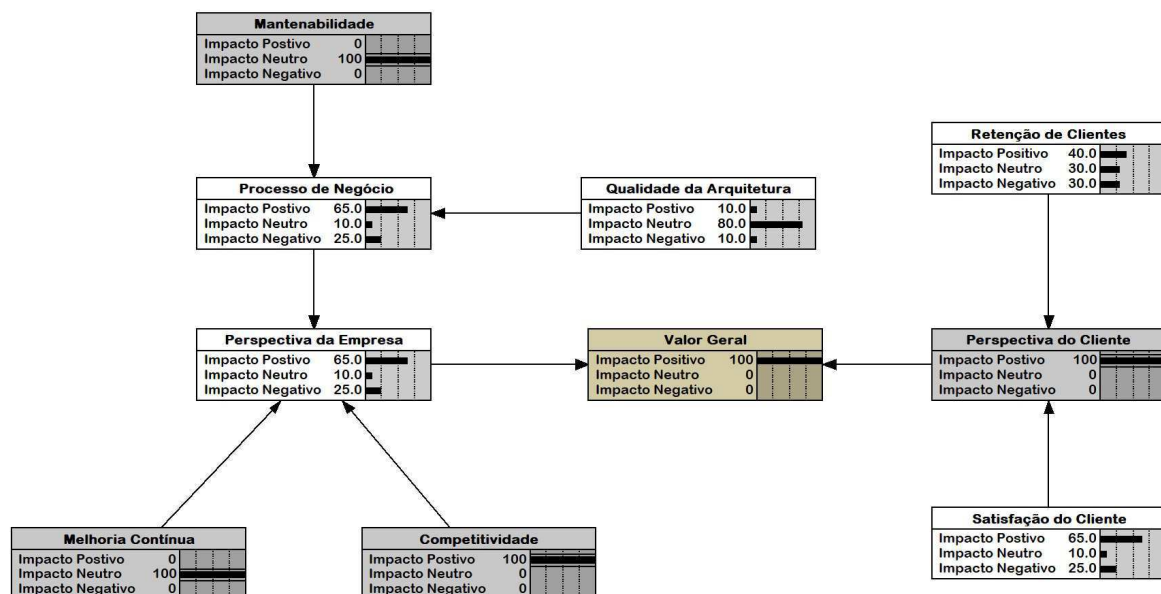


Figura 3.2: Exemplo de rede Bayesiana para estimativa de valor

## 3.4 Coleta de Dados

A coleta de dados para responder à pergunta de pesquisa consiste na construção e validação das redes Bayesianas. Sua execução está em alinhamento com o *framework* VALUE, mais especificamente, com os passos A3 e A4. No passo A3, os dados históricos de tomadas de decisões e informações fornecidas diretamente pelos especialistas foram utilizados para construir a estrutura da rede Bayesiana e quantificá-la.

Os dados referentes a frequência de tomadas de decisão registradas no banco de dados da ferramenta Value foram utilizadas para obter as probabilidades a priori dos nós folha (i.e., nós que não possuem pais na rede Bayesiana). Os métodos de geração semiautomática de TPN foram usados para gerar TPN de nós com mais de um pai.

No passo A4, as redes Bayesianas construídas foram validadas com dois métodos de validação: passo a passo do modelo e acurácia de previsão (i.e., acurácia de predição em relação a dados históricos). Os dados obtidos na etapa de validação foram utilizados para comparar a acurácia dos métodos de geração de TPN. Em seguida, os métodos utilizados para validar as redes Bayesianas construídas são detalhados.

#### 3.4.0.1 Passo a passo do modelo

O método passo a passo do modelo tem como objetivo obter uma avaliação subjetiva da rede Bayesiana. Os especialistas de domínio utilizam cenários hipotéticos para checar se os resultados calculados fazem sentido. O que significa dizer que com este método é possível medir o quão bom determinado modelo é no que diz respeito à sua capacidade de fornecer ou medir o que supostamente deveria fornecer ou medir.

Na prática, isso significa que se as saídas da rede Bayesiana para o fator “Valor Geral” parecerem plausíveis aos olhos dos especialistas de domínio (i.e., o estado de maior probabilidade do fator “Valor Geral” corresponde ao que os especialistas de domínio “sentem” que deveria ser o caso) em diferentes cenários hipotéticos, a rede Bayesiana é validada.

#### 3.4.0.2 Acurácia de previsão

A avaliação da acurácia de previsões, por sua vez, usa dados históricos de reuniões passadas para construir cenários relacionadas a vários itens de decisão (e.g., escolha de funcionalidades ou *designs* de tela para implementação) e confronta os resultados destes cenários com dados calculados no modelo.

Na prática, a abordagem é similar à aplicada no passo a passo do modelo, de modo que para cada cenário (é explicado mais adiante como cada cenário é construído), verifica-se se o estado (e.g., “Alto”, “Médio”, “Baixo”) do fator “Valor Geral” com maior probabilidade corresponde às decisões finais tomadas pelos *stakeholders* em relação aos itens de decisão. De um modo geral, cada item de decisão constitui um cenário.



### 3.4.1 Caso A

#### 3.4.1.1 Construção da Estrutura

Os especialistas de domínio usaram a ferramenta Netica<sup>4</sup>, um software de redes Bayesianas, que já havia sido pré-carregado com nós representando todos os fatores de valor previamente identificados (ver Tabela 3.1), para definir as relações causais que eles consideravam adequadas. Essas relações são apresentadas na Figura 3.3 como setas apontando de um nó para outro.

De acordo com Mendes [32], em quaisquer rede Bayesianas de estimativa de valor a ser construída, é necessário incluir um nó adicional chamado “Valor Geral”. Esse nó representa a saída da rede Bayesianas. No contexto deste estudo de caso, o “Valor Geral” exibe valores (i.e., probabilidades) para os estados “Baixo”, “Médio” ou “Alto” de acordo com cada funcionalidade avaliada para implementação em determinada sprint.

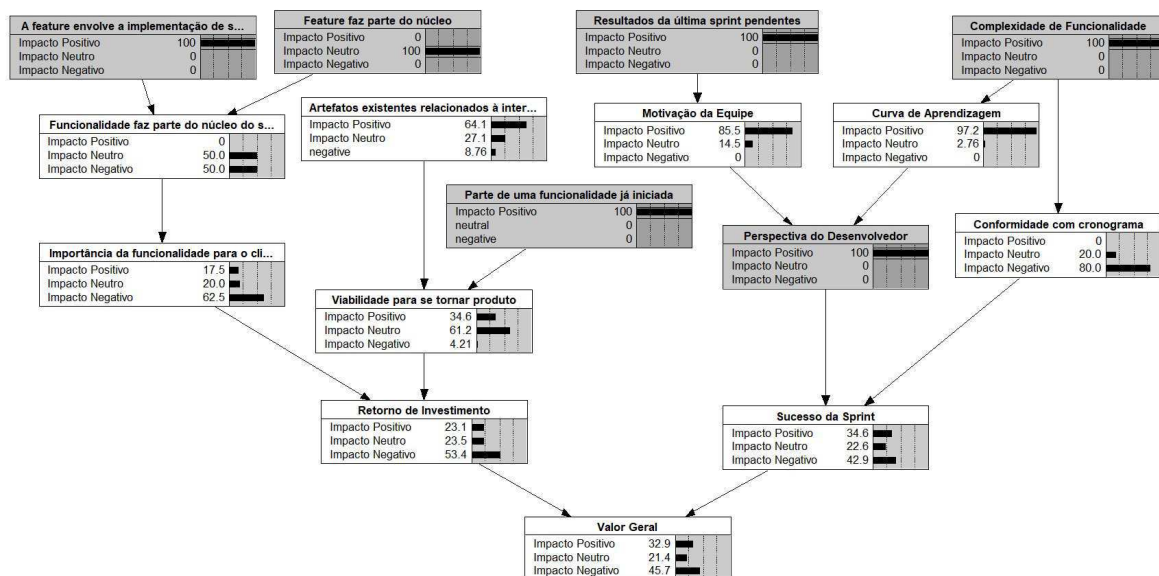


Figura 3.3: Rede Bayesiana construída no Caso A

Dessa forma, ao decidir sobre as relações de causa e efeito, os especialistas de domínio adicionaram o nó “Valor Geral” e mais dois nós – “Perspectiva do Desenvolvedor” e “Sucesso da Sprint”. Estes dois últimos não são fatores de valor em si, mas, nós que os especialistas julgaram importantes por melhorar a representação do próprio conhecimento no

<sup>4</sup><https://www.norsys.com/netica.html>

modelo.

### 3.4.1.2 Quantificação da Rede Bayesiana

Nesta etapa, foram utilizados os dados da sprint 9, com 385 registros de decisões referentes à seleção de 10 funcionalidades, para extrair as probabilidades a priori de seis nós pais: “Funcionalidade envolve implementação de sensores”; “Funcionalidade faz parte do núcleo da aplicação” (outras funcionalidades dependem dela); “Artefatos existentes relacionados à interface do usuário”; “Parte de uma funcionalidade já em desenvolvimento”; “Pendente dos resultados da última sprint”; e “Complexidade da funcionalidade”.

As probabilidades a priori foram obtidas a partir da contabilização da frequência de avaliações de itens de decisão em relação aos fatores. A Figura 3.4 apresenta um dos nós da rede Bayesiana de estimativa de valor para exemplificar o processo de extração das probabilidades a priori. Com base nas avaliações realizadas para 10 funcionalidades que foram discutidas na sprint 9, cada frequência (23%, 14%, 63%) representa, respectivamente, a porcentagem total em que uma funcionalidade foi avaliada como de “Impacto Positivo”, “Impacto Neutro” e “Impacto Negativo” em relação ao fator de valor “Complexidade da funcionalidade”. Essa distribuição de frequência sugere que a maioria das funcionalidades que foram avaliadas para a sprint 9 aparentemente aumentaram a complexidade de desenvolvimento do produto.

Complexidade da Funcionalida...	
Impacto Positivo	23.0
Impacto Neutro	14.0
Impacto Negativo	63.0

Figura 3.4: Nó pai com frequência contabilizada

Os métodos automáticos de geração de TPN foram utilizados para construir as TPN dos nós filhos. Por conseguinte, foram geradas duas versões distintas do modelo. O primeiro, construído com o método WSA e, o segundo, com o método RNM.

O principal desafio ao aplicar o método WSA está relacionado à capacidade dos especialistas (neste caso, dois gerentes de projeto) de identificar, com algum nível de confiança, as configurações de pais compatíveis (ver Figura 3.5) que são mais significativas na prática. Estudos anteriores usando o método WSA, onde as configurações de pais compatíveis foram obtidas de especialistas [1, 3], mostraram que a tarefa é vista como bastante simples pelos especialistas.

A Figura 3.5 apresenta a tela principal do plugin utilizado para aplicar o método WSA na ferramenta Netica. Do lado esquerdo são exibidos os nós da rede Bayesiana que possuem dois ou mais pais. Do lado direito, as configurações de pais compatíveis e probabilidades do nó selecionado: “Funcionalidade faz parte do núcleo do software”. Este nó possui dois pais: “A funcionalidade envolve a implementação de sensores” e “Funcionalidade faz parte do núcleo”. Na parte inferior da tela, são exibidos os campos referentes ao peso relativo de cada um dos nós pais.

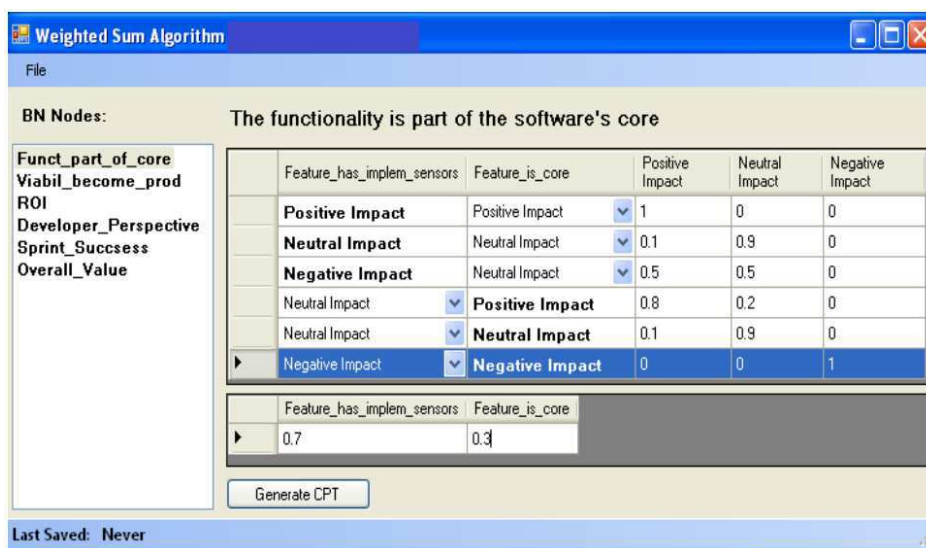


Figura 3.5: Exemplo de dados de entrada para o método WSA (Caso A)

No caso dos nós pais “Funcionalidade envolve a implementação de sensores” e “Funcionalidade faz parte do núcleo”, os especialistas primeiro escolheram os estados mais prováveis para o nó “Funcionalidade faz parte do núcleo” quando o nó “Funcionalidade envolve a implementação de sensores” está nos estados “Impacto Positivo”, “Impacto Neutro” e “Impacto Negativo” (estados em negrito na Figura 3.5).

O mesmo foi feito fixando-se os estados do nó “Funcionalidade faz parte do núcleo”. Neste caso, os especialistas escolheram os estados mais prováveis para o nó “Funcionalidade envolve a implementação de sensores” quando o nó “Funcionalidade faz parte do núcleo” está nos estados “Impacto positivo”, “Impacto neutro” e “Impacto negativo”. Depois de identificar as configurações de pais compatíveis, os especialistas forneceram as probabilidades de cada um dos estados do nó filho “Funcionalidade faz parte do núcleo do software” para cada uma das configurações identificadas.

<b>Nó-Filho</b>	<b>Minutos</b>
Funcionalidade faz parte do núcleo do software	28
Viabilidade de se tornar um produto	24
Retorno do investimento	14
Perspectiva do desenvolvedor	9
Sucesso da Sprint	8
Valor Geral	10

Tabela 3.5: Tempo necessário para aplicar o método WSA (Caso A)

Um exemplo de configuração de pais compatíveis é dado a seguir: “A funcionalidade envolve a implementação de sensores” = “Impacto Positivo”, “Funcionalidade faz parte do núcleo” = “Impacto Positivo”. Os especialistas também informaram um peso relativo para cada um dos nós pais, indicando seu grau de influência sobre o nó filho. Quanto ao peso, a única restrição existente é que a soma de todos os pesos seja igual a 1, conforme observado na seção 2.2.1. Uma vez que todas as configurações de pais compatíveis, pesos e distribuições de probabilidade foram obtidas (i.e., na prática, um subconjunto da TPN), o algoritmo do WSA calculou o restante da TPN.

Usando o método WSA, os especialistas tiveram que fornecer 15 (uma das configurações de pais compatíveis foi selecionada duas vezes) em vez de 27 probabilidades. Embora a diferença entre 15 e 27 não seja grande, há muitas circunstâncias em que as TPN dos nós pais podem exigir a entrada de milhares de probabilidades e o uso de um algoritmo como o WSA ajuda significativamente na redução desse esforço [1].

Os especialistas tiveram que preencher de forma manual as TPN de quatro nós filhos que possuíam apenas um pai - eles forneceram 36 probabilidades no total, em um período de 31 minutos. Por outro lado, o tempo gasto para coletar os parâmetros necessários para gerar as TPN dos seis nós filhos foi de 93 minutos. Detalhes são apresentados na Tabela 3.5.

No que se refere ao método RNM, o primeiro passo foi coletar dos especialistas domínio distribuições de probabilidade de combinações de estados extremos dos nós pais. Planilhas previamente preparadas foram utilizadas para este fim. Os especialistas preencheram as planilhas com os resultados esperados para cada caso. As planilhas tinham o formato do

Nós Pais		Nó Filho		
		Funcionalidade faz parte do núcleo do software		
A funcionalidade envolve a implementação de sensores	A funcionalidade faz parte do núcleo	Impacto Negativo	Impacto Neutro	Impacto Positivo
Impacto Positivo	Impacto Positivo	0	0	1
Impacto Positivo	Impacto Negativo	0	0.3	0.7
Impacto Negativo	Impacto Positivo	0	0.3	0.7
Impacto Negativo	Impacto Negativo	1	0	0

Tabela 3.6: Exemplo de tabela utilizada na coleta de dados dos especialistas para o método RNM (Caso A)

exemplo mostrado na Tabela 3.6.

Ao contrário do método WSA, no método RNM o especialista em redes Bayesianas é mais que um facilitador. É ele que define os parâmetros para o algoritmo do RNM. Dessa forma, o maior desafio para aplicação do método RNM está relacionado à capacidade do especialista em redes Bayesianas de identificar corretamente os parâmetros de entrada do algoritmo RNM (e.g., variância, pesos dos pais e expressão ponderada) para que o algoritmo do RNM possa gerar distribuições de probabilidade semelhantes às esperadas pelos especialistas de domínio para cada uma das combinações de estados dos nós pais.

Os especialistas de domínio ainda são a fonte primária de conhecimento, mas a experiência do especialista em redes Bayesianas é um fator preponderante para a aplicação adequada do método RNM. Sendo assim, o primeiro passo para identificar os parâmetros necessários para executar o RNM consiste, geralmente, em identificar a expressão ponderada que se encaixa com o padrão do domínio a ser modelado. Para isso, pode ser utilizada uma tabela verdade similar à apresentada na Tabela 3.6 para coletar os dados dos especialistas e posteriormente inferir a expressão ponderada, conforme detalhado na seção 2.2.3.

Uma vez identificada a expressão ponderada, o segundo passo consiste em identificar a variância e os pesos dos nós pais para então executar o algoritmo do RNM. Não há na literatura, até onde se sabe, um método bem definido para identificação da variância e dos pesos. Dessa forma, para descobrir os melhores parâmetros o especialista em redes Bayesianas executa várias vezes o algoritmo do RNM variando os parâmetros, em outras palavras, a estratégia aplicada é a de tentativa e erro. Portanto, a quantidade de tentativas e o quão

acertada a escolha dos parâmetros é, depende da experiência do especialista em redes Bayesianas. Além disso, não é possível garantir que os parâmetros identificados sejam os mais adequados.

Por este motivo, neste trabalho foi utilizado uma abordagem denominada “Especialista Simulado” para identificar os pesos dos nós pais e a variância do nó filho. Essa abordagem consiste em um algoritmo que recebe como entrada uma expressão ponderada e distribuições de probabilidade de referência, similar às apresentadas na Tabela 3.6 (i.e., na prática, um subconjunto da TPN) e gera como saída os parâmetros mais adequados para que o algoritmo do RNM gere resultados semelhantes à entrada dada como referência. Essa abordagem foi adotada para reduzir os efeitos que a experiência do especialista em redes Bayesianas poderia ter no estudo de caso.

O funcionamento do algoritmo do “Especialista Simulado” é simples, ele testa várias combinações de peso e variância que são incrementados ou decrementados enquanto executa o algoritmo do RNM e compara o resultado com as distribuições de probabilidade de entrada até que a primeira seja capaz de representar a última da melhor forma possível (i.e., sejam similares). Para medir a similaridade, foi utilizado como métrica o *Brier Score* (BS) (Eq 3.1). Um exemplo dos resultados obtidos com o “Especialista Simulado” é apresentado na Tabela 3.7. Na Tabela 3.8 são apresentados todos os parâmetros que foram utilizados na aplicação do método RNM neste estudo de caso. Detalhes da métrica utilizada pelo algoritmo são dados a seguir:

$$BS = \frac{1}{N} \sum_{t=1}^N \sum_{i=1}^S (f_{ti} - o_{ti})^2 \quad (3.1)$$

Onde:

- $N$  é a quantidade de casos da TPN de referência;
- $S$  é o número de estados do nó filho;
- $f$  é a probabilidade obtida com o algoritmo utilizado no RNM;
- $o$  é a probabilidade de referência (o valor real esperado).

Nós Pais		Nó Filho		
		Funcionalidade faz parte do núcleo do software		
A funcionalidade envolve a implementação de sensores	A funcionalidade faz parte do núcleo	Impacto Negativo	Impacto Neutro	Impacto Positivo
Impacto Positivo	Impacto Positivo	0	0.0033	0.9967
Impacto Positivo	Impacto Negativo	0	0.2995	0.7005
Impacto Negativo	Impacto Positivo	0	0.2995	0.7005
Impacto Negativo	Impacto Negativo	0.9965	0.0035	0

Tabela 3.7: TPN aproximada obtida com o “Especialista Simulado” (Caso A)

Resultado obtido com o “Especialista Simulado” para o nó filho “Funcionalidade faz parte do núcleo do software” usando a expressão ponderada  $WMAX$  com os seguintes parâmetros:  $\sigma^2 = 0.062$  (i.e., variância); e peso 4 para todos os nós pais. O melhor resultado obtido pelo algoritmo foi  $BS = 0.00000394$ .

Nó Filho	Expressão Ponderada	Variância	Peso dos Nós Pais	Brier Score
F16	$wmax$	0.04	$F06*4 + F15*4$	0.005
F04	$wmax$	0.062	$F01*4 + F13*4$	0.00000394
F07	$wmax$	0.1005	$F10*7 + F09*5$	0.00000659667
F06	$wmax$	0.0635	$F05*7 + F07*4$	0.000127617
F14	$wmax$	0.065	$F03*5 + F02*7$	0.0000108683
F15	$wmax$	0.135	$F14*9 + F08*9$	0.0143216

Tabela 3.8: Parâmetros utilizados com o método RNM (Caso A)

A primeira coluna identifica o nó filho. A expressão ponderada foi definida pelo especialista em redes Bayesianas. A variância e o peso dos nós pais foram obtidos com o “Especialista Simulado”. A última coluna da tabela mostra o Brier Score obtido. Quanto menor for o BS, melhor o resultado final.

### 3.4.1.3 Passo a Passo do Modelo

No contexto do Caso A, sete cenários hipotéticos foram preparados pelos especialistas em redes Bayesianas (ver Tabela 3.9). Cada cenário incluiu uma avaliação subjetiva do estado de maior probabilidade para o nó “Valor Geral” de acordo com o que é esperado pelos especialistas. Os cenários foram inseridos como evidência no modelo de estimativa de valor resultante do passo A3 (ver Figura 3.6).

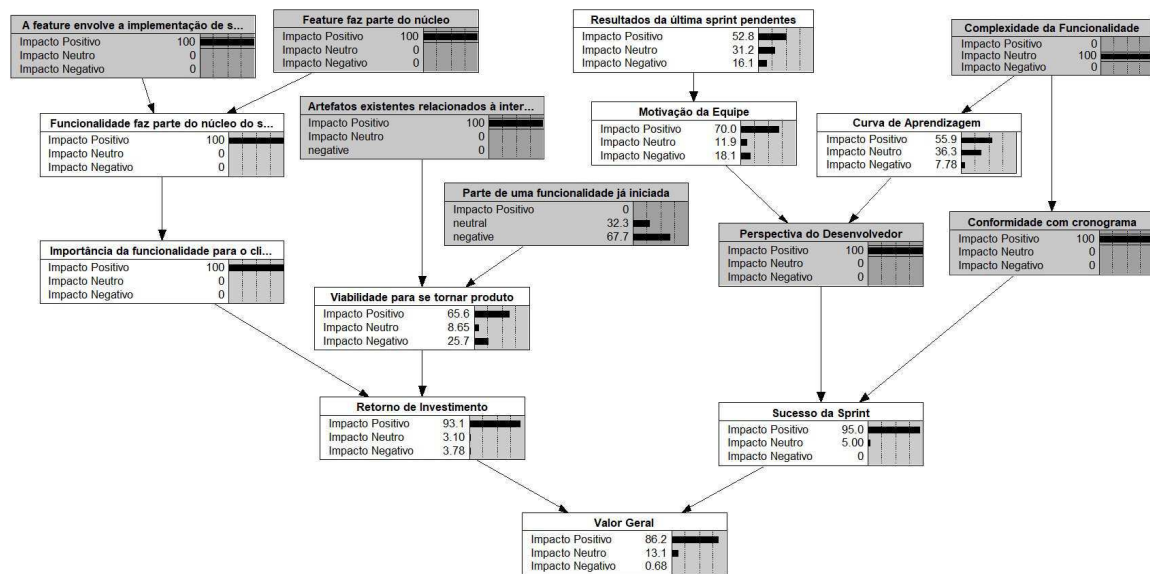


Figura 3.6: Rede Bayesiana alimentada com evidências (Caso A)

A Figura 3.6 apresenta um exemplo de cenário aplicado à rede Bayesiana (i.e., alimentada com um conjunto de evidências). Na Figura 3.6 evidência significa clicar, em um determinado nó, sobre o estado (“Impacto Positivo”, “Impacto Neutro”, “Impacto Negativo”) que corresponde à avaliação dos especialistas de domínio acerca do impacto esperado da implementação de uma determinada funcionalidade sobre algum fator em particular (i.e., o nó que recebe a evidência). Uma vez clicado, a probabilidade relativa a esse estado muda para 100%. Um conjunto de evidências dos fatores para uma funcionalidade em particular constitui um cenário.

### 3.4.1.4 Acurácia de Previsão

Nesta etapa, cenários de validação foram criados a partir de um subconjunto de tomadas de decisões extraído de dados de uma reunião que visava selecionar funcionalidades para



Fatores de Valor	Cenários Simulados						
	1	2	3	4	5	6	7
F1	Positivo	Negativo	Positivo	Negativo	-	Positivo	Negativo
F3	Positivo	Negativo	-	Negativo	-	-	Negativo
F4	Positivo	-	Positivo	Negativo	-	Positivo	Negativo
F5	Positivo	-	Positivo	-	Positivo	Positivo	-
F7	Positivo	-	Positivo	Positivo	Positivo	Positivo	-
F8	Positivo	Negativo	-	-	Positivo	Positivo	-
F9	-	-	-	-	Positivo	-	-
F10	Positivo	Negativo	Positivo	Positivo	Positivo	Positivo	-
F11	-	-	Positivo	-	-	-	Negativo
F12	-	-	Positivo	-	Positivo	-	-
F13	Positivo	Neutro	Positivo	Negativo	-	Positivo	Negativo
F14	Positivo	-	-	-	Neutro	Positivo	Positivo

Tabela 3.9: Cenários simulados para validação do modelo construído no Caso A

implementação na sprint 10, ou seja, foram utilizados dados históricos de projetos finalizados conforme recomendado por Mendes [30]. Os dados utilizados na construção desses cenários não foram usados na Parte A3 deste estudo de caso.

Neste método de avaliação, para cada cenário criado verifica-se o estado (e.g., “Alto”, “Médio”, “Baixo”) de maior probabilidade do nó “Valor Geral” resultante da execução da rede Bayesiana e compara-o com o estado que corresponde a decisão final dos especialistas. O processo de construção de cada cenário é explicado adiante com a ajuda da Tabela 3.10.

Na Tabela 3.10, temos:

- $N$  stakeholders,  $S_n$ , que anotamos:  $S_1, \dots, S_N$
- $M$  Funcionalidades,  $F_m$ , que anotamos:  $F_1, \dots, F_M$
- $L$  Fatores,  $V_l$ , que anotamos:  $V_1, \dots, V_L$

$AO$  representa o resultado agregado de todos os fatores de valor, por funcionalidades, para as avaliações de todos os stakeholders em relação à esta funcionalidade. A avaliação

<b>Dados de uma Reunião de Tomada de Decisão</b>					
$S_n$	$F_m$	$V_1$	$V_2$	...	$V_L$
$S_1$	$F_1$	$k : \text{positivo}$	$k : \text{positivo}$	...	$k : \text{positivo}$
$S_2$	$F_1$	$k : \text{positivo}$	$k : \text{neutro}$	...	$k : \text{negativo}$
...	...	...	...	...	...
$S_N$	$F_1$	$k : \text{positivo}$	$k : \text{neutro}$	...	$k : \text{negativo}$
AO		$k' : \text{positivo}$	$k' : \text{neutro}$	...	$k' : \text{negativo}$
$S_1$	$F_M$	$k : \text{neutro}$	$k : \text{neutro}$	...	$k : \text{neutro}$
$S_2$	$F_M$	$k : \text{neutro}$	$k : \text{neutro}$	...	$k : \text{negativo}$
...	...	...	...	...	...
$S_N$	$F_M$	$k : \text{positivo}$	$k : \text{negativo}$	...	$k : \text{negativo}$
AO		$k' : \text{neutro}$	$k' : \text{neutro}$	...	$k : \text{negativo}$

Tabela 3.10: Matriz utilizada para criar cenários a partir de dados históricos

é definida como “positivo”, “neutro” ou “negativo” de acordo com a maior frequência de  $k$ : positivo/neutro/negativo, considerando todos os *stakeholders* e funcionalidades. Caso haja um empate,  $k$  é definido como o estado menos otimista (negativo > neutro > positivo). Neste estudo de caso, os fatores de valor possuem a mesma importância.

A Tabela 3.11 mostra os cenários extraídos das tomadas de decisões relacionadas à sprint 10, dispostos de acordo com a Tabela 3.10. Na Tabela 3.11 as funcionalidades sombreadas em tom escuro foram selecionadas para implementação na *Sprint 10* e as funcionalidades sombreadas em tom claro foram rejeitadas.

Para validar o modelo, e por conseguinte, os métodos, cada funcionalidade relacionada na Tabela 3.11 constituiu um cenário, de forma que as evidências obtidas a partir do valor de cada célula relacionando os fatores de valor com as funcionalidades foram utilizadas para alimentar a rede Bayesiana. Na prática, no entanto, ao aplicar as propriedades de separação direcional (d-separação) [40] foi observada a existência de apenas 4 cenários distintos dentre os 7 existentes (i.e., os cenários 2, 3, 4 e 5 são iguais). Os nós que se tornaram irrelevantes nos cenários extraídos são descritos em seguida.

Considerando os conjuntos de nós  $X = \{F1, F13, F4, F5, F10, F9, F7\}$ ,  $Y = \{F16\}$

Fatores de Valor	Funcionalidades avaliadas na reunião de tomada de decisão para Sprint 10						
	1	2	3	4	5	6	7
F2	Negativo	Negativo	Negativo	Negativo	Negativo	Negativo	Negativo
F3	Neutro	Negativo	Negativo	Negativo	Negativo	Negativo	Negativo
F6	Positivo	Negativo	Negativo	Negativo	Negativo	Neutro	Positivo
F8	Positivo	Positivo	Positivo	Positivo	Positivo	Positivo	Positivo
F11	Negativo	Negativo	Negativo	Negativo	Negativo	Negativo	Negativo

Tabela 3.11: Dados de decisões tomadas para a Sprint 10, dispostos em conformidade com a Tabela 3.10

e  $E = \{F6\}$ ,  $X$  e  $Y$  são condicionalmente independentes em  $E$ . Em outras palavras, o conjunto de nós  $E$  d-separa os conjuntos de nós  $X$  e  $Y$ , haja vista que todos os caminhos de um nó pertencente a  $X$  para um nó pertencente a  $Y$  são bloqueados dado  $E$  (i.e.,  $F6$  possui evidência em todos os cenários). Dessa forma, os nós pertencentes ao conjunto  $X$  são irrelevantes para o resultado final (i.e., resultado do nó  $F16$ ). De forma análoga, o nó  $F12$  e  $F16$  são condicionalmente independentes dado a existência de evidência no nó  $F3$  em todos os cenários. Portanto, o nó  $F12$ , é também, irrelevante na execução dos cenários.

## 3.4.2 Caso B

### 3.4.2.1 Construção da Estrutura

Para criar a estrutura da rede Bayesiana do Caso B, foram realizados os mesmos passos do Caso A. Dessa forma, os especialistas utilizaram a ferramenta Netica, que havia sido pré-carregado com nós representando todos os fatores de valor previamente identificados (ver Tabela 3.2), para definir as relações de causa e efeito.

Durante a definição das relações, foram acrescentados mais três nós, assim como no caso A. O primeiro nó foi “Valor Geral”, que representa a saída do modelo. Neste estudo de caso, o “Valor Geral” pode ser “Baixo”, “Médio” ou “Alto” e varia de acordo com o design de tela avaliado para implementação. Mais dois nós foram adicionados – “Redução de Esforço” e “Usabilidade”. O objetivo destes nós, assim como no Caso A, foi aumentar a capacidade de representação do modelo.

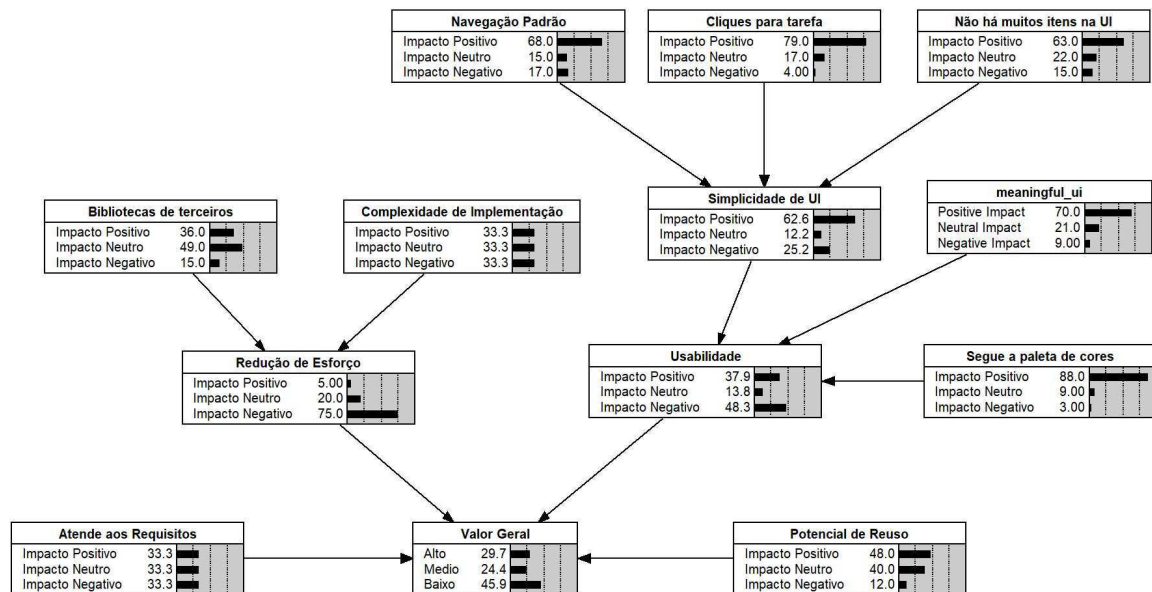


Figura 3.7: Rede Bayesiana construída no Caso B

### 3.4.2.2 Quantificação da Rede Bayesiana

Para quantificar a rede Bayesiana do Caso B, foram utilizados 1037 registros de dados oriundos de cinco reuniões de tomada de decisão. Dos 1037 registros, 744 registros foram utilizados para calibrar a rede Bayesiana e 293 para validá-la. Os registros utilizados para calibrar a rede Bayesiana estão relacionados às reuniões de identificador 4, 18 e 20, nas quais foram avaliados 18 itens. Os dados utilizados para validar a rede Bayesiana estão relacionados as reuniões de identificador 16 e 17, nas quais o objetivo foi selecionar o design para duas telas do produto em desenvolvimento.

Uma vez finalizada a estrutura da rede Bayesiana, os dados de decisão relacionados as reuniões 4, 18 e 20 foram utilizados para preencher as TPN dos nove nós folhas do modelo (i.e., probabilidades a priori) – “Atende aos requisitos da interface do usuário”, “Complexidade de implementação”, “Disponibilidade de bibliotecas de terceiros”, “Potencial de reuso”, “Cores correspondem à paleta de cores do projeto”, “Não possui muitos itens na interface do usuário”, “Segue o fluxo de navegação padrão”, “Número de cliques para realizar a tarefa” e “Acessibilidade da interface de usuário”.

A Figura 3.8 apresenta um dos nós da rede Bayesiana de estimativa de valor. Com base em todas as avaliações feitas sobre os 18 designs de tela discutidos nas reuniões 4, 18 e 20, cada frequência (79%, 17%, 4%) representa, respectivamente, a porcentagem total em que

designs de tela foram avaliados como de “Impacto Positivo”, “Impacto Neutro” e “Impacto Negativo” em relação ao “Número de clicks para realizar tarefa”. Essa frequência sugere que a maioria dos designs avaliados não adicionaram complexidade indesejada ao produto.

Cliques para realizar tarefa	
Impacto Positivo	79.0
Impacto Neutro	17.0
Impacto Negativo	4.00

Figura 3.8: Exemplo de nó com probabilidades a priori

Depois que todas as frequências foram inseridas nas TPN dos nós folha, a próxima etapa foi iniciar a construção das TPN dos quatro nós filhos. Nesta etapa, os métodos semiautomáticos de geração de TPN foram aplicados. Dessa forma, duas versões da rede Bayesiana foram geradas. A primeira versão foi gerada com o método WSA, assim como no Caso A.

A Figura 3.9 apresenta a tela principal do plugin utilizado para aplicar o método WSA neste estudo de caso. Dado os nós pais “Simplicidade da interface do usuário”, “Interface de usuário significativa” e “Cores correspondem à paleta de cores do projeto” o gerente de projeto e o desenvolvedor escolheram os estados mais prováveis para os nós pais “Interface de usuário significativa” e “Cores correspondem à paleta de cores do projeto” quando o nó pai “Simplicidade da interface do usuário” está nos estados “Impacto Positivo”, “Impacto Neutro” e “Impacto Negativo” (estados em negrito na Figura 3.9). O mesmo foi feito fixando os estados dos demais nós pais e escolhendo os estados mais prováveis dos outros dois. Para cada uma das combinações de pais compatíveis identificadas, os especialistas inseriram as probabilidades de cada um dos estados do nó filho “Usabilidade”.

Dado os nós pais “Funcionalidade envolve a implementação de sensores” e “Funcionalidade faz parte do núcleo”, os dois gerentes primeiro escolheram os estados mais prováveis para o nó pai “Funcionalidade faz parte do núcleo” quando o nó-pai “A funcionalidade envolve a implementação de sensores” está nos estados “Impacto Positivo”, “Impacto Neutro” e “Impacto Negativo”. O mesmo foi feito para cada um dos estados do nó pai “Funcionalidade faz parte do núcleo”. Para cada uma das combinações de pais compatíveis identificadas os especialistas de domínio inseriram as probabilidades de cada um dos estados do nó filho “Funcionalidade faz parte do núcleo do software”.

Um exemplo de configuração de pais compatíveis é “A funcionalidade envolve a implementação de sensores” = “Impacto Positivo”, “Funcionalidade faz parte do núcleo” =

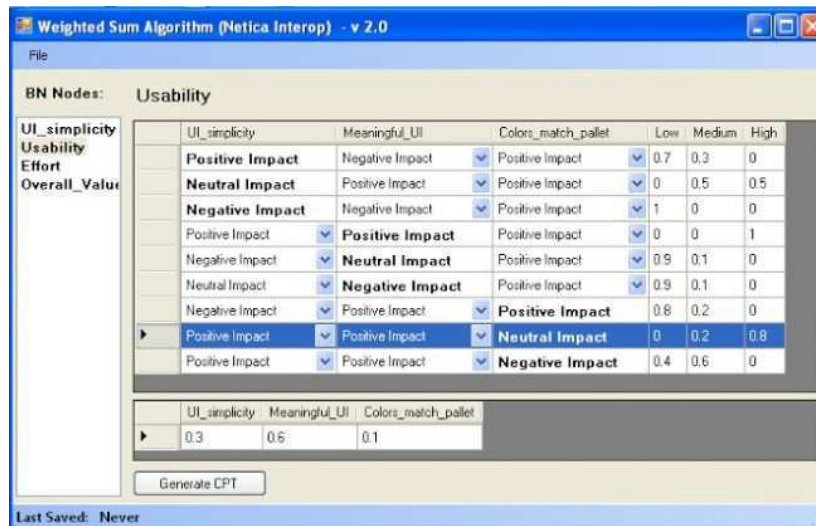


Figura 3.9: Exemplo de dados de entrada para o método WSA (Caso B)

Nó-Filho	Minutos
Simplicidade da interface do usuário	19
Usabilidade	12
Redução do esforço	3
Valor geral	7

Tabela 3.12: Tempo necessário para aplicar o método WSA (Caso B)

“Impacto Positivo”. Os *stakeholders* também escolheram um peso relativo para cada um dos nós pais, indicando seu grau de influência sobre o nó filho. Uma vez que ambos os passos foram realizados, a TPN completa foi gerada usando o algoritmo WSA.

Usando o método WSA, os especialistas de domínio precisaram fornecer 27 em vez de 81 probabilidades. Observe que a diferença agora é mais significativa do que a do exemplo da Figura 3.9. Ou seja, quanto mais complexo for a TPN, maior será o ganho do WSA em relação à redução do esforço necessário quantificá-la.

O tempo gasto para identificar configurações de pais compatíveis, inserir as probabilidades e pesos dos nós pais, para cada uma das quatro TPN dos nós filhos foi 41 minutos. Detalhes são apresentados na Tabela 3.12.

No que se refere ao método RNM, foi utilizada a mesma abordagem aplicada no Caso A. Um exemplo dos dados coletados é mostrado na Tabela 3.13. Um exemplo dos dados

Nós Pais			Nó Filho		
			Simplicidade da interface do usuário		
Número de cliques para realizar a tarefa	Não possui muitos itens na interface do usuário	Segue o fluxo de navegação padrão	Impacto Negativo	Impacto Neutro	Impacto Positivo
Impacto Positivo	Impacto Positivo	Impacto Positivo	0	0	1
Impacto Positivo	Impacto Positivo	Impacto Negativo	0.7	0.3	0
Impacto Positivo	Impacto Negativo	Impacto Positivo	0.7	0.3	0
Impacto Positivo	Impacto Negativo	Impacto Negativo	1	0	0
Impacto Negativo	Impacto Positivo	Impacto Positivo	0.7	0.3	0
Impacto Negativo	Impacto Positivo	Impacto Negativo	0.8	0.2	0
Impacto Negativo	Impacto Negativo	Impacto Positivo	0.8	0.2	0
Impacto Negativo	Impacto Negativo	Impacto Negativo	0.1	0	0

Tabela 3.13: Exemplo de tabela utilizada na coleta de dados dos especialistas do Caso B

gerados pelo “Especialista Simulado” é apresentado na Tabela 3.14.

### 3.4.2.3 Passo a passo do modelo

A validação dos modelos criados no Caso B, com cenários simulados, seguiu o mesmo protocolo do Caso A. Foram preparados oito cenários hipotéticos (ver Tabela 3.16), um a mais em relação ao Caso A.

### 3.4.2.4 Acurácia de Previsão

Nesta etapa, foi utilizado um subconjunto de casos reais extraídos de dados relacionados a tomadas de decisão das reuniões de identificador 16 e 17, com os quais foram gerados os cenários de validação.

A Tabela 3.17 apresenta os dados coletados, dispostos de acordo com a Tabela 3.12. Nas Tabelas 3.17 e 3.21 os itens de decisão sombreadas em verde foram selecionados para implementação e os itens sombreadas em vermelho não foram selecionados.

Assim como no Caso A, alguns fatores de valor (i.e., F6, F7 e F8) se tornaram irrelevantes devido a existência de evidência para o fator de valor F10.

Nós Pais			Nó Filho		
			Simplicidade da interface do usuário		
Número de cliques para realizar a tarefa	Não possui muitos itens na interface do usuário	Segue o fluxo de navegação padrão	Impacto Negativo	Impacto Neutro	Impacto Positivo
Impacto Positivo	Impacto Positivo	Impacto Positivo	0	0.0368	0.9632
Impacto Positivo	Impacto Positivo	Impacto Negativo	0.7291	0.2709	0
Impacto Positivo	Impacto Negativo	Impacto Positivo	0.6932	0.3068	0
Impacto Positivo	Impacto Negativo	Impacto Negativo	0.8874	0.1126	0
Impacto Negativo	Impacto Positivo	Impacto Positivo	0.6932	0.3068	0
Impacto Negativo	Impacto Positivo	Impacto Negativo	0.8874	0.1126	0
Impacto Negativo	Impacto Negativo	Impacto Positivo	0.8767	0.1233	0
Impacto Negativo	Impacto Negativo	Impacto Negativo	0.9664	0.0336	0

Tabela 3.14: TPN aproximada obtida com o especialista simulado (Caso B)

Resultado obtido com o especialista simulado para o nó filho “Simplicidade da interface do usuário”. A expressão ponderada utilizada foi  $WMIN$  com os seguintes parâmetros:  $\sigma^2 = 0.0915$  (i.e., variância); e pesos 9-9-10. O melhor resultado obtido pelo “Especialista Simulado” foi  $BS = 0.00246858$ .

Nó Filho	Expressão Ponderada	Variância	Nós Pais	Brier Score
F13	$wmin$	0.0405	$F01*10 + F12*10 + F04*1 + F11*1$	0.0255071
F11	$wmax$	0.026	$F03*4 + F02*10$	0.000000375
F12	$wmin$	0.072	$F10*10 + F09*7 + F05*5$	0.00208559
F10	$wmin$	0.0915	$F08*9 + F06*9 + F07*10$	0.00246858

Tabela 3.15: Parâmetros utilizados com o método RNM (Caso B)

A primeira coluna identifica o nó filho. A expressão ponderada foi definida pelo especialista em redes Bayesianas. A variância e o peso dos nós pais foram obtidos com o “Especialista Simulado”. A última coluna da tabela mostra o BS obtido. Quanto menor for o BS, melhor o resultado final.



Fatores de Valor	Cenários Simulados							
	1	2	3	4	5	6	7	8
F1	Negativo	Neutro	Neutro	Negativo	Positivo	Negativo	Positivo	Positivo
F2	Positivo	Negativo	Positivo	Negativo	Positivo	Negativo	Positivo	Negativo
F3	-	Negativo	-	-	-	Negativo	-	Neutro
F4	-	Negativo	Neutro	Negativo	Positivo	Negativo	Positivo	Negativo
F6	-	-	-	-	-	Negativo	-	Positivo
F7	-	-	Neutro	-	-	-	-	Neutro
F8	-	-	-	-	-	-	-	Positivo
F9	-	-	Negativo	-	-	Negativo	-	Positivo
F11	-	Baixo	-	-	-	Baixo	-	-

Tabela 3.16: Cenários simulados para validação do modelo construído no Caso B

Fatores de Valor	Designs de tela avaliados nas reuniões de tomada de decisão 16 e 17								
	1	2	3	4	5	6	7	8	9
F1	Positivo	Positivo	Positivo	Positivo	Positivo	Positivo	Positivo	Positivo	Positivo
F2	Neutro	Neutro	Positivo	Negativo	Neutro	Negativo	Neutro	Negativo	Negativo
F3	Neutro	Neutro	Neutro	Negativo	Neutro	Neutro	Neutro	Neutro	Neutro
F6	Neutro	Positivo	Positivo	Positivo	Neutro	Neutro	Positivo	Positivo	Neutro
F8	Positivo	Neutro	Neutro	Neutro	Positivo	Positivo	Neutro	Positivo	Positivo
F7	Positivo	Positivo	Positivo	Negativo	Negativo	Negativo	Positivo	Neutro	Neutro
F4	Neutro	Positivo	Neutro	Positivo	Positivo	Positivo	Positivo	Positivo	Positivo
F9	Positivo	Positivo	Negativo	Positivo	Neutro	Positivo	Positivo	Positivo	Positivo
F10	Positivo	Positivo	Neutro	Positivo	Positivo	Neutro	Positivo	Positivo	Neutro

Tabela 3.17: Dados de decisões tomadas para a Sprint 10, dispostos em conformidade com a Tabela 3.10

## 3.5 Resultados e Discussão

### 3.5.1 Caso A

No que se refere à validação com passo a passo do modelo (i.e., cenários simulados), o método WSA apresentou uma acurácia superior ao método RNM. A Tabela 3.18 mostra os resultados esperados pelos especialistas de domínio em cada um dos cenários hipotéticos criados para validar os modelos, bem como os resultados obtidos pelos métodos de geração de TPN.

Como é possível observar na Tabela 3.18, dos sete cenários criados, três (i.e., cenários 2, 4, 7) apresentaram conflito entre o estado de maior probabilidade esperado pelos especialistas de domínio para o nó “Valor Geral” e o estado de maior probabilidade do nó “Valor Geral” resultante do modelo construído com o método RNM, que apresentou uma precisão de apenas 57,14%. Por outro lado, o modelo construído com o método WSA apresentou uma acurácia de 100%, estimando corretamente o estado de maior probabilidade para o nó “Valor Geral” em todos os cenários considerados.

No que se refere à validação com dados históricos (i.e., acurácia da previsão), os métodos foram considerados equivalentes. A tabela 3.19 mostra os resultados obtidos nos cenários construídos com dados referentes à decisões tomadas na escolha de funcionalidades para a sprint 10. Na Tabela 3.19 as linhas sombreadas em tom escuro indicam as funcionalidades que foram selecionadas e as linhas sombreadas em tom claro representam as funcionalidades rejeitadas. As funcionalidades 2, 3, 4 e 5 obtiveram resultados equivalentes, dado que, na prática, as evidências dos nós que diferenciavam os cenários se tornaram irrelevantes (i.e., ao aplicar as propriedades de d-separação).

Fonte	Estado de maior probabilidade em F15 / Cenários Simulados							Acurácia
	1	2	3	4	5	6	7	
Especialista	Positivo	Negativo	Positivo	Negativo	Positivo	Positivo	Negativo	
WSA	Positivo	Negativo	Positivo	Negativo	Positivo	Positivo	Negativo	7/7
RNM	Positivo	Neutro	Positivo	Positivo	Positivo	Positivo	Positivo	4/7

Tabela 3.18: Resultado da validação com cenários simulados (Caso A)

Funcionalidade	Valor Geral					
	WSA			RNM		
	Baixo	Médio	Alto	Baixo	Médio	Alto
1	17,3%	13,4%	69,3%	0%	0%	100%
2	38%	20%	42%	0%	39%	61%
3	38%	20%	42%	0%	39%	61%
4	38%	20%	42%	0%	39%	61%
5	38%	20%	42%	0%	39%	61%
6	25%	19%	56%	0%	25%	75%
7	17%	13%	70%	0%	0%	100%

Tabela 3.19: Resultado da validação com dados históricos da sprint 10 (Caso A)

De acordo com os resultados, as funcionalidades selecionadas foram, de fato, as que apresentaram maior probabilidade para o estado “Alto” do nó “Valor Geral”, com exceção da funcionalidade 3 que, apesar de equivalente às funcionalidades 2, 3 e 4, ainda foi selecionada. A seleção da funcionalidade 3 talvez pudesse ser explicada por um fator não modelo na rede Bayesiana ou talvez trate-se de uma decisão de exceção. No entanto, dado que a rede Bayesiana serve de suporte a tomada de decisão, os resultados obtidos com os dois métodos ainda são considerados promissores. Um efeito observado que vale a pena mencionar neste estudo é que o modelo construído com o método RNM foi mais otimista do que o modelo construído com o método WSA em todos os casos em que a previsão foi acertada.

### 3.5.2 Caso B

A validação com passo a passo dos modelos criados no Caso B apresentaram um maior equilíbrio entre os métodos, apesar do modelo criado com o método WSA ainda ter apresentado uma acurácia superior ao modelo construído com o método RNM. Na Tabela 3.20 podemos observar que, dos oito cenários criados, quatro (i.e., cenários 1, 2, 3 e 8) apresentaram conflito entre o estado de maior valor esperado pelos especialistas de domínio para o nó “Valor Geral” e o estado de maior valor obtido no modelo construído com o método RNM para o mesmo nó, apresentando uma precisão de 50%. Por outro lado, o método WSA apresentou

Fonte	Estado de maior probabilidade em F15 / Cenários Simulados								Acurácia
	1	2	3	4	5	6	7	8	
Especialista	Baixo	Baixo	Baixo	Baixo	Alto	Baixo	Alto	Baixo	
WSA	Baixo	Alto	Médio	Baixo	Alto	Baixo	Alto	Baixo	6/8
RNM	Médio	Médio	Médio	Baixo	Alto	Baixo	Alto	Médio	4/8

Tabela 3.20: Resultado da validação com cenários simulados (Caso B)

Design de Tela	Valor Geral					
	WSA			RNM		
	Baixo	Médio	Alto	Baixo	Médio	Alto
1	77%	3%	20%	0%	11%	89%
2	21%	3%	76%	0%	6%	94%
3	77%	14%	9%	0%	100%	0%
4	3%	39%	58%	0%	11%	89%
5	34%	31,5%	34,48%	0%	46%	54%
6	24%	6%	70%	0%	95%	5%
7	21%	3%	76%	0%	6%	94%
8	24%	4%	72%	0%	11%	89%
9	24%	6%	70%	0%	95%	5%

Tabela 3.21: Resultado da validação com dados históricos (Caso B)

uma precisão de 75%, estimando corretamente o estado de maior valor para o nó “Valor Geral” em seis dos oito cenários criados (i.e., apresentou conflito nos cenários 2 e 3).

No que se refere à validação com dados históricos, os métodos foram considerados equivalentes. A Tabela 3.21 apresenta os resultados obtidos nos cenários construídos com dados de decisões referentes à escolha de designs de tela para implementação em determinado produto de software. Na Tabela 3.21, as linhas sombreadas em tom escuro indicam os designs aprovados para implementação e as linhas sombreadas em tom claro representam os designs rejeitados. Os dois designs de tela selecionadas tiveram a maior probabilidade para o estado “Alto” em ambos os métodos.

### 3.6 Ameaças à Validade

Uma das ameaças deste estudo de caso está relacionada à medida escolhida para mensurar valor, haja vista que não foi possível utilizar uma escala numérica; ou seja, a escolha foi entre uma escala nominal ou ordinal. Os especialistas de domínio escolheram medir o valor como “Valor de Impacto”, medido em uma escala ordinal de três pontos. Não obstante, o processo que foi usado para construir e validar os modelos de estimativa de valor já foram empregados em outros estudos e obtiveram bons resultados [1,3,29]. Além disso, a validação dos modelos, e por conseguinte, dos métodos, foi baseada não apenas na entrada direta de especialistas de domínio (i.e., passo a passo cognitivo do modelo), mas também usando dados históricos de reuniões de tomada de decisões.

Outra ameaça à validade está relacionada a não homogeneidade do ambiente em que foram aplicados os métodos. Embora o método WSA tenha sido aplicado com a presença dos especialistas em redes Bayesianas, que atuaram como facilitadores, os dados para aplicação do método RNM foram obtidos de forma remota, por meio de planilhas eletrônicas que foram utilizadas para coletar os dados de entrada para o “Especialista Simulado”, a partir do qual foram gerados os parâmetros de entrada para o algoritmo do RNM. Por fim, o conjunto de fatores de valor e os modelos de estimativa de valor são soluções específicas da empresa. Portanto, os resultados estão vinculados ao contexto da empresa para a qual eles foram feitos. No entanto, eles talvez possam ser aplicáveis a outras empresas que desenvolvem aplicações similares àsquelas aqui detalhadas.

# Capítulo 4

## Experimento

Neste capítulo, são apresentados os resultados de um estudo empírico comparativo entre métodos de geração semiautomática de TPN. Os métodos comparados foram: RNM, WSA, AHP e uma combinação dos métodos WSA e AHP.

### 4.1 Definição

O estudo realizado é do tipo experimental, e teve como objetivo analisar métodos de geração semiautomática de TPN com a intenção de compará-los a respeito de acurácia e capacidade de modelagem de TPN do ponto de vista de especialistas em redes Bayesianas no contexto de quantificação de incertezas na avaliação da coesão de equipes ágeis de desenvolvimento de software. Os dados foram coletados de especialistas de domínio que utilizaram os métodos e definiram manualmente TPN de referência para comparar os resultados dos mesmos.

#### 4.1.1 Objetivos específicos

1. Comparar a acurácia dos métodos semiautomáticos de geração de TPN;
2. Comparar a capacidade de modelagem dos métodos semiautomáticos de geração de TPN.

## 4.2 Visão Geral

Este experimento foi realizado em quatro etapas para cumprir os objetivos definidos na Seção 4.1. A Figura 4.1 apresenta um diagrama com uma visão geral das atividades realizadas.

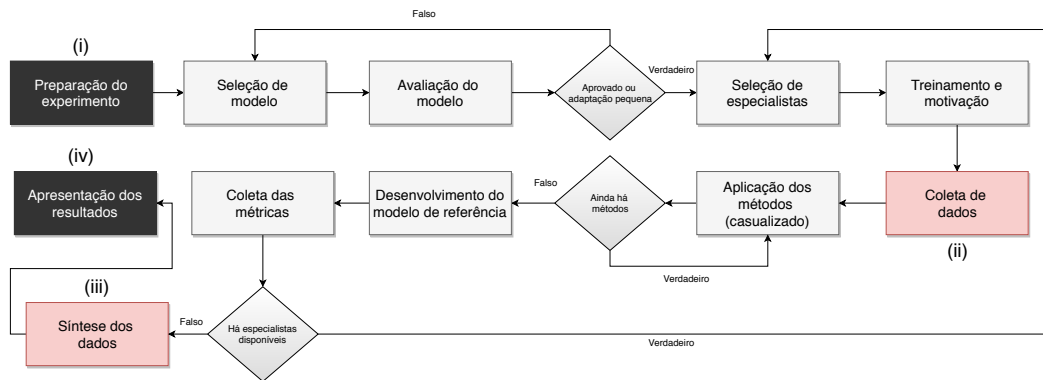


Figura 4.1: Visão geral das atividades realizadas no experimento

As atividade realizadas foram:

1. Planejamento e instrumentação;
2. Seleção do modelo;
3. Avaliação do modelo;
4. Seleção de especialistas;
5. Treinamento e motivação;
6. Aplicação dos métodos;
7. Desenvolvimento do modelo de referência;
8. Coleta de métricas;
9. Análise e interpretação dos dados;
10. Apresentação dos resultados.

As atividades 1, 2, 3 e 5 estão relacionadas com a etapa (i), as atividades 6, 7 e 8 estão relacionados com a etapa (ii), atividade 9 está relacionada com a etapa (iii), e finalmente a atividade 10 que está relacionada com a etapa (iv) e converge na escrita deste capítulo.

## 4.3 Contexto

Os métodos de geração semiautomática de TPN foram aplicados por alunos graduandos do curso de Ciência da Computação, que atuam como desenvolvedores júnior em projetos de desenvolvimento de software no Laboratório Embedded – um laboratório de desenvolvimento de software localizado na Universidade Federal de Campina Grande, Brasil.

No Laboratório Embedded, os projetos são executados em parceria com outras empresas para desenvolver produtos de software ou protótipos. Os desenvolvedores seguem processos de desenvolvimento de software e os projetos são executados seguindo um processo baseado em Scrum.

Os desenvolvedores participantes deste experimento utilizaram os métodos semiautomáticos de geração de TPN para quantificar incertezas em uma rede Bayesiana destinada à avaliação e melhoria da coesão de equipes ágeis de desenvolvimento de software, adaptada de Freire et al. [15].

## 4.4 Planejamento

Nesta seção, as questões de pesquisa e hipóteses, a rede Bayesiana utilizada, a seleção de participantes, o design do experimento e a instrumentação do mesmo são descritos em detalhes.

### 4.4.1 Questões de pesquisa e hipóteses

Conforme sugere a discussão apresentada em Zagorecki e Druzdzal [52], não há uma medida padrão estabelecida para medir a similaridade de TPN. No entanto, há um conjunto de medidas que são geralmente utilizadas para este fim. Neste estudo são utilizadas três dessas medidas de similaridade e uma medida de proporção para comparar a acurácia das TPN geradas pelos métodos semiautomáticos. Dessa forma, as perguntas de pesquisa são:

*PI.* A capacidade de modelagem de TPN dos métodos semiautomáticos é diferente?

O que nos leva às seguintes hipóteses:



$H1_0$ : A capacidade de modelagem de TPN dos métodos semiautomáticos é a mesma quando as TPN são comparadas a partir do *Brier Score*;

$H1_1$ : A capacidade de modelagem de TPN dos métodos semiautomáticos é diferente quando as TPN são comparadas a partir do *Brier Score*;

$H2_0$ : A capacidade de modelagem de TPN dos métodos semiautomáticos é a mesma quando as TPN são comparadas usando a Distância Euclidana;

$H2_1$ : A capacidade de modelagem de TPN dos métodos semiautomáticos é diferente quando as TPN são comparadas usando a Distância Euclidana;

$H3_0$ : A capacidade de modelagem de TPN dos métodos semiautomáticos é a mesma quando as TPN são comparadas usando o Erro Médio Absoluto;

$H3_1$ : A capacidade de modelagem de TPN dos métodos semiautomáticos é diferente quando as TPN são comparadas usando o Erro Médio Absoluto.

Caso a hipótese nula seja refutada - indicando que há diferença - as médias das medidas de similaridade serão analisadas para que se possa estabelecer o método que proporciona a maior capacidade de modelagem de TPN.

**P2.** A acurácia dos métodos semiautomáticos é diferente?

O que nos leva às seguintes hipóteses:

$H4_0$ : A acurácia dos métodos semiautomáticos é a mesma quando consideramos a proporção de acertos dos métodos;

$H4_1$ : A acurácia dos métodos semiautomáticos é diferente quando consideramos a proporção de acertos dos métodos.

No que se refere a acurácia, caso a hipótese nula seja refutada - indicando que há diferença - a proporção de acertos será analisada para que se possa estabelecer o método que proporciona maior acurácia, de um ponto de vista prático.

### 4.4.2 Definição Formal das Hipóteses

Formalmente, as quatro hipóteses descritas anteriormente podem ser definidas conforme a Tabela 4.1:

<b>Hipótese</b>	<b>Hipótese Nula</b> ( $I$ é o número de métodos)	<b>Hipótese Alternativa</b> (para pelo menos um $i$ )
$H1$	$H1_0 : BS(t_1) = BS(t_2) = \dots = BS(t_I) = 0$	$H1_1 : BS(t_i) \neq 0$
$H2$	$H2_0 : ED(t_1) = ED(t_2) = \dots = ED(t_I) = 0$	$H2_1 : ED(t_i) \neq 0$
$H3$	$H3_0 : EM(t_1) = EM(t_2) = \dots = EM(t_I) = 0$	$H3_1 : EM(t_i) \neq 0$
$H4$	$H4_0 : PA(t_1) = PA(t_2) = \dots = PA(t_I) = 0$	$H4_1 : PA(t_i) \neq 0$

Tabela 4.1: Hipóteses do experimento

### 4.4.3 Seleção do Modelo

A rede Bayesiana utilizada neste experimento trata-se de uma adaptação do modelo proposto por Freire et al [15]. A mesma foi simplificada para adequar-se a este experimento (i.e., de acordo com a análise prévia do perfil dos possíveis participantes e do contexto do experimento), a fim de fornecer um conjunto de etapas que são realizáveis e significativas para os especialistas de domínio participantes do experimento.

A rede Bayesiana adaptada contém duas TPN com dois pais (baixa complexidade) e uma com três pais (média complexidade). Estas TPN são menos suscetíveis a vieses dos especialistas de domínio quando comparadas à TPN de alta complexidade (i.e., no que diz respeito à definição manual). Entretanto, apesar do tamanho, estas TPN são consideradas não triviais. Uma TPN é considerada não trivial quando possui, pelo menos, dois nós pais [2,50]. A rede Bayesiana utilizada no experimento é apresentada na Figura 4.2.

### 4.4.4 Estudo Piloto

Antes da execução do experimento, um estudo preliminar foi realizado. Este estudo teve os seguintes objetivos:

1. Avaliar a rede Bayesiana adaptada de Freire et al. [15];

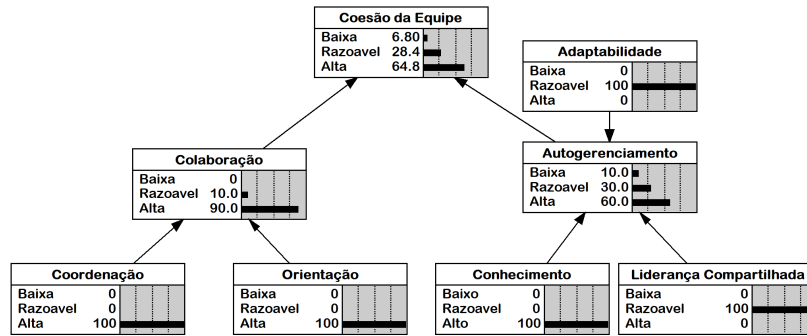


Figura 4.2: Rede Bayesiana utilizada no experimento

- Validar a aplicação do método RNM sem especialista em redes Bayesianas (i.e., os parâmetros de entrada do algoritmo do RNM são dados por um “Especialista Simulado”).

Este estudo foi realizado com alunos da disciplina Projeto em Computação I - que tiveram seu primeiro contato com o trabalho em equipe em um contexto de desenvolvimento de softwares. Os métodos aplicados foram RNM e WSA.

#### 4.4.4.1 Metodologia

Os dados dos participantes foram coletados remotamente por meio de planilhas eletrônicas. Na primeira etapa, foi disponibilizado um formulário online para recrutar voluntários, contendo os seguintes campos: nome; e-mail; e projeto desenvolvido durante a disciplina. O formulário foi divulgado ao término da disciplina no canal de comunicação da disciplina.

Na segunda etapa, links individuais foram enviados via e-mail para cada um dos participantes, promovendo acesso aos seguintes artefatos: planilha para preenchimento de probabilidades de combinações de estados extremos (i.e., dados de entrada do “Especialista Simulado”, que tem como saída os parâmetros do algoritmo do RNM); planilha para preenchimento de configurações de pais compatíveis e pesos dos nós pais (i.e., dados de entrada para o método WSA); vídeo explicativo com informações e instruções sobre como os dados deveriam ser preenchidos, bem como toda a semântica relacionada às variáveis da rede Bayesiana; uma planilha com cenários simulados para validação da rede Bayesiana.

A Figura 4.3a mostra a planilha utilizada para coletar os dados dos participantes, para, então, aplicar o método RNM. Neste caso em específico, o participante deveria raciocinar acerca da influência da “Expertise” e da “Liderança Compartilhada” na “Auto-Organização”

da equipe. Na Figura 4.3b é possível observar um exemplo de combinações de pais compatíveis com distribuições de probabilidade e peso dos nós pais, denotando os dados de entrada para o método WSA.

RNM		Resultados Esperados			STATUS
Expertise	Liderança Compartilhada	Baixa	Média	Alta	
Alta	Alta	0,1	0,15	0,75	válido
Alta	Baixa	0,3	0,45	0,25	válido
Baixa	Alta	0,25	0,45	0,3	válido
Baixa	Baixa	0,85	0,1	0,05	válido

WSA		Resultados Esperados			STATUS
Expertise	Liderança Compartilhada	Auto-Organização			
		Baixa	Média	Alta	
Baixa	Alta	0,5	0,25	0,25	válido
Média	Média	0,33	0,33	0,34	válido
Alta	Baixa	0,3	0,35	0,35	válido
Expertise	Liderança Compartilhada	Baixa	Média	Alta	
Alta	Baixa	0,3	0,35	0,35	válido
Média	Média	0,33	0,33	0,34	válido
Baixa	Alta	0,5	0,25	0,25	válido

Pesos		STATUS
Expertise	Liderança Compartilhada	
0,5	0,5	válido

(a) Dados de entrada para o método RNM

(b) Dados de entrada para o método WSA

Figura 4.3: Planilhas utilizadas na coleta de dados no estudo piloto

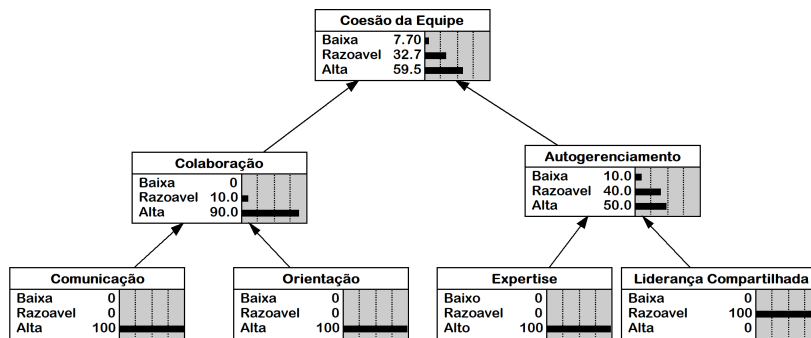


Figura 4.4: Rede Bayesiana utilizada no estudo piloto

Para validar as redes Bayesianas geradas a partir dos dados de entrada extraídos dos participantes, foi utilizado o método de cenários simulados (i.e., passo a passo do modelo). Conforme podemos observar na Tabela 4.2, foram criados 10 cenários hipotéticos. Estes cenários foram utilizados para comparar as estimativas calculadas por cada uma das redes Bayesianas construídas e os resultados esperados por cada um dos participantes do estudo.

**4.4.4.2 Resultados e Discussão**

O experimento piloto contou com a participação de cinco alunos, oriundos de quatro equipes de desenvolvimento de software (i.e., Buzzticket; Alocca; Hipster; e Event Planner). Na Tabela 4.3 é possível observar os resultados dos tratamentos aplicados. A acurácia neste estudo piloto é dado pelo número de acertos dividido pelo número de cenários. Para ilustrar

Cenário	Evidência dos nós pais				Resultado Esperado
	Comunicação	Expertise	Liderança Compartilhada	Orientação	Coesão da Equipe
1	<i>Alta</i>	<i>Alta</i>	<i>Alta</i>	<i>Alta</i>	?
2	<i>Alta</i>	<i>Alta</i>	<i>Baixa</i>	<i>Alta</i>	?
3	<i>Alta</i>	<i>Média</i>	<i>Alta</i>	<i>Baixa</i>	?
4	<i>Baixa</i>	<i>Baixa</i>	<i>Média</i>	<i>Baixa</i>	?
5	<i>Alta</i>	<i>Alta</i>	<i>Média</i>	<i>Média</i>	?
6	<i>Baixa</i>	<i>Alta</i>	<i>Média</i>	<i>Média</i>	?
7	<i>Média</i>	<i>Média</i>	<i>Média</i>	<i>Média</i>	?
8	<i>Alta</i>	<i>Alta</i>	<i>Baixa</i>	<i>Baixa</i>	?
9	<i>Baixa</i>	<i>Baixa</i>	<i>Alta</i>	<i>Média</i>	?
10	<i>Baixa</i>	<i>Baixa</i>	<i>Baixa</i>	<i>Baixa</i>	?

Tabela 4.2: Cenários simulados (Estudo Piloto)

os resultados, tomemos como exemplo o participante número 1, na primeira linha da Tabela 4.3.

Diante dos resultados obtidos, é possível observar que o tratamento com RNM produziu um modelo que estimou corretamente 4 dos 10 cenários hipotéticos, e portanto, alcançou uma acurácia de 40%. O segundo tratamento, com o método WSA, resultou em um modelo com uma maior capacidade de representação do conhecimento do participante 1, alcançando uma acurácia de 60%.

De forma geral, os resultados obtidos ajudaram a identificar mudanças necessárias ao modelo antes da realização do experimento principal (e.g., necessidade de um nó com três pais, alterações de nós e mudança de nomenclatura). A Figura 4.2 já reflete as alterações aplicadas. Ademais, os resultados validaram o uso do modelo no experimento principal (i.e., com alterações), pois são similares aos resultados obtidos no estudo de caso apresentado no Capítulo 3.

A abordagem com “Especialista Simulado” foi adotada no estudo de caso para descobrir a variância e o pesos dos nós pais (i.e., ainda havia necessidade de um especialista em

Participantes	Número de Acertos		Precisão	
	RNM	WSA	RNM	WSA
1	4	5	40%	50%
2	8	6	80%	60%
3	6	7	60%	70%
4	8	6	80%	60%
5	8	3	80%	30%

Tabela 4.3: Resultado dos Cenários Simulados (Estudo Piloto)

redes Bayesianas para definir a expressão ponderada). Neste estudo piloto, o “Especialista Simulado” também foi utilizado para descobrir a expressão ponderada (i.e., eliminando a necessidade do especialista em redes Bayesianas). Os resultados validaram esta abordagem.

#### 4.4.5 Seleção dos Participantes

Os participantes são alunos do curso de Ciência da Computação da UFCG, que atuam como desenvolvedores júnior em projetos de desenvolvimento de software no Laboratório Embedded. A participação no experimento foi voluntária. A única restrição para participação foi possuir experiência mínima de três meses com a metodologia ágil de desenvolvimento de software, Scrum. O tempo disponibilizado ao experimento por parte dos participantes, foi convertido em horas de trabalho contabilizadas na carga horária semanal dos mesmos no Laboratório Embedded.

No contexto deste experimento, os desenvolvedores são caracterizados como especialistas de domínio. Participaram deste experimento quatro especialistas de domínio. Na Tabela 4.4 é apresentado os dados demográficos dos participantes. Eles possuem entre 6 e 24 meses de experiência com a metodologia ágil, Scrum.

#### 4.4.6 Delineamento do Experimento

Neste experimento, o fator de interesse  $F$  são os métodos de geração semiautomática de TPN com quatro níveis: WSA, RNM, AHP e WSA-AHP. A acurácia e similaridade das TPN geradas pelos métodos, medida com diferentes métricas, é a variável resposta (i.e.,

Categoria Demográfica	Resultados	
Idade	Média	22.2
	Desvio Padrão	1.6
Grau de Escolaridade	Graduando	10
Experiência Profissional	6-12 meses	4
	13-24 meses	4
	25-36 meses	2
Papel no Projeto	Desenvolvedor	9
	Testador	1

Tabela 4.4: Dados demográficos dos participantes do experimento

variável dependente).

Toda e qualquer variável que possa interferir na variável resposta ou dependente deve ser mantida constante ou ter seu efeito mitigado sempre que possível. Neste experimento, a variável que poderia interferir na variável resposta é a experiência dos participantes em redes Bayesianas. Em outras palavras, o fator-estorvo, ou fonte de variação indesejada  $X$ , é caracterizado como a experiência dos participantes em relação a redes Bayesianas.

A blocagem é uma alternativa útil para reduzir a influência de fontes de variação sistemáticas e melhorar a eficiência de experimentos - por meio do arranjo de unidades experimentais em grupos homogêneos. Portanto, para cancelar o efeito de  $X$ , foi adotado um delineamento em blocos casualizados (DBC) com balanceamento. Dessa forma, o fator-estorvo  $X$  foi bloqueado e o fator de interesse  $F$  foi randomizado, ou seja,  $X$  teve uma chance igual de influenciar todos os níveis do fator  $F$ . Na Tabela 4.5 é apresentado o resultado da casualização dos níveis de  $F$  entre blocos.

Para medir a capacidade de modelagem das TPN geradas pelos métodos é possível calcular a similaridade para cada uma das distribuições de probabilidade condicionais nas TPN com alguma medida padrão e, em seguida, usar a soma dos valores para descrever a similaridade das TPN. Ou seja, é possível medir a similaridade tratando as distribuições como vetores nos quais são aplicadas funções geométricas (e.g., distância euclidiana, distância de Manhattan) [28].

No entanto, para Zagorecki e Druzdzal [52] os valores de similaridade relacionados a diferentes distribuições de probabilidades não possuem igual importância, e dependem do quão provável seja a combinação correspondente de estados dos nós pais na vida real. Ou seja, é preciso realizar uma soma ponderada das medidas de similaridade.

Diante do exposto em Zagorecki e Druzdzal [52], para medir a capacidade de modelagem dos métodos semiautomáticos (i.e., o quão bem as TPN construídas com os métodos semiautomáticos representam o conhecimento dos especialistas de domínio) foi tomada a decisão de não medir a similaridade dos vetores de probabilidade, mas sim, de toda a TPN, assim como em Laitila [25].

Dessa forma, a similaridade de duas TPN  $\Theta$  e  $\tilde{\Theta}$  com  $N$  combinações de estados dos nós pais (i.e., distribuições de probabilidade ou vetores de probabilidade) é medida considerando cada uma das células das TPN como pontos. Sobre qual medida de similaridade usar, a discussão apresentada em Zagorecki e Druzdzal [52] indica que não há uma medida padrão estabelecida para medir a similaridade de duas TPN.

Neste experimento foram adotadas as seguintes medidas de similaridade: Distância Euclidiana (ED) (Eq. 4.1), *Brier Score* (BS) (Eq. 4.2) e Erro Médio Absoluto (EM) (Eq. 4.3). Estas medidas foram utilizadas para comparar os pontos (i.e., células) das TPN (i.e.,  $\Theta_{ij}$  e  $\tilde{\Theta}_{ij}$ ). A divergência de Kullback-leibler foi considerada mas sua limitação para lidar com casos em que a probabilidade estimada é zero e a probabilidade esperada não é zero [52] fez com ela não fosse adotada <sup>1</sup>. As equações das medidas adotadas são apresentadas em seguida:

$$ED = \sqrt{\sum_{i=1}^N \sum_{j=1}^S (\Theta_{ij} - \tilde{\Theta}_{ij})^2} \quad (4.1)$$

$$BS = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^S (\Theta_{ij} - \tilde{\Theta}_{ij})^2 \quad (4.2)$$

$$EM = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^S |\Theta_{ij} - \tilde{\Theta}_{ij}| \quad (4.3)$$

<sup>1</sup>Uma medida comumente adotada nesses casos é substituir zero por um número próximo de zero, mas na prática, foi observada uma punição muito elevada para incidências desses casos no experimento (i.e., apesar de terem ocorrido em apenas dois blocos)



em que:

- $\Theta$  é a TPN de referência;
- $\tilde{\Theta}$  é a TPN gerada por um método semiautomático;
- $N$  é o número de combinação de nós pais;
- $S$  é o número de estados do nó filho;
- $N$  é o número de probabilidades de  $\Theta$  e  $\tilde{\Theta}$ .

Geralmente, modelos baseados em redes Bayesianas servem como suporte à tomada de decisões, de forma que, muitas vezes, o que importa para o especialista de domínio é o estado mais provável para um determinada variável do modelo (i.e., o estado com maior probabilidade em um nó). Um exemplo disso são os modelos de estimativa de valor desenvolvidos no estudo de caso apresentado no Capítulo 3.

Dessa forma, a moda das distribuições de probabilidade se torna um critério relevante para comparar os métodos. Em outras palavras, os métodos de geração semiautomática de TPN conseguem gerar TPN que expressam corretamente a moda de distribuições de probabilidade relacionadas a combinações de estados dos nós pais assim como modelado nas TPN de referência? Quão bem eles conseguem realizar esta tarefa? Para responder a estas indagações é considerada a Proporção de Acertos (PA) dos métodos, conforme é apresentado na Equação 4.4:

$$PA = \frac{A}{N} \quad (4.4)$$

em que:

- $PA$  é a proporção de acertos;
- $A$  é o número de distribuições de probabilidade de  $\tilde{\Theta}$  em que a moda das distribuição de probabilidade é equivalente à obtida em  $\Theta$ ;
- $N$  é o número de distribuições de probabilidades de  $\Theta$  e  $\tilde{\Theta}$

Assim,  $PA$  torna-se uma métrica para medir a acurácia prática dos métodos<sup>2</sup>, relevante para este este experimento.

<sup>2</sup>Em alguns momentos no texto a Proporção de Acertos será referenciada como acurácia e vice-versa.

Desenvolvedores	Níveis de $F$			
	WSA+AHP	WSA	AHP	RNM
$X_1$	WSA+AHP	WSA	AHP	RNM
$X_2$	WSA+AHP	RNM	WSA	AHP
$X_3$	WSA	RNM	WSA+AHP	AHP
$X_4$	RNM	WSA+AHP	WSA	AHP
$X_5$	WSA+AHP	RNM	WSA	AHP
$X_6$	WSA+AHP	WSA	AHP	RNM
$X_7$	RNM	WSA	AHP	WSA+AHP
$X_8$	WSA+AHP	RNM	AHP	WSA
$X_9$	AHP	RNM	WSA+AHP	WSA
$X_{10}$	WSA	RNM	WSA+AHP	AHP

Tabela 4.5: Blocagem dos desenvolvedores com randomização dos níveis de  $F$ 

#### 4.4.7 Instrumentação

De acordo com Jedlitschka e Pfahl [22] há dois tipos de instrumentos em experimentos controlados em engenharia de software: manuais e instrumentos de medição (e.g., questionários, ferramentas de coleta de dados). Nesta seção, são apresentados os artefatos necessários à aplicação do experimento.

##### 4.4.7.1 Manual do Participante

Uma animação em vídeo, de caráter instrucional, foi desenvolvida para explicar as variáveis da rede Bayesiana - fornecendo uma descrição inicial dos nós (i.e., variáveis) e seus possíveis estados (i.e., valores). Um manual com informações acerca dos nós e seus possíveis estados foi confeccionado. Este manual foi disponibilizado aos especialistas de domínio durante a aplicação de cada tratamento.

##### 4.4.7.2 Planilhas Eletrônicas

Planilhas eletrônicas foram utilizadas para registrar informações referentes ao tempo de aplicação dos tratamentos, dados de entrada do método RNM e anotações diversas sobre os elementos do experimento.

### 4.4.7.3 Protótipo para calibragem do método RNM

Neste experimento, uma nova abordagem para elicitare conhecimento de especialistas e aplicar o método RNM é apresentada. Esta abordagem pode ser dividida em duas etapas: (i) elicitação das probabilidades, de forma direta ou indireta, dos especialistas de domínio; e (ii) uso do “Especialista Simulado” para extração dos parâmetros de entrada necessários para aplicar o método RNM (i.e., extraídos a partir dos dados obtidos em (i)). Em (i) os especialistas de domínio usam uma ferramenta gráfica que permite raciocinar sobre as distribuições de probabilidade em termos numéricos ou de proporção.

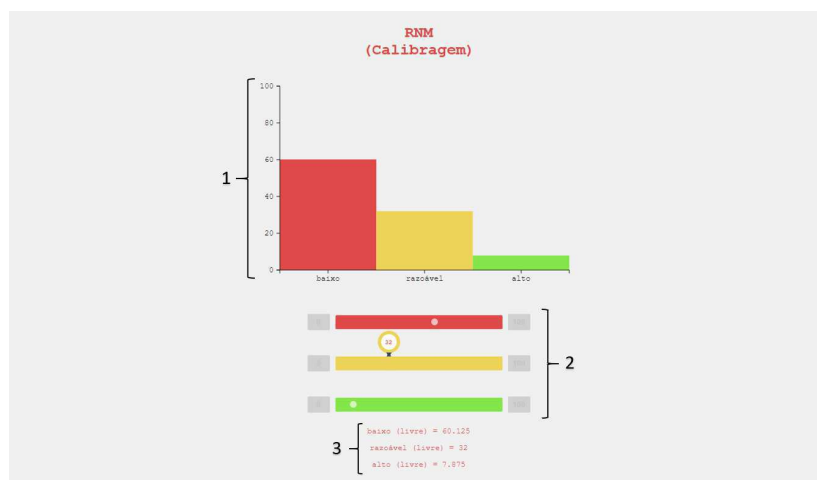


Figura 4.5: Protótipo para calibragem do método RNM

Na Figura 4.5 é apresentado o protótipo utilizado para elicitare o conhecimento dos especialistas de domínio. Para cada combinação de casos extremos do nós pais, os especialistas de domínio informam a distribuição de probabilidade esperada por meio de uma interface gráfica (i.e., diretamente no protótipo).

Para informar cada distribuição de probabilidade esperada, os especialistas de domínio podem usar o gráfico de barra (1), disposto verticalmente, ou o gráfico de barra (2), disposto horizontalmente. Os especialistas de domínio interagem diretamente com as barras do gráfico de forma que, ao aumentar ou diminuir uma das barras, as demais se adequam automaticamente. Por exemplo, na Figura 4.5, ao ajustar a barra referente ao estado razoável (em amarelo), os demais estados são auto-ajustados. Em outras palavras, ao aumentar a probabilidade de razoável em 10%, o gráfico diminui 5% dos demais estados.

Uma estratégia comumente empregada por especialistas de domínio consiste em, primei-

ramente, ajustar a barra referente ao estado que ele tem uma maior segurança em estimar e adequar os demais estados de acordo. Caso o especialista de domínio queira aplicar esta estratégia, o elemento (1) na Figura 4.5 permite travar o estado, de forma que as alterações nos demais estados não o modifiquem. Os especialistas podem fornecer os dados necessários ao “Especialista Simulado”, raciocinando em termos de proporção, caso em que ele ignora as informações numéricas do protótipo, ou de forma probabilística, utilizando os indicativos do elemento (2) e (3) na Figura 4.5.

#### 4.4.7.4 Plugin para aplicar o método WSA

Para aplicação do tratamento com WSA, foi desenvolvido um plugin na ferramenta BayesWeb<sup>3</sup>. A Figura 4.6 ilustra o uso do plugin por um dos participantes, no qual são identificadas as configurações de pais compatíveis e as suas respectivas probabilidades. O plugin tem como objetivo reduzir o esforço para aplicação do método WSA (i.e., esforço relativo ao processo de elicitación de conhecimento dos especialistas de domínio participantes).

Para redução do esforço, algumas estratégias foram empregadas ao plugin. Por exemplo, é comum aos especialistas de domínio identificar configurações de pais compatíveis equivalentes - o que diminui a quantidade de parâmetros para execução do WSA. Na Figura 4.6 a configuração de pais compatíveis do nó “Coordenação” = “razoável” é equivalente a configuração de pais compatíveis do nó “Orientação” = “razoável”. Neste caso, a interface do plugin fornece um feedback visual que indica relações de equivalência ao especialista de domínio, e automaticamente realiza uma transposição das probabilidades para evitar inconsistências.

#### 4.4.7.5 Plugin para aplicar o método AHP

Para aplicação dos tratamentos com o método AHP, foi desenvolvido um processo de elicitación de conhecimento que viabiliza o uso do método por especialistas de domínio, sem que estes, necessariamente, possuam um conhecimento prévio acerca do funcionamento interno do método (i.e., seus conceitos matemáticos).

O processo de elicitación de conhecimento para uso do método AHP baseou-se em um tipo de questionário, no formato de *wizard*<sup>4</sup>, em que o especialista de domínio responde a

<sup>3</sup>Uma ferramenta para construção de redes Bayesianas desenvolvida no Laboratório Embedded.

<sup>4</sup>Um tipo de interface de usuário que apresenta ao usuário uma sequência de caixas de diálogo que levam

← Weighted Sum Algorithm

Impossible Improbable Uncertain Fifty-Fifty Expected Probable Certain

Colaboração REVERSE ↺

Coordenação	Baixa	Razoável	Alta	Baixa	Razoável	Razoável
Orientação	Razoável	Razoável	Alta	Baixa	Razoável	Alta
Baixa	0.7	0.2	0	0.85	0.2	0
Razoável	0.3	0.6	0.2	0.15	0.6	0.35
Alta	0	0.2	0.8	0	0.2	0.65

SAVE WSA

Parent's weight Coordenação: 0.45    Parent's weight Orientação: 0.55

Figura 4.6: Plugin para aplicação do método WSA

uma série de perguntas em sequência. O *wizard* serve para guiar os especialistas por todos os passos até que o método seja completamente aplicado e uma TPN seja gerada.

Na Figura 4.7 são apresentadas duas telas do plugin para aplicação do método, implementado na ferramenta BayesWeb. Na Figura 4.7a é possível observar uma descrição de um cenário específico e uma pergunta. Nesta tela, o especialista de domínio participante deve selecionar o estado mais provável, dado o cenário apresentado, e informar quão mais provável este estado é em relação a outro estado. A comparação entre estados é feita através do uso de uma escala dupla com números e palavras. Nesta tela, ainda há um indicador do processo, que fornece feedback acerca do progresso da aplicação do método e botões de ação para avançar, retroceder ou reiniciar o processo.

Na Figura 4.7b o especialista de domínio verifica o resultado de suas respostas. Nesta tela, há duas informações relevantes: a distribuição de probabilidade resultante das comparações pareadas entre estados do nó filho e um índice de inconsistência, que serve como um indicativo de quão consistentes foram as respostas dadas pelo especialista de domínio.

#### 4.4.8 Procedimento de Coleta de Dados

As TPN dos modelos criados foram coletadas diretamente do banco de dados da ferramenta BayesWeb - utilizada na aplicação dos tratamentos. Para realizar esta tarefa foi utilizada a interface do usuário por uma série de etapas bem definidas. Tarefas complexas, realizadas com pouca frequência ou desconhecidas podem ser mais fáceis de executar usando um wizard.



(a) Perguntas do método AHP

(b) Resultado do método AHP

Figura 4.7: Plugin para aplicação do método AHP

ferramenta Robo 3T<sup>5</sup>. Os dados extraídos do banco de dados foram convertidos em arquivos no formato CSV<sup>6</sup>. A ferramenta R Studio<sup>7</sup> foi utilizada durante a análise dos dados.

Os início e término de cada tratamento foram registrados em planilhas eletrônicas. A interação dos especialistas com os métodos foi registrada utilizando o software de captura de vídeo, Camtasia<sup>8</sup>.

#### 4.4.9 Procedimento de Análise

Neste experimento é adotado um delineamento em blocos casualizados no qual há  $v$  tratamentos dispostos em  $N = vb$  unidades experimentais dispostos em  $b$  blocos de tamanho  $v$  no qual cada tratamento aparece uma vez em cada bloco.

Neste experimento, cada observação  $Y_{ij}$  pode ser decomposta conforme o modelo matemático descrito a seguir:

$$Y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}, \quad i = 1, 2, \dots, v; \quad j = 1, 2, \dots, b. \quad (4.5)$$

em que:

- $Y_{ij}$  é a observação (i.e., resposta) da  $i$ -ésima unidade experimental recebendo o tratamento no  $j$ -ésimo bloco;

<sup>5</sup><https://www.robomongo.org/>

<sup>6</sup>Arquivos de texto de formato regulamentado pelo RFC 4180, que faz uma ordenação de bytes ou um formato de terminador de linha, separando valores com vírgulas.

<sup>7</sup><https://www.rstudio.com/>

<sup>8</sup><http://www.techsmith.com>

- $\mu$  é uma constante (i.e., estimativa da média geral);
- $\tau_i$  é o efeito do  $i$ -ésimo tratamento;
- $\beta_j$  é o efeito do  $j$ -ésimo bloco;
- $\epsilon_{ij}$  é o erro aleatório na observação.

O interesse aqui é testar a igualdade do efeito dos tratamentos. Em outras palavras, o objetivo é testar a hipótese nula  $H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4$  contra a hipótese alternativa  $H_1$ : de que pelo menos dois dos  $\tau_i$ 's tratamentos são diferentes. É importante salientar que, ao longo do texto, sempre que houver uma menção a  $H_0$ , estarão sendo referenciadas na verdade, todas as hipóteses nulas das hipóteses elencadas na Tabela 4.1. De forma equivalente, sempre que houver uma menção a  $H_1$ , estarão sendo referenciadas todas as hipóteses alternativas deste experimento.

Neste experimento as observações  $Y_{ij}$  foram transformadas em  $\ln(Y_{ij})$  com o intuito de: (i) deixar as distribuições mais próximas da normalidade; (ii) trazer outliers (i.e., valores extremos) para mais perto do resto dos valores da população. Contudo, os valores apresentados na análise descritiva não estão transformados, visto que o uso dos dados transformados podem tornar a interpretação dos dados mais difícil.

## 4.5 Execução

Nesta seção é descrito o passo a passo para execução das atividades planejadas. Foram definidas quatro reuniões para cada um dos participantes. A primeira reunião foi reservada para a realização de um treinamento básico sobre redes Bayesianas, seguido de uma sessão de motivação, na qual o pesquisador explicou o objetivo do experimento e como este objetivo seria alcançado (i.e., que tipo de colaboração era necessário dos especialistas de domínio). Além disso, a motivação englobou a apresentação da rede Bayesiana que seria quantificada com os métodos. O primeiro tratamento foi realizado ainda na primeira reunião.

A animação em vídeo, descrita na seção 4.4.7.1, foi utilizada para apresentar os nós (i.e., variáveis) da rede Bayesiana e seus possíveis estados (i.e., valores). O objetivo do uso de uma animação em vídeo foi apresentar uma descrição completa do modelo, de forma didática

e padronizada. Os participantes foram instruídos a pausar a animação sempre que tivessem dúvidas. Para alguns dos participantes, houve um momento de discussão com o pesquisador, enquanto outros se limitaram a ver o vídeo por completo, pois se diziam preparados para participar do treinamento sobre o método que seria aplicado em seguida. Evidentemente que este comportamento variou de acordo com o nível de familiaridade dos participantes para com o domínio (i.e., experiência com métodos ágeis).

Além disso, o intuito da animação em vídeo - e discussão, com alguns participantes - foi prover uma base de conhecimento comum a todos os especialistas de domínio que participaram do experimento (i.e., entendimento comum acerca dos nós e seus possíveis estados). Por exemplo, para o nó “Colaboração”, o vídeo apresentou uma descrição do que é colaboração em um contexto de equipes ágeis de desenvolvimento de software, bem como o que significa uma colaboração baixa, razoável ou alta.

Contudo, é importante salientar que os especialistas de domínio foram instruídos a manter sua interpretação acerca das variáveis do domínio, sempre que houvesse divergência com o que foi exposto no vídeo instrucional. Em outras palavras, a importância do vídeo instrucional foi, além de promover uma base de conhecimento comum, invocar diversos cenários e estimular os especialistas de domínio. Aqueles com opiniões divergentes deveriam apenas se manter consistentes durante o experimento (i.e., durante a aplicação dos métodos e construção manual das TPN).

Ao término de cada reunião, a próxima era agendada de acordo com a disponibilidade do especialista de domínio. Este foi o procedimento adotado até que todos os métodos fossem aplicados. Foi estabelecido um intervalo mínimo de três dias e máximo de sete dias entre reuniões. Esta decisão teve como objetivo evitar possíveis vieses que poderia afetar o experimento caso o participante ficasse entediado ou fadigado, por exemplo.

Na última reunião, na qual o último método foi aplicado, foram dadas instruções ao especialista acerca do último passo do experimento. O último passo do experimento consiste na definição manual das TPN, que serve como parâmetro de referência para as TPN geradas pelos métodos. Foi acertado com cada um dos especialistas o seguinte: eles teriam que definir manualmente as TPN, mas este procedimento poderia ser realizado sem acompanhamento do pesquisador, de preferência em um momento que o próprio especialista de domínio julgasse adequado. Foi definido apenas um prazo limite. Esta decisão em particular, teve como



objetivo garantir que os especialistas de domínio definissem as TPN no momento em que estivessem mais à vontade para tal, dado a importância das TPN de referência.

Dessa forma, dias após a aplicação do último método, os especialistas receberam acesso a uma planilha eletrônica para definição manual das TPN. Esta planilha não foi enviada imediatamente após a aplicação do último método por dois motivos: o primeiro, foi evitar que o cansaço ou outro fator externo, subjetivo, afetasse a definição manual das TPN; o segundo, foi a necessidade de preparar as TPN para definição manual. Por exemplo, as combinações de estados já utilizadas como entrada na aplicação do tratamento com o método WSA (i.e., probabilidades definidas manualmente) não precisaram ser elicitadas novamente. Portanto, para evitar inconsistências e diminuir o esforço dos participantes, estas probabilidades foram transpostas para a TPN manual.

A duração do treinamento sobre redes Bayesianas, motivação e aplicação dos métodos foi registrada em planilhas eletrônicas, bem como observações acerca do comportamento do especialistas de domínio durante a aplicação dos métodos. Por outro lado, o treinamento sobre cada método, que antecedeu sua aplicação de fato, não foi registrada. Em relação à aplicação dos métodos, a Tabela 4.6 mostra o tempo transcorrido na aplicação de cada tratamento, e o total geral. É importante salientar, no entanto, que a Tabela 4.6 não reflete o tempo total gasto na reunião, apenas o tempo gasto na aplicação dos métodos de fato.

Na prática, as reuniões tiveram duração superior à exibida nas tabelas, devido a necessidade de apresentar em mais detalhes (em comparação com a introdução do experimento realizada na primeira reunião) os métodos aos especialistas. Na Tabela 4.6 é possível observar que o método WSA-AHP foi o método de maior duração, com um tempo médio de aplicação de 34.9 minutos. Os métodos WSA e RNM foram ambos aplicados em um tempo médio de 20.3 minutos. O método AHP foi aplicado em um tempo médio de 25.3 minutos.

Após a realização de todas estas atividades, os dados coletados foram compilados e preparados para realização da análise, que é apresentada na seção seguinte.

Participante	Métodos				Treinamento	Motivação	Total
	WSA	WSA-AHP	AHP	RNM			
X1	16	28	30	18	5	25	122
X2	25	81	28	18	7	9	168
X3	29	32	31	20	10	17	139
X4	27	41	27	30	16	12	153
X5	9	52	20	23	13	11	128
X6	41	25	35	22	6	10	139
X7	16	19	16	16	6	12	85
X8	14	24	22	27	0	10	97
X9	12	13	26	13	0	18	82
X10	14	34	18	16	9	19	110
Total	203	349	253	203	72	143	1223

Tabela 4.6: Registro de tempo da execução do experimento (em minutos)

## 4.6 Análise dos Dados

Nesta seção, são apresentados o resumo dos dados coletados no experimento e os procedimentos realizados na análise de variância.

### 4.6.1 Análise Descritiva

#### 4.6.1.1 Brier Score

Uma breve análise sob os dados apresentados nas Tabelas 4.7 e 4.8 sugere que, de acordo com a medida de similaridade, *Brier Score*, o método RNM é superior aos demais, seguido pelos métodos WSA, WSA-AHP e AHP. Este último, apresentando um desempenho inferior em relação aos demais métodos.

O boxplot apresentado na Figura 4.8c reforça a ideia de que o método de menor capacidade de modelagem de TPN seja o AHP. No entanto, ainda não há evidências suficientes para se chegar a esta conclusão. Somente o teste estatístico poderá confirmar se essas diferenças possuem significância estatística e se o método AHP realmente possui a menor capacidade de modelagem de TPN (i.e., considerando esta medida de similaridade).

Estatística	N	Média	Desvio Padrão	Min	Max
WSA	10	0.03209349	0.0160529	0.01785802	0.06077432
WSA-AHP	10	0.03368596	0.01563935	0.0162316	0.05885354
AHP	10	0.06875261	0.02416669	0.04533854	0.1052835
RNM	10	0.02855719	0.02279669	0.01118815	0.0898061

Tabela 4.7: Sumário dos dados (BS). Legenda: azul > verde > preto > vermelho

#### 4.6.1.2 Distância Euclidiana

Uma breve análise sob os dados apresentados nas Tabelas 4.9 e 4.10 sugere que, de acordo com a distância euclidiana entre as TPN geradas pelos métodos e as TPN de referência, o

	WSA	WSA-AHP	AHP	RNM
1	0.06077432	0.05541539	0.1042138	0.03268251
2	0.02719151	0.03145276	0.1052835	0.02450732
3	0.01786418	0.02025455	0.06162503	0.0171046
4	0.03234614	0.03385527	0.04533854	0.02061901
5	0.01785802	0.02431274	0.05193	0.02649935
6	0.02366031	0.02327819	0.04641398	0.03325619
7	0.03404023	0.02259454	0.08983505	0.01518141
8	0.02706509	0.05885354	0.05448695	0.01472725
9	0.06056257	0.05061101	0.08143015	0.0898061
10	0.01957253	0.0162316	0.04696904	0.01118815

Tabela 4.8: Dados brutos coletados (BS). Legenda: azul > verde > preto > vermelho

método RNM apresenta uma capacidade de modelagem de TPN superior aos demais, seguido pelos métodos WSA, WSA-AHP e AHP. Este último, apresentando, mais uma vez, o pior desempenho entre os métodos.

O boxplot apresentado na Figura 4.8b reforça a ideia de que o método de menor capacidade de modelagem de TPN seja o AHP. No entanto, ainda não há evidências suficientes para se chegar a esta conclusão. Somente o teste estatístico poderá confirmar se essas diferenças possuem significância estatística e se o método AHP realmente possui a menor capacidade de modelagem de TPN.

#### 4.6.1.3 Erro Médio Absoluto

Uma breve análise sob os dados apresentados nas Tabelas 4.11 e 4.12 sugere que, de acordo com a erro médio absoluto obtido ao comparar as TPN geradas pelos métodos e as TPN de referência, o método RNM, mais uma vez, apresenta uma capacidade de modelagem de TPN superior aos demais, seguido pelos métodos WSA, WSA-AHP e AHP. Este último, apresentando, pela terceira vez, o pior desempenho entre os métodos.

Estatística	N	Média	Desvio Padrão	Min	Max
WSA	10	1.572655	0.3746541	1.202705	2.21872
WSA-AHP	10	1.613567	0.3726241	1.14663	2.183377
AHP	10	2.328319	0.4053683	1.916356	2.920268
RNM	10	1.44619	0.4962837	0.9519666	2.69709

Tabela 4.9: Sumário dos dados (ED). Legenda: azul &gt; verde &gt; preto &gt; vermelho

	WSA	WSA-AHP	AHP	RNM
1	2.21872	2.118643	2.905395	1.627047
2	1.484086	1.596143	2.920268	1.408933
3	1.202913	1.280866	2.234195	1.177061
4	1.618653	1.655982	1.916356	1.292339
5	1.202705	1.403329	2.050934	1.465076
6	1.384372	1.373147	1.938951	1.641265
7	1.6605	1.352833	2.697525	1.108916
8	1.480632	2.183377	2.10082	1.092203
9	2.214852	2.02472	2.568237	2.69709
10	1.259117	1.14663	1.950511	0.9519666

Tabela 4.10: Dados brutos coletados (ED). Legenda: azul &gt; verde &gt; preto &gt; vermelho

O boxplot apresentado na Figura 4.8a reforça a ideia de que o método de menor capacidade de modelagem de TPN seja realmente o AHP. No entanto, ainda não há evidências suficientes para se chegar a esta conclusão. Somente o teste estatístico poderá confirmar se essas diferenças possuem significância estatística e se o método AHP realmente possui a menor capacidade de modelagem de TPN.

Estatística	N	Média	Desvio Padrão	Min	Max
WSA	10	0.1219101	0.02966282	0.08987654	0.177679
WSA-AHP	10	0.1406646	0.03617692	0.09871852	0.1996077
AHP	10	0.1981249	0.03720482	0.1588691	0.2711122
RNM	10	0.1203499	0.03543112	0.08439753	0.2082568

Tabela 4.11: Sumário dos dados (EM). Legenda: azul > verde > preto > vermelho

	WSA	WSA.AHP	AHP	RNM
1	0.177679	0.1918825	0.2373654	0.1359062
2	0.1093933	0.1267867	0.2711122	0.1022783
3	0.09113595	0.1086308	0.1909127	0.1049482
4	0.127284	0.14224	0.1588691	0.1144667
5	0.08987654	0.1237383	0.1742296	0.122684
6	0.1063503	0.1168399	0.1607528	0.1374577
7	0.1277272	0.1199099	0.2326	0.09840617
8	0.118062	0.1996077	0.1853556	0.09469788
9	0.1672099	0.1782914	0.1979753	0.2082568
10	0.1043827	0.09871852	0.1720765	0.08439753

Tabela 4.12: Dados brutos coletados (EM). Legenda: azul > verde > preto > vermelho

#### 4.6.1.4 Proporção de Acertos

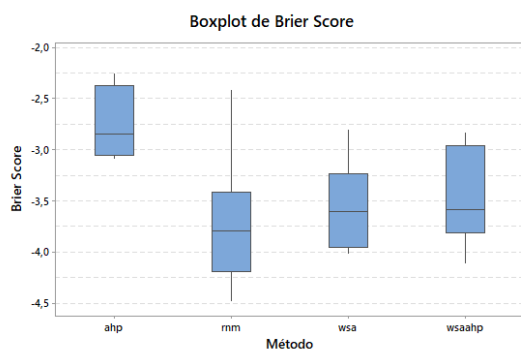
Uma breve análise sob os dados apresentados nas Tabelas 4.13 e 4.14 sugere que, de acordo com a proporção de acertos, o método RNM é superior aos demais em relação a acurácia, seguido pelos métodos AHP, WSA-AHP e WSA. O que é interessante neste resultado é o resultado do método AHP, que até então vinha apresentando o pior desempenho acerca da capacidade de modelagem, e agora apresenta o segundo melhor índice de acurácia. No entanto, o boxplot apresentado na Figura 4.8d não apresenta evidências suficientes para chegar a esta conclusão. Somente o teste estatístico poderá confirmar se essas diferenças possuem significância estatística.

Estatística	N	Média	Desvio Padrão	Min	Max
WSA	10	0,6312	0,1260719475	0,4444	0,7692
WSA-AHP	10	0,5977	0,1277322499	0,3478	0,7600
AHP	10	0,6583	0,08807105179	0,4615	0,7308
RNM	10	0,7579	0,1206166418	0,5769	0,9231

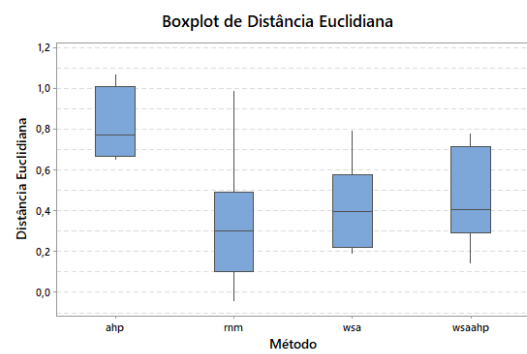
Tabela 4.13: Sumário dos dados (PA). Legenda: azul > verde > preto > vermelho

	WSA	WSA-AHP	AHP	RNM
1	0,4444	0,4074	0,5926	0,5926
2	0,5769	0,5769	0,5769	0,8462
3	0,6538	0,5769	0,4615	0,5769
4	0,7692	0,6538	0,7308	0,9231
5	0,7600	0,6800	0,6800	0,7200
6	0,6800	0,7600	0,6800	0,6800
7	0,4444	0,6667	0,7037	0,8519
8	0,5217	0,3478	0,6957	0,6957
9	0,7308	0,6538	0,7308	0,8462
10	0,7308	0,6538	0,7308	0,8462

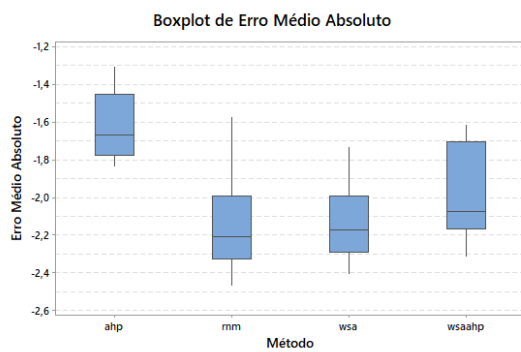
Tabela 4.14: Dados brutos coletados (PA). Legenda: azul > verde > preto > vermelho



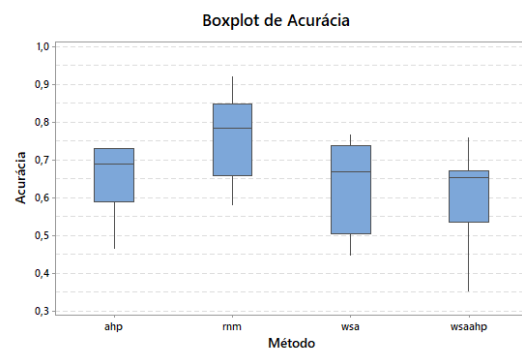
(a) Brier Score



(b) Distância Euclidiana



(c) Erro Médio Absoluto



(d) Proporção de Acertos

Figura 4.8: Boxplot das medidas de similaridade e proporção



### 4.6.2 Validação do Modelo

Para realizar a análise de variância é preciso que as variáveis envolvidas no modelo matemático adotado obedeçam alguns pré-requisitos:

1. Independência dos erros;
2. Normalidade;
3. Homocedasticidade.

Nesta seção serão mostrados os resultados da análise gráfica e testes realizados nas variáveis do modelo logarítmico natural <sup>9</sup>.

A normalidade refere-se ao formato da distribuição dos resíduos do modelo estatístico e sua correspondência à distribuição normal. Para descobrir se os resíduos correspondem a uma distribuição normal, é possível realizar análises gráficas ou testes de hipóteses. Neste experimento, a análise gráfica foi realizada utilizando histogramas dos resíduos e gráficos quantil-quantil normal relativos a cada medida de similaridade.

A Figura 4.9 apresenta os gráficos do histograma de cada uma das medidas de similaridade coletadas no experimento, incluindo a métrica de proporção de acertos. No gráfico é apresentado a curva esperada para uma distribuição normal (em azul) e a curva obtida pela distribuição dos resíduos (em preto), considerando cada uma das medidas de similaridade. A partir dos gráficos, é possível observar uma similaridade entre as curvas, o que sugere que os resíduos assumem uma distribuição normal.

Uma outra análise gráfica que pode ser utilizada para refutar ou reforçar os resultados obtidos é observar o gráfico quantil-quantil normal. Este gráfico possibilita a comparação de duas distribuições. Neste caso, uma das distribuições seria a normal e a outra a da variável em questão. Para afirmar que a variável assume uma distribuição normal é preciso que os pontos no gráfico referentes a distribuição da variável se aproximem da reta diagonal traçada no gráfico, esta que representa a distribuição normal.

---

<sup>9</sup>Embora a transformação logarítmica tenha sido aplicada nas medidas de similaridade, ao se referir a estas medidas nos gráficos e ao longo do texto em nenhum momento será apresentado explicitamente o uso da transformada logarítmica (i.e. em vez de mencionar medida de similaridade  $\ln(\text{BS})$  será dito apenas medida de similaridade BS)

A Figura 4.10 apresenta os gráficos quantil-quantil normal dos resíduos obtidos na análise de variância de cada uma das medidas de similaridade. A partir dos gráficos, os resíduos da análise de variância relacionadas a cada uma das medidas de similaridade assumem uma distribuição normal.

Para apoiar ou refutar os resultados obtidos na análise gráfica, que sugere a normalidade da distribuição dos resíduos, foi realizado o teste de Shapiro-Wilk. Neste teste é avaliada a hipótese nula de que os resíduos da análise de variância tem uma distribuição normal, contra a hipótese alternativa de que ela não tem uma distribuição normal. Se o p-valor está abaixo do limiar determinado (neste experimento  $\alpha$  igual a 0.05), então a hipótese nula é rejeitada e a hipótese alternativa prevalece.

A Tabela 4.15 mostra o teste de Shapiro-Wilk aplicado aos resíduos obtidos a partir da análise de variância de cada medida de similaridade. Os resultados do teste apoiam os resultados da análise gráfica, e pode-se dizer que os resíduos assumem uma distribuição normal (i.e., a hipótese nula prevalece), considerando as quatro medidas de similaridade.

O termo homocedasticidade designa a variância constante dos erros experimentais  $\epsilon_{ij}$  para observações distintas ou a suposição de que a variável dependente exibe um mesmo nível de variância ao longo dos valores assumidos pelas variáveis independentes. Uma forma de análise gráfica que pode ser feita é observar o gráfico dos resíduos pelos valores ajustados da variável resposta.

O gráfico dos resíduos versus valores ajustados é uma das principais técnicas utilizadas para verificar as suposições dos resíduos. Além da detecção de heteroscedasticidade (i.e., ausência de homoscedasticidade), esse gráfico pode indicar que não existe uma relação linear entre as variáveis explicativas com a variável resposta por meio de alguma tendência nos pontos.

Para o diagnóstico de heteroscedasticidade, tentamos encontrar alguma tendência no gráfico. Por isso, se os pontos estão aleatoriamente distribuídos em torno da reta horizontal que passa na origem do eixo das ordenadas, sem nenhum comportamento ou tendência, temos indícios de que a variância dos resíduos é homoscedástica. Já a presença de “funil” ou “parábola” é um indicativo da presença de heteroscedasticidade.

Este mesmo gráfico e comportamento também é esperado para o pré-requisito da independência entre variáveis. Em outros termos, o gráfico dos resíduos pelos valores ajustados

serve para analisar os pré-requisitos 1 e 3 ao mesmo tempo.

A Figura 4.11 apresenta o gráfico dos resíduos pelos valores ajustados. A partir dele é possível observar que a dispersão dos resíduos não apresenta um padrão e portanto, pode-se dizer que a variável dependente apresenta o mesmo nível de variância ao longo dos valores assumidos pelas variáveis independentes e não há interação entre estas variáveis independentes.

Desta maneira, o modelo matemático é validado através de uma análise residual. As suposições de normalidade; independência dos erros e variação constante dos resíduos foram verificadas para cada uma das variáveis resposta da pesquisa. Em seguida este modelo é utilizado para testar estatisticamente as hipóteses do experimento através do método ANOVA (análise de variância).

Medida de Similaridade	W	p-valor
<i>Brier Score</i>	0.95815	0.1447
Distância Euclidiana	0.95815	0.1447
Erro Médio Absoluto	0.96468	0.2412
Proporção de Acertos	0.97931	0.6641

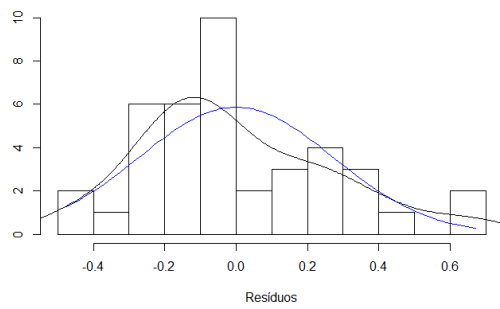
Tabela 4.15: Teste de normalidade dos resíduos (Shapiro-Wilk)

### 4.6.3 Análise da Variância

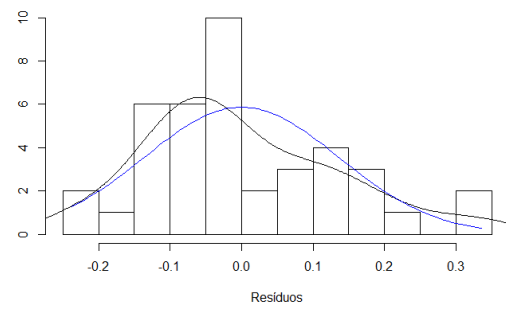
As hipóteses da investigação científica foram testadas através da técnica de análise de variância (Analysis of Variance - ANOVA).

#### 4.6.3.1 Alocação de variação

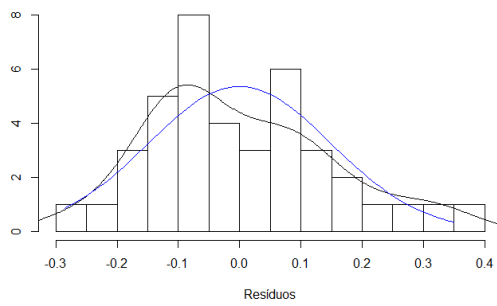
Conforme pode ser visto na Tabela 4.16, considerando a medida de similaridade BS e ED, 43% da variação da saída é explicada pelo fator método semiautomático de geração de TPN (i.e.,  $\tau$ ), os blocos explicam 36% da variação (i.e.,  $\beta$ , a diferença entre os especialistas de domínio que participaram do experimento) e o erro experimental (i.e.,  $\epsilon$ ) que sobra explica 21%.



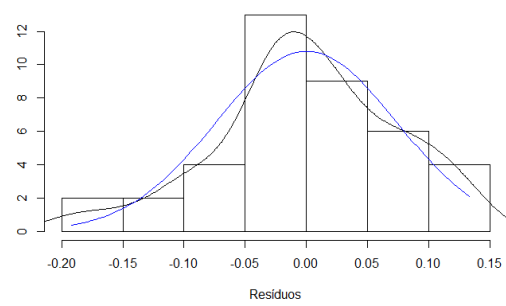
(a) Brier Score



(b) Distância Euclidiana

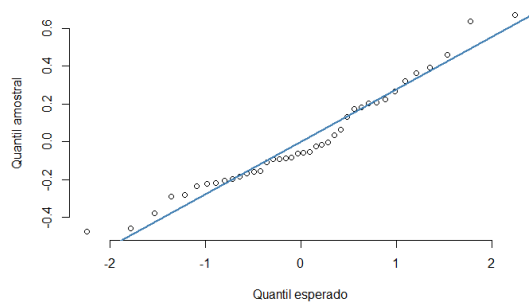


(c) Erro Médio Absoluto

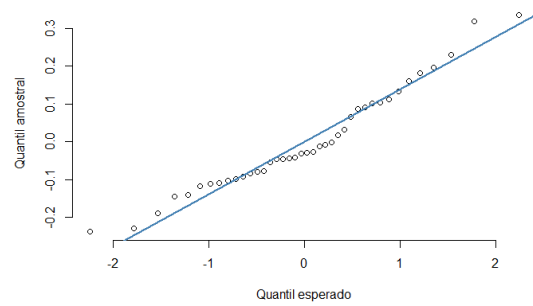


(d) Acurácia

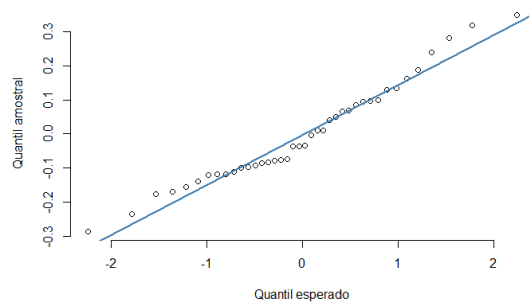
Figura 4.9: Histograma dos resíduos da análise de variância (TPN Adaptabilidade)



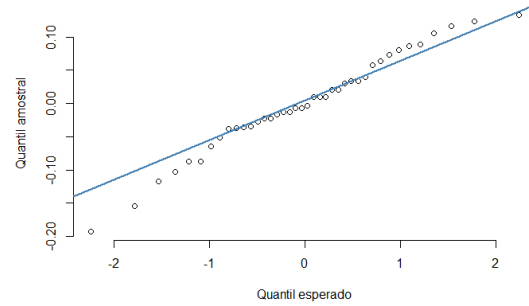
(a) Brier Score



(b) Distância Euclidiana

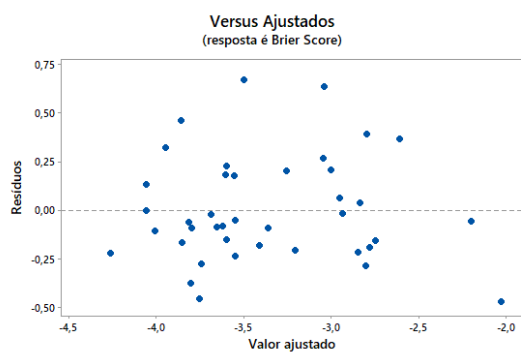


(c) Erro Médio Absoluto

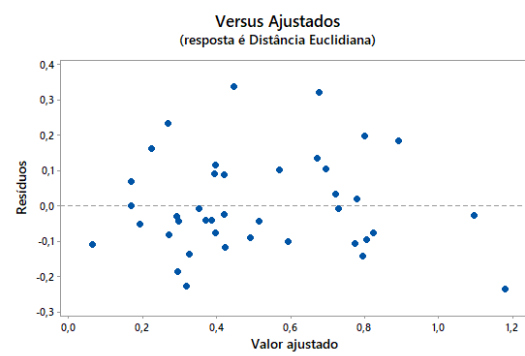


(d) Acurácia

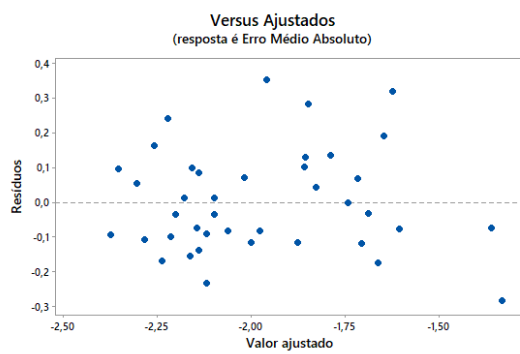
Figura 4.10: Quantil-quantil normal dos resíduos da análise de variância (TPN Adaptabilidade)



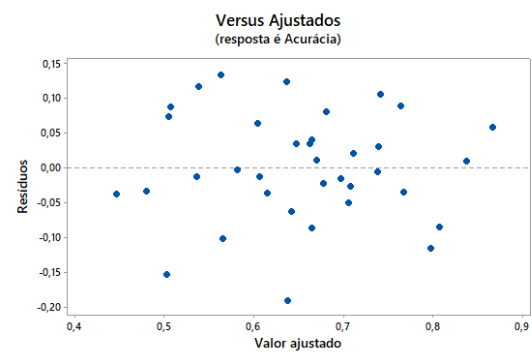
(a) Brier Score



(b) Distância Euclidiana



(c) Erro Médio Absoluto



(d) Acurácia

Figura 4.11: Gráfico dos resíduos pelos valores preditivos (TPN Adaptabilidade)

Em relação à medida de similaridade EM, 47% da variação da saída é explicada pelo fator método semiautomático de geração de TPN, os blocos explicam 29% da variação e o erro experimental explica 24%. Em relação a medida PS, 23% da variação da saída é explicada pelo fator método semiautomático de geração de TPN, os blocos explicam 44% da variação e o erro experimental explica 34%.

Em resumo, os fatores explicam 79% da variação da saída usando a medida BS e ED, 76% usando a medida EM e 66% usando a proporção PS (os valores dessa tabela estão arredondados). Neste design espera-se que o fator de bloqueio explique bastante variação, portanto se a média quadrática (MQ) do fator de bloqueio for maior que a média quadrática do erro conclui-se que a bloqueio foi eficiente. Este foi o caso para todas as medidas de similaridade e para a métrica de acurácia.

Efeito	Medidas de Similaridade / Acurácia			
	Brier Score (BS)	Distância Euclidiana (ED)	Erro Médio Absoluto (EM)	Acurácia (PA)
$\tau$	42,97%	42,97%	47,31%	43,91%
$\beta$	35,99%	35,99%	28,73%	43,91%
$\epsilon$	21,04%	21,04%	23,96%	33,55%

Tabela 4.16: Alocação de variação do experimento

#### 4.6.3.2 Significância dos efeitos

Para o fator de interesse (i.e., o efeito dos métodos) foi testada as hipóteses apresentadas na Tabela 4.1. Realizando um teste F com nível de significância 0.05, a hipótese nula  $H_0$  é rejeitada para o efeito das alternativas dos métodos semiautomáticos de geração de TPN (com todas as medidas de similaridade e proporção). Isto é, é possível afirmar que há uma diferença significativa entre pelo menos dois métodos.

Exemplo (teste F): conforme podemos observar na Tabela 4.17, para verificar se existe diferença entre os efeitos das alternativas (i.e., linha destacada em negrito) dos métodos semiautomáticos de geração de TPN, utilizando como medida de similaridade o *Brier Score*, temos:  $F > F_{crítico}$ . Assim, rejeitamos  $H_{10}$ . Logo, existe diferença entre os efeitos das alternativas dos métodos semiautomáticos de geração de TPN.

Os dados do teste F utilizando a medida de similaridade Distância Euclidiana (i.e.,  $H_2$ )

são apresentados na Tabela 4.18. Os dados do teste F utilizando a medida de similaridade Erro Médio Absoluto (i.e.,  $H3$ ) são apresentados na Tabela 4.19. Os dados do teste F utilizando a Proporção de Acertos (i.e.,  $H4$ ) são apresentados na Tabela 4.19.

### Brier Score

<i>Fonte da variação</i>	<i>SQ</i>	<i>gl</i>	<i>MQ</i>	<i>F</i>	<i>valor-P</i>	<i>F crítico</i>
Desenvolvedores	4,924213	9	0,547135	5,132507	0,000435	2,250131
<b>Métodos</b>	<b>5,878199</b>	<b>3</b>	<b>1,9594</b>	<b>18,38054</b>	<b>1,06E-06</b>	<b>2,960351</b>
Erro	2,87825	27	0,106602			

Tabela 4.17: Teste F com nível de significância 0.05 (*Brier Score*)

### Distância Euclidiana

<i>Fonte da variação</i>	<i>SQ</i>	<i>gl</i>	<i>MQ</i>	<i>F</i>	<i>valor-P</i>	<i>F crítico</i>
Desenvolvedores	1,231053	9	0,136784	5,132507	0,000435	2,250131
<b>Métodos</b>	<b>1,46955</b>	<b>3</b>	<b>0,48985</b>	<b>18,38054</b>	<b>1,06E-06</b>	<b>2,960351</b>
Erro	0,719563	27	0,02665			

Tabela 4.18: Teste F com nível de significância 0.05 (Distância Euclidiana)

### Erro Médio Absoluto

<i>Fonte da variação</i>	<i>SQ</i>	<i>gl</i>	<i>MQ</i>	<i>F</i>	<i>valor-P</i>	<i>F crítico</i>
Desenvolvedores	1,037852	9	0,115317	3,597507	0,004674	2,250131
<b>Métodos</b>	<b>1,709014</b>	<b>3</b>	<b>0,569671</b>	<b>17,77186</b>	<b>1,43E-06</b>	<b>2,960351</b>
Erro	0,865476	27	0,032055			

Tabela 4.19: Teste F com nível de significância 0.05 (Erro Médio Absoluto)



<b>Acurácia</b>						
<i>Fonte da variação</i>	<i>SQ</i>	<i>gl</i>	<i>MQ</i>	<i>F</i>	<i>valor-P</i>	<i>F crítico</i>
Desenvolvedores	0,278166	9	0,030907	3,925331	0,002732	2,250131
<b>Métodos</b>	<b>0,142865</b>	<b>3</b>	<b>0,047622</b>	<b>6,048106</b>	<b>0,002745</b>	<b>2,960351</b>
Erro	0,212593	27	0,007874			

Tabela 4.20: Teste F com nível de significância 0.05 (Acurácia)

#### 4.6.3.3 Verificação das Hipóteses

Sabendo que todos os fatores possuem significância estatística sobre as variáveis resposta, é simples verificar e tomar uma decisão sobre as hipóteses apresentadas na Tabela 4.1 que motivaram o experimento:

O teste F citado na Seção 4.6.3.2 indicou que existe diferença nas medidas de similaridade e acurácia em decorrência das alternativas dos métodos semiautomáticos de geração de TPN, e não devido ao acaso. Assim, rejeita-se  $H_0$  e conclui-se que a capacidade de modelagem e a acurácia dos métodos é diferente.

#### 4.6.3.4 Testes post-hoc

Após a realização da análise de variância (ANOVA), testes post hoc podem ser realizados para identificar quais dos pares de métodos diferem com relação à capacidade de modelagem e acurácia. Como o número de observações são iguais entre os métodos, é possível aplicar o teste Tukey para avaliar a diferença entre as médias das medidas de similaridade e as médias de proporção de acertos dos métodos.

Na Tabela 4.21 são apresentados os resultados do teste Tukey para as diferenças de médias da medida de similaridade BS. Na Tabela 4.21 (olhando para a coluna “Diferença de Médias” e “Valor-P Ajustado”) podemos ver quais os métodos que possuem diferenças significantes. Por exemplo, podemos concluir que:

- Não há diferenças significantes entre os métodos WSA-RNM (p-valor 0,502), WSAHP-RNM (p-valor 0,310) e WSAHP-WSA (p-valor 0,984);

- Há diferenças significativas entre os métodos RNM-AHP, WSA-AHP e WSAAHP-AHP. Todos com p-valor ajustado para múltiplas comparações menor que o nível de significância (0.05).

Olhando para a coluna “Diferença de Médias” podemos ver a diferença das médias e olhando para a coluna “IC de 95% simultâneo” podemos ver que os métodos que apresentam diferenças significativas na média são aqueles em que os intervalos de confiança não incluem o zero. Na Figura 4.12 podemos visualizar pares de métodos e analisar diferenças significativas de forma mais rápida. Diferenças significativas são aquelas que não cruzam o valor zero (tracejado verde).

De modo geral, os resultados obtidos com as medidas de similaridade são equivalentes. Na Figura 4.12 podemos observar que, considerando as medidas de similaridade BS (Figura 4.12a), ED (Figura 4.12b) e EM (Figura 4.12c) o método AHP apresenta uma capacidade de modelagem inferior a todos os outros métodos. Os resultados detalhados para as medidas de similaridade ED e EM são apresentados nas Tabelas 4.22 e 4.23, respectivamente.

No que se refere à acurácia dos métodos, no entanto, os resultados são diferentes. Observando a Figura 4.24 e a Tabela 4.24 podemos concluir que:

- Não há diferenças significantes entre os métodos RNM-AHP (p-valor 0,081), WSA-AHP (p-valor 0,903), WSAAHP-AHP (p-valor 0,436) e WSAAHP-WSA (p-valor 0,833);
- Há diferenças significativas entre os métodos WSA-RNM e WSAAHP-RNM. Em ambos os casos o método RNM se sobressai aos demais.

**Testes Simultâneos de Tukey para as Diferenças de Médias (Brier Score)**

Diferença de Níveis	Diferença de Médias	EP da Diferença	IC de 95% simultâneo	Valor-T	Valor-P Ajustado
rnm - ahp	-1,012	0,146	(-1,411; -0,612)	-6,93	0,000
wsa - ahp	-0,805	0,146	(-1,205; -0,406)	-5,51	0,000
wsaahp - ahp	-0,753	0,146	(-1,153; -0,354)	-5,16	0,000
wsa - rnm	0,206	0,146	(-0,193; 0,606)	1,41	0,502
wsaahp - rnm	0,258	0,146	(-0,141; 0,658)	1,77	0,310
wsaahp - wsa	0,052	0,146	(-0,348; 0,451)	0,35	0,984

*Nível de confiança individual = 98,92%*

Tabela 4.21: Testes simultâneos de Tukey para as diferenças de médias (BS)

**Testes Simultâneos de Tukey para as Diferenças de Médias (Distância Euclidiana)**

Diferença de Níveis	Diferença de Médias	EP da Diferença	IC de 95% simultâneo	Valor-T	Valor-P Ajustado
rnm - ahp	-0,5058	0,073	(-0,7056; -0,3060)	-6,93	0,000
wsa - ahp	-0,4026	0,073	(-0,6024; -0,2028)	-5,51	0,000
wsaahp - ahp	-0,3767	0,073	(-0,5765; -0,1770)	-5,16	0,000
wsa - rnm	0,1032	0,073	(-0,0966; 0,3030)	1,41	0,502
wsaahp - rnm	0,1291	0,073	(-0,0707; 0,3288)	1,77	0,310
wsaahp - wsa	0,0259	0,073	(-0,1739; 0,2256)	0,35	0,984

*Nível de confiança individual = 98,92%*

Tabela 4.22: Testes simultâneos de Tukey para as diferenças de médias (ED)

**Testes Simultâneos de Tukey para as Diferenças de Médias (Erro Médio Absoluto)**

Diferença de Níveis	Diferença de Médias	EP da Diferença	IC de 95% simultâneo	Valor-T	Valor-P Ajustado
rnm - ahp	-0,516	0,0801	(-0,7351; -0,2968)	-6,44	0
wsa - ahp	-0,4953	0,0801	(-0,7144; -0,2762)	-6,19	0
wsaahp - ahp	-0,3557	0,0801	(-0,5748; -0,1366)	-4,44	0,001
wsa - rnm	0,0207	0,0801	(-0,1984; 0,2398)	0,26	0,994
wsaahp - rnm	0,1602	0,0801	(-0,0589; 0,3793)	2	0,213
wsaahp - wsa	0,1395	0,0801	(-0,0796; 0,3586)	1,74	0,322

*Nível de confiança individual = 98,92%*

Tabela 4.23: Testes simultâneos de Tukey para as diferenças de médias (EM)

**Testes Simultâneos de Tukey para as Diferenças de Médias (Acurácia)**

Diferença de Níveis	Diferença de Médias	EP da Diferença	IC de 95% simultâneo	Valor-T	Valor-P Ajustado
rnm - ahp	0,0996	0,0397	(-0,0090; 0,2082)	2,51	0,081
wsa - ahp	-0,0271	0,0397	(-0,1357; 0,0815)	-0,68	0,903
wsaahp - ahp	-0,0606	0,0397	(-0,1692; 0,0480)	-1,53	0,436
wsa - rnm	-0,1267	0,0397	(-0,2353; -0,0181)	-3,19	0,018
wsaahp - rnm	-0,1602	0,0397	(-0,2688; -0,0516)	-4,04	0,002
wsaahp - wsa	-0,0335	0,0397	(-0,1421; 0,0751)	-0,84	0,833

*Nível de confiança individual = 98,92%*

Tabela 4.24: Testes simultâneos de Tukey para as diferenças de médias (PA)

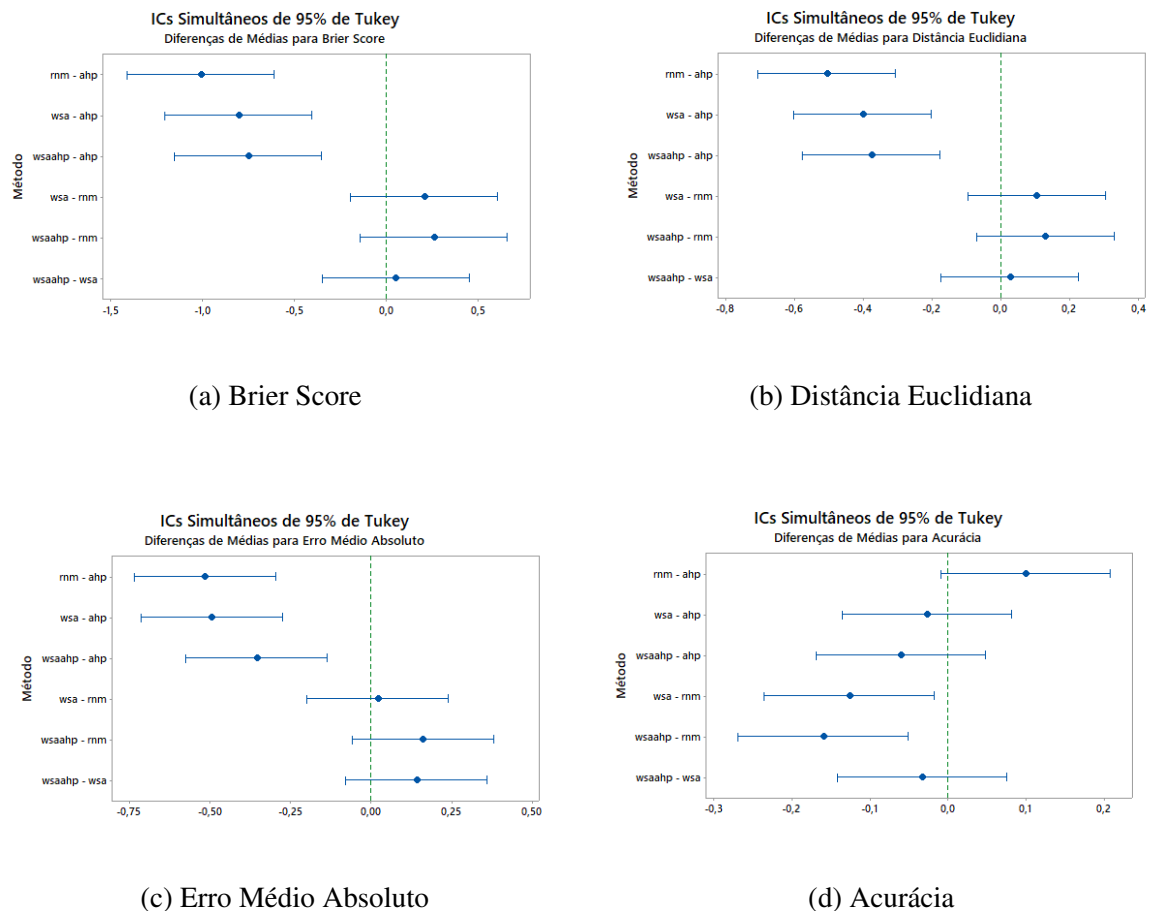


Figura 4.12: Teste de Tukey (diferença de médias entre métodos)

## 4.7 Resultados e Discussão

O objetivo desta seção é interpretar as descobertas da análise apresentada na seção anterior. Isto inclui uma visão geral dos resultados e uma descrição das ameaças à validade.

### 4.7.1 Avaliação dos Resultados e Implicações

No que diz respeito a capacidade de modelagem de TPN, os resultados obtidos mostram a superioridade dos métodos WSA, WSA-AHP e RNM em relação ao método AHP. As duas vantagens do método AHP são: a elicitación indireta de probabilidades e a redução do esforço cognitivo dos especialistas (i.e., as comparações pareadas dos estados são condicionadas a um único pai de cada vez).

Não obstante, a derivação da TPN completa a partir do produto das probabilidades de

nós filhos (i.e., estratégia aplicada pelo método AHP) apresentou um potencial inferior aos outros métodos. Por outro lado, o uso de comparações pareadas entre estados para extração de probabilidades combinado com o método WSA se mostrou promissor, haja vista que não houve diferença significativa entre WSA e WSA-AHP em nenhum dos teste  $F$  aplicados. Este achado é interessante, pois valida uma extensão ao método WSA para elicitación de probabilidades de forma indireta, útil para situações em que os especialistas de domínio não se sentem confortáveis em estimar probabilidades de forma direta.

De um ponto de vista prático, ao observar as Tabelas 4.7, 4.9 e 4.11, é possível notar uma superioridade clara do método RNM em relação aos demais métodos. O método WSA apresenta uma leve vantagem sobre o método WSA-AHP e fica claro a inferioridade do método AHP. No que se refere a acurácia dos métodos, é possível observar a superioridade do método RNM, que se sobressai aos métodos WSA e WSA-AHP (com diferença estatística significativa). Contudo, curiosamente para este critério o método AHP se saiu melhor do que os métodos WSA e WSA-AHP, pois não há diferença significativa entre o método AHP e RNM. Este resultado chama atenção, haja vista que o método AHP foi o pior em relação a capacidade de modelagem de TPN.

Conclui-se que o método que apresenta o melhor resultado no geral é o RNM, seguido dos métodos WSA e WSA-AHP. No entanto, diante dos resultados obtidos em relação à acurácia, o método AHP não pode ser descartado, apesar de sua capacidade limitada de modelar o conhecimento de especialistas de domínio. De um modo geral, os resultados sugerem que apesar do método AHP gerar distribuições de probabilidades mais dispersas daquelas esperadas pelos especialistas de domínio, ele ainda é capaz de modelar a tendência central da moda das distribuições.

#### 4.7.2 Ameaças à Validade

Uma das ameaças à validade presentes neste experimento está intrinsecamente relacionada a natureza do método WSA e ao meio de medição da acurácia, uma vez que as entradas para o método WSA foram utilizadas na TPN manual com o propósito de manter a consistência. Por conseguinte, os vetores de probabilidade utilizados na aplicação do WSA automaticamente registraram acertos de 100%. Para minimizar este possível viés, as TPN menores não foram consideradas na análise de acurácia devido à acentuação deste efeito. Isto posto, considera-se

que a TPN utilizada é considerada grande o suficiente para que este efeito tenha sido reduzido a um patamar aceitável.

Outra ameaça está relacionada ao contexto em que o experimento foi aplicado e ao perfil de seus participantes, haja vista que este experimento contou com a participação de alunos de graduação que atuam como desenvolvedores júnior em um Laboratório de desenvolvimento de software. Consequentemente, os resultados do mesmo estão vinculados ao contexto em que foi aplicado e ao tipo de especialistas de domínio, que ainda são estudantes. Por fim, das quatro expressões ponderadas disponíveis ao método RNM, foram consideradas apenas três, dado que a solução utilizada foi a proposta em Nunes et al. [38], que ainda não suporta a expressão ponderada *MIXMINMAX*.

# Capítulo 5

## Discussão

Nesta seção são discutidos os resultados obtidos no estudo de caso e no experimento, traçando assim um paralelo entre os resultados apresentados nos Capítulos 3 e 4. Na Seção 5.1 são discutidos os resultados acerca da acurácia dos métodos, de acordo com o primeiro objetivo de pesquisa. Na Seção 5.2 são discutidos os resultados acerca da capacidade de modelagem dos métodos, de acordo com o segundo objetivo de pesquisa.

### 5.1 Acurácia

O estudo de caso apresentado no Capítulo 3 contou com duas unidades de análise, Caso A e Caso B. Em relação aos resultados obtidos com o passo a passo cognitivo do modelo, no Caso A o método WSA apresentou uma acurácia de 100% e o método RNM uma acurácia de 57,14%, no Caso B o método WSA apresentou uma acurácia de 75% e o método RNM uma acurácia de 50%. Em relação à avaliação com dados históricos, os dois métodos alcançaram resultados equivalentes em ambas unidades de análise.

É importante salientar, no entanto, que neste caso, a ameaça à validade indicada na Seção 3.6 do Capítulo 3, pode ter enviesado os resultados de alguma forma, de modo que não é possível descartar completamente a possibilidade do método WSA ter tido uma leve vantagem, haja vista que os especialistas de domínio foram acompanhados por especialistas em redes Bayesianas durante o processo de definição dos parâmetros para aplicação do método WSA, e o mesmo não ocorreu com o método RNM, que devido a incompatibilidade de horário dos participantes do estudo de caso, foi aplicado de forma remota.



Em contrapartida, no estudo piloto descrito na Seção 4.4.4 do Capítulo 4, os dois métodos foram aplicados em condições homogêneas, na qual os especialistas de domínio forneceram as probabilidade de entrada de ambos os métodos de forma remota, sem o acompanhamento dos especialistas em redes Bayesianas. Dessa forma, os métodos foram comparados a partir da aplicação do passo a passo cognitivo (também de forma remota). Ao considerarmos os resultados deste, o RNM foi superior ao WSA, apresentando uma acurácia de 68% contra 54% do WSA.

Isto posto, com os resultados expostos até o momento, ainda não seria possível chegar a uma conclusão de qual dos métodos apresenta maior acurácia, apenas que são equivalentes, haja vista que o método WSA foi melhor no estudo de caso e o método RNM foi melhor no estudo piloto (considerando o passo a passo cognitivo dos modelos). Em relação aos resultados obtidos com dados históricos, no entanto, é possível afirmar, diante dos resultados apresentados, que os dois métodos são equivalentes.

Em relação aos resultados obtidos no experimento apresentado no Capítulo 4, por sua vez, considerando o gráfico das diferenças nas médias de acurácia apresentados na Figura 4.12d do Capítulo 4 é possível observar que o intervalo de confiança entre os métodos WSA e RNM se estende de -0.2353 à -0.0181. Este intervalo não inclui o zero, o que indica que a diferença entre as médias é estatisticamente significativa. Ademais, este resultado é favorável ao método RNM.

Dessa forma, é possível afirmar com um nível de confiança individual de 98,92% que a diferença média na acurácia do método WSA em relação ao método RNM é de -12,67%. Esta diferença é equivalente àquela das médias obtidas no estudo piloto (-14% de diferença do método WSA em relação ao método RNM). Assim sendo, levando em consideração os resultados obtidos no estudo de caso e no experimento controlado, conclui-se que o método RNM foi superior ao método WSA em termos de acurácia quando aplicados diretamente por especialistas de domínio.

Em relação aos métodos RNM e AHP, considerando as diferenças nas médias de acurácia apresentados na Figura 4.12d no Capítulo 4, os intervalos de confiança para as diferenças entre as médias dos mesmos incluem o zero, o que indica que as diferenças entre as médias não são estatisticamente significativas. Em outras palavras, não é possível afirmar que haja diferenças entre os métodos RNM e AHP em relação à acurácia. Entretanto, na prática, é

possível observar a superioridade do método RNM.

Por outro lado, considerando as diferenças nas médias de acurácia apresentados na Figura 4.12d no Capítulo 4 é possível observar que o intervalo de confiança entre o método RNM e os métodos WSA e WSA-AHP não incluem o valor zero, o que indica que a diferença entre essas médias é estatisticamente significativa. A diferença entre os métodos RNM e WSA já foi discutida, por este motivo será discutido em seguida os resultados referentes a diferença entre as médias do método RNM e WSA-AHP.

No gráfico apresentado na Figura 4.12d do Capítulo 4 é possível observar que o intervalo de confiança entre o método WSA-AHP e RNM se estende de -0.2688 à -0.0516. Assim como aconteceu na comparação entre WSA e RNM, este intervalo não inclui o zero, o que indica que a diferença entre essas médias é estatisticamente significativa. Portanto, é possível afirmar com um nível de confiança individual de 98,92% que a diferença média na acurácia entre os métodos WSA-AHP e RNM é de -16,02%. Essa diferença é superior àquela relatada para as médias obtidas entre WSA e RNM.

De um ponto de vista prático, o método RNM obteve melhores resultados em relação aos demais métodos, mas houve diferença estatisticamente significante apenas entre este e os métodos WSA e WSA-AHP. Por outro lado, apesar do método AHP ter sido o único que não obteve resultados significativamente diferentes do método RNM, também não teve diferença significativa quando comparado aos métodos WSA e WSA-AHP (assim como teve o método RNM). Dessa forma, de modo geral o método RNM apresentou o melhor resultado, seguido do método AHP, WSA e WSA-AHP.

## 5.2 Capacidade de modelagem

No que tange a capacidade de modelagem de TPN, os intervalos de confiança para as diferenças entre as médias dos métodos WSA, WSA-AHP e RNM incluem o zero, o que indica que as diferenças entre as médias não são estatisticamente significativas. Em outras palavras, não é possível afirmar que haja diferenças na capacidade de modelagem dos métodos WSA, WSA-AHP e RNM. Por outro lado, o método AHP apresentou uma capacidade de modelagem inferior com diferenças significativas em relação aos demais métodos.

Sob outra perspectiva, os resultados obtidos com os métodos WSA e WSA-AHP suge-

rem que a extensão proposta para o método WSA com a integração de elementos do método AHP (i.e., elicitación de probabilidades de forma indireta) é capaz de alcançar resultados similares aos resultados obtidos com o método WSA em sua forma original. Este resultado é promissor, pois indica para a possibilidade do uso do método WSA em domínios em que os especialistas de domínio não se sentem confortáveis com a definição direta de probabilidades.

De um modo geral, apesar de não haver diferença estatisticamente significativa entre os métodos RNM, WSA e WSA-AHP, de um ponto de vista prático o método RNM alcançou melhores resultados, seguido dos métodos WSA, WSA-AHP e AHP. De acordo com os resultados confrontantes do método AHP, que obteve uma maior acurácia em relação aos métodos WSA e WSA-AHP mas apresentou capacidade inferior de modelagem de TPN, observa-se que sua capacidade de definir corretamente a moda dos vetores de probabilidade é comparável a do método RNM, mas sua capacidade de expressar a incerteza dos especialistas é limitada (e.g., no RNM o parâmetro  $\sigma^2$  expressa o grau de incerteza).

# Capítulo 6

## Considerações Finais

Neste trabalho, foi realizado um estudo empírico comparativo entre métodos de geração semiautomática de TPN. O propósito deste trabalho foi medir e comparar a capacidade de modelagem e acurácia dos métodos no que diz respeito à tarefa de geração de TPN. Para isto, foi realizado um estudo de caso com duas unidades de análise, no qual os métodos RNM e WSA foram comparados à partir de dados coletados na validação dos modelos construídos, e um experimento controlado, no qual TPN geradas pelos métodos RNM, WSA, WSA-AHP e AHP foram comparadas com TPN de referência definidas pelos próprios especialistas de domínio.

No estudo de caso, foram construídas duas redes Bayesianas para estimativa de valor em um contexto de engenharia de software baseado em valor, utilizando conhecimento de especialistas de domínio. As redes Bayesianas construídas foram validadas com passo a passo cognitivo do modelo e dados históricos de projetos relacionados às redes desenvolvidas, disponíveis em um banco de dados de uma ferramenta web. Participaram deste estudo de caso três gerentes de projetos e um desenvolvedor. Os dados obtidos com a validação dos modelos foram utilizados como parâmetros para comparar os métodos RNM e WSA.

No experimento controlado, dez desenvolvedores de software aplicaram os métodos RNM, WSA, WSA-AHP e AHP para quantificar incertezas em uma rede Bayesiana para avaliação da coesão de equipes ágeis de desenvolvimento de software. As redes Bayesianas quantificadas pelos métodos foram então comparadas com redes Bayesianas de referência, também construídas pelos desenvolvedores. A rede Bayesiana de referência teve suas TPN definidas de forma manual.

Os critérios utilizados para comparar os métodos foram: capacidade de modelagem de TPN e acurácia. Para comparar a capacidade de modelagem foram utilizadas as medidas de similaridade: *Brier Score*; Distância Euclidiana; e Erro Médio Absoluto. Para comparar a acurácia dos métodos foi utilizada a proporção de acertos da moda de cada vetor de probabilidade das TPN.

No que tange a capacidade de modelagem de TPN, não houve diferenças estatisticamente significantes entre os métodos RNM, WSA e WSA-AHP, mas houve entre o método AHP e os demais (AHP apresentou capacidade inferior de modelagem de TPN). No que se refere à acurácia, o método RNM apresentou diferenças estatisticamente significantes em relação aos métodos WSA e WSA-AHP, mas não com relação ao método AHP.

Para aplicação do método RNM foi adotada uma nova abordagem, denominada “Especialista Simulado”. Esta abordagem se mostrou promissora quando utilizada para identificar os dados de entrada para o algoritmo do RNM, removendo assim a necessidade de um especialista em RNM para aplicar o método.

A combinação dos métodos WSA e AHP também apresentou bons resultados, mostrando assim, a viabilidade do uso de métodos indiretos de elicitación de probabilidades que podem ser utilizados para estender o método WSA, viabilizando sua aplicação em domínios nos quais os especialistas de domínio não se sentem confortáveis com a definição direta de probabilidades.

De um ponto de vista prático, o método RNM sobressaiu-se aos demais. Talvez este resultado possa ser explicado pelo menor número de parâmetros necessários para aplicar o método, que reduz o número de interações necessárias com os especialistas de domínio para gerar as TPN. Além disso, é possível que vieses dos especialistas possam ser ainda transferidos na forma de dados de entrada aos métodos semiautomáticos, de forma que um menor número de parâmetros de entrada pode, neste caso, minimizar a incidência dos mesmos.

Isto posto, diante dos resultados aferidos não é possível descartar a utilidade dos outros métodos, que também podem ser utilizados na prática. Todavia, é sugerido que antes de escolher um método semiautomático para quantificar incertezas durante o processo de construção de redes Bayesianas, seja feita uma análise das especificidades de cada um deles, em especial do método AHP.

A principal contribuição deste trabalho consiste na disponibilização de evidências em-

---

píricas acerca da capacidade de modelagem e acurácia de quatro métodos de geração semiautomática de TPN. Estes métodos foram comparados em um contexto comum, compartilhado por múltiplos especialistas de domínio que participaram desta pesquisa. Dessa forma, espera-se que os resultados obtidos sirvam como parâmetros úteis à tomada de decisão, especialmente de especialistas em redes Bayesianas, quando da escolha de um método semiautomático para quantificação de incertezas em redes Bayesianas.

Outras contribuições desta pesquisa são: uma nova abordagem para aplicação do método RNM sem que seja necessário a presença de um especialista em redes Bayesianas; uma extensão ao método WSA para elicitação de probabilidades de forma indireta utilizando o método AHP; a disponibilização de uma ferramenta gratuita para criação de redes Bayesianas com suporte ao método RNM; disponibilização de um serviço web para realização de cálculos do método WSA. De um modo geral, é disponibilizado um arcabouço para aplicar os métodos e replicar este estudo.

Como trabalho futuro, fica em aberto a possibilidade de avaliar os métodos usando redes Bayesianas com informações sintéticas e sobre o mundo real, já estabelecidas na literatura. O desafio, neste caso, consiste em descobrir como aplicar o método WSA, especificamente, como descobrir as configurações de pais compatíveis a serem utilizadas. Tal abordagem viabilizaria a avaliação de um número maior de métodos em variados contextos, haja vista que não seria necessária a presença de especialistas de domínio para a aplicação dos métodos.

# Bibliografia

- [1] Simon Baker. *Towards the construction of large Bayesian networks for Web cost estimation*. PhD thesis, University of Auckland, 2009.
- [2] Simon Baker and Emilia Mendes. Assessing the weighted sum algorithm for automatic generation of probabilities in bayesian networks. In *Information and Automation (ICIA), 2010 IEEE International Conference on*, pages 867–873. IEEE, 2010.
- [3] Simon Baker and Emilia Mendes. Evaluating the weighted sum algorithm for estimating conditional probabilities in bayesian networks. In *SEKE*, volume 2010, pages 319–324, 2010.
- [4] Barry Boehm. Value-based software engineering: reinventing. *ACM SIGSOFT Software Engineering Notes*, 28(2):3, 2003.
- [5] Kwai-Sang Chin, Da-Wei Tang, Jian-Bo Yang, Shui Yee Wong, and Hongwei Wang. Assessing new product development project risk by bayesian network with a systematic probability generation methodology. *Expert Systems with Applications*, 36(6):9879–9890, 2009.
- [6] Lee G Cooper. Strategic marketing planning for radically new products. *Journal of marketing*, 64(1):1–16, 2000.
- [7] Balaram Das. Generating conditional probabilities for bayesian networks: Easing the knowledge acquisition problem. *arXiv preprint cs/0411034*, 2004.
- [8] Balaram Das. Generating Conditional Probabilities for Bayesian Networks: Easing the Knowledge Acquisition Problem. *Computing Research Repository*, cs.AI/0411034, 2004.

- 
- [9] Ana CV de Melo and Adilson J Sanchez. Software maintenance project delays prediction using bayesian networks. *Expert Systems with Applications*, 34(2):908–919, 2008.
- [10] MJ Druzdel and Linda C Van Der Gaag. Building probabilistic networks:"where do the numbers come from?". *IEEE Transactions on knowledge and data engineering*, 12(4):481–486, 2000.
- [11] Norman Fenton, William Marsh, Martin Neil, Patrick Cates, Simon Forey, and Manesh Tailor. Making resource decisions for software projects. In *Software Engineering, 2004. ICSE 2004. Proceedings. 26th International Conference on*, pages 397–406. IEEE, 2004.
- [12] Norman Fenton, Martin Neil, and Jose Galan Caballero. Using ranked nodes to model qualitative judgments in bayesian networks. *IEEE Trans. on Knowl. and Data Eng.*, 19(10):1420–1432, October 2007.
- [13] Norman Fenton, Martin Neil, William Marsh, Peter Hearty, David Marquez, Paul Krause, and Rajat Mishra. Predicting software defects in varying development lifecycles using bayesian nets. *Information and Software Technology*, 49(1):32–43, 2007.
- [14] Norman E Fenton, Martin Neil, and Jose Galan Caballero. Using ranked nodes to model qualitative judgments in bayesian networks. *IEEE Transactions on Knowledge & Data Engineering*, (10):1420–1432, 2007.
- [15] Arthur Freire, Mirko Perkusich, Renata Saraiva, Hyggo Almeida, and Angelo Perkusich. A bayesian networks-based approach to assess and improve the teamwork quality of agile teams. *Information and Software Technology*, 100:119–132, 2018.
- [16] Vitor Freitas, Emilia Mendes, and Burak Turhan. Providing tool-support for value-based decision-making: a usability assessment. In *Software Engineering and Advanced Applications (SEAA), 2016 42th Euromicro Conference on*, pages 34–41. IEEE, 2016.
- [17] Vitor Freitas, Mirko Perkusich, Emilia Mendes, Pilar Rodríguez, and Markku Oivo. Value-based decision-making using a web-based tool: A multiple case study. In *Asia-*



- Pacific Software Engineering Conference (APSEC), 2017 24th*, pages 279–288. IEEE, 2017.
- [18] Nir Friedman, Dan Geiger, and Moisés Goldszmidt. Bayesian network classifiers. *Machine Learning*, 29(2-3):131–163, 1997.
- [19] Wolfgang Garn and Panos Louvieris. Conditional probability generation methods for high reliability effects-based decision making. *arXiv preprint arXiv:1512.08553*, 2015.
- [20] José P González-Brenes, John T Behrens, Robert J Mislevy, Roy Levy, and Kristen E DiCerbo. Bayesian networks. *The Wiley Handbook of Cognition and Assessment: Frameworks, Methodologies, and Applications*, page 328, 2016.
- [21] David Heckerman. Learning in graphical models. chapter A Tutorial on Learning with Bayesian Networks, pages 301–354. MIT Press, Cambridge, MA, USA, 1999.
- [22] Andreas Jedlitschka and Dietmar Pfahl. Reporting guidelines for controlled experiments in software engineering. In *Empirical Software Engineering, 2005. 2005 International Symposium on*, pages 10–pp. IEEE, 2005.
- [23] Daniel Kahneman and Amos Tversky. The simulation heuristic. Technical report, STANFORD UNIV CA DEPT OF PSYCHOLOGY, 1981.
- [24] JinHyung Kim and Judea Pearl. A computational model for causal and diagnostic reasoning in inference systems. In *International Joint Conference on Artificial Intelligence*, pages 0–0, 1983.
- [25] Pekka Laitila. Improving the use of ranked nodes in the elicitation of conditional probabilities for bayesian networks. 2013.
- [26] Pekka Laitila and Kai Virtanen. Improving construction of conditional probability tables for ranked nodes in bayesian networks. *IEEE Transactions on Knowledge and Data Engineering*, 28(7):1691–1705, 2016.
- [27] Eunchang Lee, Yongtae Park, and Jong Gye Shin. Large engineering project risk management using a bayesian belief network. *Expert Syst. Appl.*, 36(3):5880–5887, April 2009.

- [28] Lillian Lee. On the effectiveness of the skew divergence for statistical language analysis. In *AISTATS*. Citeseer, 2001.
- [29] Emilia Mendes. Using knowledge elicitation to improve web effort estimation: lessons from six industrial case studies. In *Proceedings of the 34th International Conference on Software Engineering*, pages 1112–1121. IEEE Press, 2012.
- [30] Emilia Mendes. *Practitioner’s Knowledge Representation*. Springer, 2014.
- [31] Emilia Mendes, Mirko Perkusich, Vitor Freitas, and João Nunes. Using bayesian network to estimate the value of decisions within the context of value-based software engineering. In *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018*, pages 90–100. ACM, 2018.
- [32] Emilia Mendes, Pilar Rodriguez, Vitor Freitas, Simon Baker, and Mohamed Amine Atoui. Towards improving decision making and estimating the value of decisions in value-based software engineering: the value framework. *Software Quality Journal*, pages 1–50, 2017.
- [33] Emilia Mendes, Burak Turhan, Pilar Rodríguez, and Vitor Freitas. Estimating the value of decisions relating to managing and developing software-intensive products and projects. In *Proceedings of the 11th International Conference on Predictive Models and Data Analytics in Software Engineering*, page 7. ACM, 2015.
- [34] Lusine Mkrtchyan, Luca Podofillini, and Vinh N Dang. Methods for building conditional probability tables of bayesian belief networks from limited judgment: an evaluation for human reliability application. *Reliability Engineering & System Safety*, 151:93–112, 2016.
- [35] Stefano Monti and Giuseppe Carenini. Dealing with the expert inconsistency in probability elicitation. *IEEE Transactions on Knowledge and Data Engineering*, 12(4):499–508, 2000.
- [36] Martin Neil, Norman Fenton, and Manesh Tailor. Using bayesian networks to model expected and unexpected operational losses. *Risk Analysis*, 25(4):963–972, 2005.

- [37] Martin Neil, Bob Malcolm, and Roger Shaw. Modelling an air traffic control environment using bayesian belief networks. In *21st International System Safety Conference, Ottawa, Ontario, Canada, 2003*.
- [38] Joao Nunes, Renan Willamy, Mirko Perkusich, Renata Saraiva, Kyller Gorgonio, Hyggo Almeida, and Angelo Perkusich. An algorithm to define the node probability functions of bayesian networks based on ranked nodes.
- [39] J. Pearl and S. Russell. Bayesian networks. *Handbook of brain theory and neural networks*, 1995.
- [40] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier, 2014.
- [41] Mirko Perkusich, Kyller Gorgonio, Hyggo Almeida, and Angelo Perkusich. Assisting the continuous improvement of scrum projects using metrics and bayesian networks. *Journal of Software: Evolution and Process*, 2016. Article in Press.
- [42] Mirko Perkusich, Gustavo Soares, Hyggo Almeida, and Angelo Perkusich. A procedure to detect problems of processes in software development projects using bayesian networks. *Expert Systems with Applications*, 42(1):437 – 450, 2015.
- [43] Silja Renooij and Caecilia Liesbeth Maria Witteman. *Talking probabilities: communicating probabilistic information with words and numbers*, volume 1999. Utrecht University: Information and Computing Sciences, 1999.
- [44] Per Runeson and Martin Höst. Guidelines for conducting and reporting case study research in software engineering. *Empirical software engineering*, 14(2):131, 2009.
- [45] Thomas L Saaty. How to make a decision: the analytic hierarchy process. *Interfaces*, 24(6):19–43, 1994.
- [46] Roland W Scholz and Olaf Tietje. *Embedded case study methods: Integrating quantitative and qualitative knowledge*. Sage, 2002.
- [47] Jeff Sutherland and Ken Schwaber. The scrum guide. *The definitive guide to scrum: The rules of the game. Scrum.org*, 268, 2013.

- 
- [48] Amos Tversky and Daniel Kahneman. Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2):207–232, 1973.
- [49] BW Wisse, Sicco Pier van Gosliga, Nicole P van Elst, and Ana Isabel Barros. Relieving the elicitation burden of bayesian belief networks. In *BMA*, 2008.
- [50] Owen Woodberry, Ann E Nicholson, Kevin B Korb, and Carmel Pollino. Paramete-  
rising bayesian networks. In *Australasian Joint Conference on Artificial Intelligence*,  
pages 1101–1107. Springer, 2004.
- [51] Adam Zagorecki and Marek J Druzdzel. An empirical study of probability elicitation  
under noisy-or assumption. In *Flairs conference*, pages 880–886, 2004.
- [52] Adam Zagorecki and Marek J Druzdzel. Knowledge engineering for bayesian  
networks: How common are noisy-max distributions in practice? *IEEE Transactions  
on Systems, Man, and Cybernetics: Systems*, 43(1):186–195, 2013.

## **Apêndice A**

### **Artefatos Utilizados no Experimento**

## Autogerenciamento

Equipes que são adaptáveis, possuem liderança compartilhada e redundância.

Adaptabilidade		
A adaptabilidade da equipe é sua habilidade de identificar mudanças no ambiente de trabalho e ajustar-se as estratégias.		
Baixa	Razoável	Alta
A equipe <b>não consegue ao menos identificar mudanças necessárias</b> ao ambiente de trabalho.	A equipe <b>consegue identificar mudanças necessárias ao ambiente de trabalho.</b>	A equipe <b>consegue identificar mudanças necessárias ao ambiente de trabalho.</b>
<b>A equipe não consegue ajustar-se à novas estratégias.</b>	A equipe <b>nem sempre consegue ajustar-se à novas estratégias.</b>	A equipe <b>consegue ajustar-se à novas estratégias com facilidade.</b>

Expertise/Conhecimento		
A soma de conhecimento dos indivíduos que formam a equipe (conhecimento); A capacidade dos membros da equipe de realizar as tarefas de outros e substituí-los caso seja necessário (redundância); A equipe possui todas as competências necessárias para completar o trabalho sem depender de outros que não fazem parte da equipe (multifuncional).		
Baixa	Razoável	Alta
A equipe <b>não é capaz de substituir ou realizar as atividades de um membro ausente.</b>	A equipe <b>não é capaz de substituir ou realizar as atividades de um membro ausente.</b>	A equipe <b>consegue realizar todas as tarefas de um membro ausente e se necessário, substituí-lo.</b>
A equipe <b>depende de terceiros para completar o trabalho.</b>	A equipe <b>não depende de terceiros para completar o trabalho.</b>	A equipe <b>não depende de terceiros para completar o trabalho.</b>

Liderança compartilhada		
Liderança compartilhada trata-se de compartilhar a autoridade para tomar decisões.		
Baixa	Razoável	Alta
As <b>decisões</b> são <b>centralizadas</b> em uma única pessoa.	As <b>decisões são centralizadas</b> em uma única pessoa.	As <b>decisões e a liderança são compartilhadas</b> por toda equipe.
<b>As opiniões</b> de membros da equipe <b>não são levadas em consideração.</b>	As <b>opiniões</b> dos membros da equipe <b>são levadas em consideração.</b>	Todos têm <b>poder de decisão</b> nas tarefas que lhes cabem.

## Coesão da Equipe

A coesão da equipe refere-se à atração interpessoal dos membros da equipe, o comprometimento dos mesmos com as atividades da equipe e o espírito de equipe.

Colaboração		
Para uma equipe colaborar, a mesma deve ser coordenada. A coordenação da equipe refere-se aos membros da equipe executarem suas atividades de forma sincronizada e integrada. Além disso, ela deve ser orientada. A orientação trata-se do respeito mútuo entre os membros das equipes e priorização dos objetivos da equipe sobre os objetivos individuais. Para uma equipe coordenar suas atividades, é necessário se comunicar.		
Baixa	Razoável	Alta
A equipe <b>não realiza reuniões diárias para sincronizar as atividades.</b>	A equipe <b>as vezes realiza as reuniões diárias para sincronizar as atividades.</b>	A equipe <b>realiza reuniões diárias para sincronizar as atividades.</b>
A equipe <b>é distribuída.</b>	A equipe <b>trabalha no mesmo local físico.</b>	A equipe <b>trabalha no mesmo local físico.</b>
A equipe <b>não possui canais de comunicação.</b>	A equipe <b>não possui canais de comunicação.</b>	A equipe <b>possui canais de comunicação</b> como o slack.
Os membros da equipe <b>não se respeitam.</b>	Os membros da equipe <b>se respeitam.</b>	Os membros da equipe <b>se respeitam.</b>
<b>Objetivos individuais sobrepõem os objetivos coletivos</b> da equipe.	<b>Objetivos individuais sobrepõem os objetivos coletivos</b> (da equipe).	<b>Objetivos da equipe sobrepõem os objetivos individuais.</b>

Autogerenciamento		
Equipes que são adaptáveis, possuem liderança compartilhada e redundância.		
Baixa	Razoável	Alta
<b>Dependem de terceiros</b> para completar o trabalho.	Os membros do time possuem o <b>conhecimento necessário para completar o trabalho, mas não há redundância na equipe.</b>	Os membros da equipe têm o <b>conhecimento necessário para completar o trabalho, com redundância.</b>
As <b>decisões</b> são <b>concentradas em um único membro da equipe</b> que <b>não considera as opiniões de outros.</b>	As <b>decisões</b> são <b>concentradas em um único membro</b> que <b>leva em consideração a opinião de outros.</b>	As <b>decisões e a liderança</b> são <b>compartilhadas</b> por toda equipe.
A equipe <b>não consegue ao menos identificar mudanças necessárias</b> ao ambiente de trabalho.	A equipe <b>consegue identificar mudanças necessárias ao ambiente de trabalho.</b>	A equipe <b>consegue identificar mudanças necessárias ao ambiente de trabalho.</b>
<b>A equipe não consegue ajustar-se à novas estratégias.</b>	A equipe <b>nem sempre consegue ajustar-se à novas estratégias.</b>	A equipe <b>consegue ajustar-se à novas estratégias com facilidade.</b>

## Colaboração

Para uma equipe colaborar, a mesma deve ser coordenada. A coordenação da equipe refere-se aos membros da equipe executarem suas atividades de forma sincronizada e integrada. Além disso, ela deve ser orientada. A orientação trata-se do respeito mútuo entre os membros das equipes e priorização dos objetivos da equipe sobre os individuais. Para uma equipe coordenar suas atividades, é necessário se comunicar.

Coordenação		
A coordenação da equipe refere-se aos membros da equipe executarem suas atividades de forma sincronizada e integrada. Para uma equipe coordenar suas atividades, é necessário se comunicar. A qualidade da comunicação da equipe depende dos canais de comunicação e distribuição da equipe.		
Baixa	Razoável	Alta
A equipe <b>não realiza reuniões diárias para sincronizar as atividades.</b>	A equipe <b>as vezes realiza as reuniões diárias para sincronizar as atividades.</b>	A equipe <b>realiza reuniões diárias para sincronizar as atividades.</b>
A equipe <b>é distribuída.</b>	A equipe <b>trabalha no mesmo local físico.</b>	A equipe <b>trabalha no mesmo local físico.</b>
A equipe <b>não possui canais de comunicação.</b>	A equipe <b>não possui canais de comunicação.</b>	A equipe <b>possui canais de comunicação</b> como o slack.

Orientação		
A orientação trata-se do respeito mútuo entre os membros da equipe e priorização dos objetivos da equipe sobre os individuais [1].		
Baixa	Razoável	Alta
Os membros da equipe <b>não se respeitam.</b>	Os membros da equipe <b>se respeitam.</b>	Os membros da equipe <b>se respeitam.</b>
Os membros da equipe <b>não possuem características em comum e apresentam dificuldades de convivência.</b>	Os membros da equipe <b>não possuem muitas características em comum, mas não apresentam problemas sérios de convivência ou há um membro conciliador.</b>	Os membros da equipe <b>possuem características em comum e apresentam um ótimo relacionamento.</b>
<b>Objetivos individuais sobrepõem os objetivos coletivos</b> da equipe.	<b>Objetivos individuais sobrepõem os objetivos coletivos</b> da equipe.	<b>Objetivos da equipe sobrepõem os objetivos individuais.</b>