

UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

**Seleção de covariância para o modelo grafo gaussiano via
*reversible jump***

Eriton Barros dos Santos

Tese de Doutorado do Programa Interinstitucional de Pós-Graduação em
Estatística (PIPGEs)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Eriton Barros dos Santos

Seleção de covariância para o modelo grafo gaussiano via
reversible jump

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP e ao Departamento de Estatística – DEs-UFSCar, como parte dos requisitos para obtenção do título de Doutor em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística. *VERSÃO REVISADA*

Área de Concentração: Estatística

Orientador: Prof. Dr. Luis Aparecido Milan

USP – São Carlos
Abril de 2023

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

S237s Santos, Eriton Barros
Seleção de covariância para o modelo grafo
gaussiano via reversible jump / Eriton Barros
Santos; orientador Luis Aparecido Milan. -- São
Carlos, 2023.
104 p.

Tese (Doutorado - Programa Interinstitucional de
Pós-graduação em Estatística) -- Instituto de Ciências
Matemáticas e de Computação, Universidade de São
Paulo, 2023.

1. Estatística. 2. Modelo grafo. 3. Modelo grafo
gaussiano. 4. Inferência bayesiana. 5. Algoritmo
reversible jump. I. Milan, Luis Aparecido, orient.
II. Título.

Eriton Barros dos Santos

Covariance selection for graphical gaussian model via
reversible jump

Thesis submitted to the Institute of Mathematics
and Computer Science – ICMC-USP and to the
Department of Statistics – DEs-UFSCar – in
accordance with the requirements of the Statistics
Interagency Graduate Program, for the degree of
Doctor in Statistics. *FINAL VERSION*

Concentration Area: Statistics

Advisor: Prof. Dr. Luis Aparecido Milan

USP – São Carlos
April 2023

AGRADECIMENTOS

Primeiramente, aos meus pais, Eridan e Nilton, minha irmã, Sarah, e minha namorada, Terezinha, por todo o apoio e amor e por acreditarem em minhas escolhas.

Ao meu orientador, Luis Aparecido Milan, por contribuir de forma significativa para a minha formação acadêmica, como, também, pelas sugestões, contribuições e por acreditar que fosse capaz de desenvolver esse trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

RESUMO

SANTOS, E. B. **Seleção de covariância para o modelo grafo gaussiano via *reversible jump***. 2023. 104 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2023.

O objetivo do modelo grafo gaussiano consiste em encontrar a estrutura de covariância que representa a relação entre variáveis aleatórias, cuja distribuição conjunta é uma normal multivariada. Isso é uma ferramenta usada para modelar grafos gaussianos. A inferência de parâmetros desse tipo de modelagem geralmente é baseada na estimativa da máxima verossimilhança. No entanto, esse tipo de metodologia requer o ajuste de todos os modelos possíveis para verificar qual modelo melhor representa a relação entre as variáveis. No caso de qualquer modelo, dentre todos os possíveis, apresentar problema de estimativa, o resultado pode não representar a verdadeira relação entre as variáveis. Nós propomos alterações no procedimento baseado no algoritmo *Reversible Jump* de Dobra et al. (2011) para seleção e ajuste do modelo grafo gaussiano. Nós também criamos indicadores para avaliar resultados de simulações provenientes de um modelo grafo gaussiano. Os resultados obtidos no trabalho são favoráveis a nossa proposta apresentada, no qual observou um melhoramento no método de seleção de modelo diminuindo o erro ao buscar a estrutura de covariância.

Palavras-chave: Modelo Grafo Gaussiano, Modelo Grafo, Grafo Gaussiano, Metropolis-Hastings, *Reversible Jump*, Seleção de Modelo, Seleção de Covariância, Correlação Parcial.

ABSTRACT

SANTOS, E. B. **Covariance selection for graphical gaussian model via reversible jump.** 2023. 104 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2023.

The purpose of the Graphical Gaussian model is to find the covariance structure that represents the relationship between random variables, whose joint distribution is a multivariate normal. This is a tool used to modeling Gaussian graphs. The inference of parameters of this type of modeling is commonly based on maximum likelihood estimation. However, this type of methodology requires the adjustment of all possible models to verify which model best represents the relationship between the variables. In case any model, among all the possibilities, presents an estimation problem, the result may not represent the true relationship between the variables. We propose alterations in the procedure based on the Reversible Jump algorithm of Dobra et al. (2011) for selecting and fitting the Graphical Gaussian model. We also create indicators to evaluate simulation results from a Graphical Gaussian model. The results obtained in this work are favorable to our proposal presented, in which an improvement in the model selection method was observed, reducing the error when searching for the covariance structure.

Keywords: Graphical Gaussian Model, Graphical Model, Gaussian Graphs, Metropolis-Hastings, Reversible Jump, Model Selection, Covariance Selection Model, Partial Correlation.

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Objetivo	17
1.2	Organização do Trabalho	17
2	MODELO GRAFO GAUSSIANO	19
2.1	Conceitos Gerais	19
2.2	Modelo Grafo	20
2.2.1	<i>Modelo Grafo Gaussiano</i>	20
2.2.1.1	<i>Modelo Grafo Gaussiano de Dobra, Lenkoski e Rodriguez</i>	20
3	ESTIMAÇÃO VIA METROPOLIS-HASTINGS	23
3.1	Definição	23
3.2	Metropolis-Hastings no Modelo Grafo Gaussiano	24
4	REVERSIBLE JUMP	25
4.1	Definição	25
4.2	<i>Reversible Jump</i> no Modelo Grafo Gaussiano	26
5	SIMULAÇÃO	31
5.1	Simulação de Dados	31
5.2	Critério de Avaliação e Comparação	32
5.2.1	<i>Critério Quantitativo</i>	32
5.2.1.1	<i>Contagem de Permanência das Arestas</i>	32
5.2.1.2	<i>Probabilidade a Posteriori de Permanência das Arestas</i>	32
5.2.1.3	<i>Índice de Erro</i>	32
5.2.1.4	<i>Soma Absoluta dos Erros</i>	33
5.2.1.5	<i>Soma Quadrática dos Erros</i>	33
5.2.1.6	<i>Índices de Semelhança</i>	34
5.2.2	<i>Critério Gráfico</i>	34
5.2.2.1	<i>Gráfico de Linhas para os Determinantes</i>	34
5.2.2.2	<i>Gráfico de Avaliação das Probabilidades a Posteriori</i>	35
5.3	Resultados das Simulações	35
5.3.1	<i>Resultados no cenário com $m = 100.000$ e $n = 100$</i>	36
5.3.1.1	<i>Algoritmo de Dobra et al. ($m = 100.000$ e $n = 100$)</i>	36

5.3.1.2	Algoritmo da Proposta 1 ($m = 100.000$ e $n = 100$)	38
5.3.1.3	Algoritmo da Proposta 2 ($m = 100.000$ e $n = 100$)	39
5.3.1.4	Algoritmo da Proposta 3 ($m = 100.000$ e $n = 100$)	40
5.3.1.5	Comparação dos Algoritmos ($m = 100.000$ e $n = 100$)	42
5.3.2	Resultados no cenário com $m = 100.000$ e $n = 250$	43
5.3.2.1	Algoritmo de Dobra et al. ($m = 100.000$ e $n = 250$)	43
5.3.2.2	Algoritmo da Proposta 1 ($m = 100.000$ e $n = 250$)	44
5.3.2.3	Algoritmo da Proposta 2 ($m = 100.000$ e $n = 250$)	46
5.3.2.4	Algoritmo da Proposta 3 ($m = 100.000$ e $n = 250$)	47
5.3.2.5	Comparação dos Algoritmos ($m = 100.000$ e $n = 250$)	48
5.3.3	Resultados no cenário com $m = 100.000$ e $n = 500$	50
5.3.3.1	Algoritmo de Dobra et al. ($m = 100.000$ e $n = 500$)	50
5.3.3.2	Algoritmo da Proposta 1 ($m = 100.000$ e $n = 500$)	51
5.3.3.3	Algoritmo da Proposta 2 ($m = 100.000$ e $n = 500$)	52
5.3.3.4	Algoritmo da Proposta 3 ($m = 100.000$ e $n = 500$)	53
5.3.3.5	Comparação dos Algoritmos ($m = 100.000$ e $n = 500$)	54
5.3.4	Resultados no cenário com $m = 200.000$ e $n = 100$	56
5.3.4.1	Algoritmo de Dobra et al. ($m = 200.000$ e $n = 100$)	56
5.3.4.2	Algoritmo da Proposta 1 ($m = 200.000$ e $n = 100$)	58
5.3.4.3	Algoritmo da Proposta 2 ($m = 200.000$ e $n = 100$)	59
5.3.4.4	Algoritmo da Proposta 3 ($m = 200.000$ e $n = 100$)	60
5.3.4.5	Comparação dos Algoritmos ($m = 200.000$ e $n = 100$)	61
5.3.5	Resultados no cenário com $m = 200.000$ e $n = 250$	63
5.3.5.1	Algoritmo de Dobra et al. ($m = 200.000$ e $n = 250$)	63
5.3.5.2	Algoritmo da Proposta 1 ($m = 200.000$ e $n = 250$)	64
5.3.5.3	Algoritmo da Proposta 2 ($m = 200.000$ e $n = 250$)	65
5.3.5.4	Algoritmo da Proposta 3 ($m = 200.000$ e $n = 250$)	67
5.3.5.5	Comparação dos Algoritmos ($m = 200.000$ e $n = 250$)	68
5.3.6	Resultados no cenário com $m = 200.000$ e $n = 500$	69
5.3.6.1	Algoritmo de Dobra et al. ($m = 200.000$ e $n = 500$)	69
5.3.6.2	Algoritmo da Proposta 1 ($m = 200.000$ e $n = 500$)	71
5.3.6.3	Algoritmo da Proposta 2 ($m = 200.000$ e $n = 500$)	72
5.3.6.4	Algoritmo da Proposta 3 ($m = 200.000$ e $n = 500$)	73
5.3.6.5	Comparação dos Algoritmos ($m = 200.000$ e $n = 500$)	75
5.3.7	Resultados no cenário com $m = 300.000$ e $n = 100$	76
5.3.7.1	Algoritmo de Dobra et al. ($m = 300.000$ e $n = 100$)	76
5.3.7.2	Algoritmo da Proposta 1 ($m = 300.000$ e $n = 100$)	78
5.3.7.3	Algoritmo da Proposta 2 ($m = 300.000$ e $n = 100$)	79
5.3.7.4	Algoritmo da Proposta 3 ($m = 300.000$ e $n = 100$)	80

5.3.7.5	<i>Comparação dos Algoritmos ($m = 300.000$ e $n = 100$)</i>	81
5.3.8	<i>Resultados no cenário com $m = 300.000$ e $n = 250$</i>	83
5.3.8.1	<i>Algoritmo de Dobra et al. ($m = 300.000$ e $n = 250$)</i>	83
5.3.8.2	<i>Algoritmo da Proposta 1 ($m = 300.000$ e $n = 250$)</i>	84
5.3.8.3	<i>Algoritmo da Proposta 2 ($m = 300.000$ e $n = 250$)</i>	85
5.3.8.4	<i>Algoritmo da Proposta 3 ($m = 300.000$ e $n = 250$)</i>	87
5.3.8.5	<i>Comparação dos Algoritmos ($m = 300.000$ e $n = 250$)</i>	88
5.3.9	<i>Resultados no cenário com $m = 300.000$ e $n = 500$</i>	89
5.3.9.1	<i>Algoritmo de Dobra et al. ($m = 300.000$ e $n = 500$)</i>	89
5.3.9.2	<i>Algoritmo da Proposta 1 ($m = 300.000$ e $n = 500$)</i>	91
5.3.9.3	<i>Algoritmo da Proposta 2 ($m = 300.000$ e $n = 500$)</i>	92
5.3.9.4	<i>Algoritmo da Proposta 3 ($m = 300.000$ e $n = 500$)</i>	93
5.3.9.5	<i>Comparação dos Algoritmos ($m = 300.000$ e $n = 500$)</i>	94
5.4	<i>Comparação dos Critérios</i>	96
6	<i>APLICAÇÃO</i>	97
7	<i>DISCUSSÃO</i>	101
7.1	<i>Discussão dos Resultados</i>	101
7.2	<i>Proposta de Pesquisa Futura</i>	102
	REFERÊNCIAS	103

INTRODUÇÃO

No meio científico, os pesquisadores sempre estão buscando relações entre fenômenos sociais, físicos, biológicos, entre outros. Essas relações ajudam a entender melhor um evento ou fenômeno aleatório. Uma ferramenta para descrever essas relações são os grafos. Eles conseguem representar de forma visual e direta conceitos de independência, independência condicional e dependência entre variáveis aleatórias. Uma forma de aplicar o conceito de grafo em banco de dados é mediante o modelo grafo.

Um tipo de grafo utilizado para representar relações é a rede. Na teoria de grafos uma rede pode ser definida como um grafo que possui características atribuídas aos vértices (por exemplo, idade, gênero, renda). O modelo de redes permite aos pesquisadores representar estruturas relacionais de indivíduos ou vértices (por exemplo, pessoas, animais, organizações, países), onde as relações são definidas por características ou combinações delas. Com isso, temos o modelo de Erdős-Rényi ([ERDOS; RENYI, 1959](#); [ERDŐS; RÉNYI *et al.*, 1960](#)) que é utilizado para estudar grafos aleatórios e que também pode ser aplicado em redes. Segundo [Daudin, Picard e Robin \(2008\)](#), o modelo de Erdős-Rényi é um dos modelos mais antigos e melhor estudado, entretanto, esse modelo não se ajusta bem a dados reais, pois a distribuição empírica do modelo não corresponde bem a suposição deste modelo. Isso acontece devido a suposição de que a distribuição do número de relações de cada vértice é a mesma para todos.

De acordo com [Nowicki e Snijders \(2001\)](#), o modelo de redes tem como foco primário no padrão de relações entre os vértices envolvidos e essas relações são frequentemente afetadas pelas características dos indivíduos. Em situações práticas, estamos interessados em estudar grupos com características diferentes, o que aumenta a complexidade do padrão observado de relacionamentos. A complexidade da situação é ainda aumentada pelo fato das características que influenciam os padrões de relacionamento não serem conhecidas *a priori*, e que essas características podem não ter sido medidas. Essas variáveis são consideradas como não observadas, ou não observáveis, ou latentes, nos modelos que representam o fenômeno. Em [Nowicki e Snijders](#)

(2001), o modelo de estruturas de blocos estocásticos (*Stochastic Blockmodel*) é proposto para corrigir esse problema. Esse modelo assume que os vértices pertencem a classes com diferentes características de conectividade. Entretanto, conforme Daudin, Picard e Robin (2008), esta suposição fornece uma estrutura probabilística adequada, mas o método de estimação proposto não pode lidar com redes compostas com mais de 200 vértices. Outros modelos foram propostos para corrigir essas deficiências (ALBERT; BARABÁSI, 2002).

Um método menos complexo e apresentado por Daudin, Picard e Robin (2008), consiste em propor uma abordagem de modelagem para mistura ao modelo de Erdős-Rényi. Com isso, iremos considerar que o grafo é dividido em classes ou misturas, em que cada vértice dentro da mesma classe possui a mesma distribuição do número de relações e vértices de classes distintas apresentam diferentes distribuições. Com esta mesclagem de metodologias, possibilitamos ao modelo de redes interagir com um maior número de configurações para as relações entre vértices.

Em Haslbeck e Waldorp (2015), é apresentado um método baseado em campos aleatórios de Markov, fatoração do grafo em cliques e família exponencial. Esta abordagem consiste em supor que os vértices pertencem a família exponencial. Através das propriedades markovianas o grafo é fatorado em cliques e o método LASSO (*Least Absolute Shrinkage and Selection Operator*) é utilizado para estimar os parâmetros e selecionar o melhor modelo grafo.

Uma modelagem mais clássica é representada pelo modelo grafo gaussiano. Este modelo consiste, principalmente, que a distribuição conjunta dos vértices segue densidade de uma normal multivariada. Segundo Whittaker (1990, p. 155), a utilização da suposição de normalidade multivariada são devidos aos resultados fechados que a distribuição conjunta apresenta sob marginalização, condicionamento e correlação. Um dos resultados importantes é que a independência na estrutura do grafo é determinada pela alocação de zeros na matriz de covariâncias ou a inversa da matriz de covariâncias.

O modelo grafo gaussiano proposto por Dobra, Lenkoski e Rodriguez (2011) consiste em decompor a inversa da matriz de covariâncias e mediante uma abordagem bayesiana é feita a estimação dos parâmetros e seleção do modelo. De acordo com os autores, este tipo de abordagem usada para inferência do modelo grafo, contribui na criação de estimadores que incorporam a incerteza da estrutura do grafo. A seleção de modelo é feita via *Reversible Jump* e a estimação dos parâmetros pelo método Metropolis–Hastings.

Para acomodar uma modelagem menos presa a suposição de normalidade dos vértices, Dobra e Lenkoski (2011) propoem uma abordagem bayesiana abrangente para determinação de modelos grafos que podem acomodar variáveis binárias, ordinais ou contínuas simultaneamente. O modelo é denominado modelo grafo gaussianos de cópula e incorpora a seleção de modelos grafos dentro de uma cópula gaussiana semiparamétrica.

Outra abordagem em modelos grafos foi proposta por Mohammadi e Wit (2015), que introduziu uma nova e eficiente estrutura bayesiana para determinação de modelos grafos

gaussianos que é uma abordagem transdimensional de Markov Chain Monte Carlo (MCMC) baseada em um processo de nascimento-morte em tempo contínuo.

1.1 Objetivo

O objetivo deste trabalho é estudar e utilizar o modelo grafo gaussiano proposto por [Dobra, Lenkoski e Rodriguez \(2011\)](#), que estima os parâmetros do modelo via o método MCMC (*Markov Chain Monte Carlo*), e propomos modificações no algoritmo RJ (*Reversible Jump*), que é usado para selecionar o melhor modelo que represente a relação entre os vértices. As modificações propostas consistem em melhorar a função geradora de candidato do algoritmo, pois queremos propor com maior frequência candidatos que trazem mais informação para o modelo grafo gaussiano. Além disso, propomos também critérios de avaliação para um estudo de simulação. E comparamos o nosso método com o proposto por [Dobra, Lenkoski e Rodriguez \(2011\)](#) via estudo de simulação.

1.2 Organização do Trabalho

No Capítulo 2, apresenta-se o conceito de modelo grafo e na Seção 2.1, alguns conceitos utilizados sobre teoria dos grafos são definidos. Na Seção 2.2, o modelo grafo é definido, assim como, o modelo grafo gaussiano e o modelo grafo gaussiano proposto por [Dobra, Lenkoski e Rodriguez \(2011\)](#).

No Capítulo 3, defini-se o procedimento de estimação Bayesiana, em que as estimativas são obtidas utilizando o algoritmo Metropolis-Hastings, para o modelo de estudo. Na Seção 3.1, apresenta-se a definição do algoritmo Metropolis-Hastings. Na Seção 3.2, o método de estimação via Metropolis-Hastings é aplicado ao modelo grafo gaussiano de [Dobra, Lenkoski e Rodriguez \(2011\)](#).

No Capítulo 4, a definição do algoritmo RJ é apresentado e aplicado ao modelo grafo gaussiano de [Dobra, Lenkoski e Rodriguez \(2011\)](#) e modificações são propostas na geradora de candidato.

No Capítulo 5, um método de simular dados referentes a um modelo grafo gaussiano é apresentado e critérios de avaliações para um estudo de simulação sob modelo grafo gaussiano é proposto. Por fim, os resultados do estudo de simulação são apresentados.

No Capítulo 6, apresenta-se os resultados da aplicação do método estudado em um banco de dados da literatura.

No Capítulo 7, apresenta-se a discussão dos resultados do Capítulo 6 e a proposta para trabalhos futuros.

MODELO GRAFO GAUSSIANO

Neste capítulo, alguns conceitos gerais sobre teoria dos grafos são definidos, assim como a definição de modelo grafo e as características de um modelo grafo gaussiano. Por fim, o modelo proposto por [Dobra, Lenkoski e Rodriguez \(2011\)](#) é definido e apresentada sua construção teórica é apresentada.

2.1 Conceitos Gerais

Os grafos são estruturas utilizadas para estudar as relações entre os elementos de um determinado conjunto. Um grafo $G = (V, E)$ é composto por dois conjuntos, V é um conjunto não vazio de objetos denominados vértices e E é um subconjunto de pares de vértices que são denominadas de arestas, tal que $E \subset V \times V$. Quando dois vértices são conectados por uma aresta esses vértices são denominados adjacentes. Se as arestas do grafo possuem sentido associado, então G é denominado de grafo direcionado. Caso contrário, o grafo é chamado de não direcionado, ou seja, se $(i, j) \in E$, então $(j, i) \in E$. Um grafo não direcionado também pode ser denominado *Markov Random Fields* ou *Markov networks*. Um grafo não direcionado é denominado completo se todos os vértices do grafo são adjacentes.

Um grafo $G' = (V', E')$ é um subgrafo de G se $V' \subseteq V$ e $E' \subseteq E$. Um subgrafo G' é um subgrafo induzido de G se para qualquer par de vértices $i, j \in V'$, então a aresta $(i, j) \in E'$ se e somente se $(i, j) \in E$. A vizinhança de um vértice i em um grafo G é um subgrafo induzido de G , constituído por todos os vértices adjacentes a i e todas as arestas ligando esses dois vértices.

Um conceito básico na teoria dos grafos é a definição de clique em um grafo não direcionado. C é um clique em G se $C \subseteq V$ tal que $\forall i, j \in C, (i, j) \in E$. Com isso, um subgrafo induzido por C é um grafo completo.

2.2 Modelo Grafo

Conforme [Lauritzen \(1996\)](#), os modelos grafo são famílias de distribuições utilizadas para representar um grafo G através de subconjuntos de V que são condicionalmente independentes. Com isso, cada vértice $v \in V$ do grafo G pode ser associado a uma variável aleatória X_v e cada subconjunto $A \subseteq V$ pode ser representado por um vetor aleatório $X_A = \{X_v : v \in A\}$. As relações de dependência condicional entre as variáveis aleatórias do grafo G são representadas pelo conjunto de arestas E .

2.2.1 Modelo Grafo Gaussiano

O modelo grafo gaussiano consiste, principalmente, que a distribuição conjunta dos vértices segue densidade de uma normal multivariada, ou seja, $X_V \sim N_{|V|}(\mu, \Sigma)$, em que $|V|$ denota o número de elementos do conjunto V . A função densidade de probabilidade da distribuição normal multivariada é dada por

$$f(X|V, \mu, \Sigma) = (2\pi)^{-|V|/2} \det(\Sigma)^{-1/2} \exp \left\{ -\frac{1}{2} (X - \mu)^\top \Sigma^{-1} (X - \mu) \right\}.$$

Segundo [Whittaker \(1990, p. 155\)](#), a utilização da suposição de normalidade multivariada é devida aos resultados fechados que a distribuição conjunta apresenta sob marginalização e condicionamento, o resultado referente a relação direta entre não correlação e independência da densidade normal, e que a independência na estrutura do grafo é determinada pela alocação de zeros na inversa da matriz de covariâncias.

Conforme [Whittaker \(1990, p. 165\)](#), para especificar um modelo grafo gaussiano é necessário um grafo G e um vetor aleatório p dimensional, em que p é o número de vértices no grafo. Com isso, um modelo grafo gaussiano pode ser definido como uma família de distribuições normais restritas às regras de independências condicionais inerentes ao grafo.

2.2.1.1 Modelo Grafo Gaussiano de Dobra, Lenkoski e Rodriguez

O modelo grafo gaussiano proposto por [Dobra, Lenkoski e Rodriguez \(2011\)](#) consiste em decompor a inversa da matriz de covariâncias e mediante uma abordagem bayesiana é feita a estimação dos parâmetros e seleção do modelo. De acordo com os autores, este tipo de abordagem usada para inferência do modelo grafo, contribui na criação de estimadores que incorporam a incerteza da sua estrutura, em que a seleção de modelo é feita via *Reversible Jump* e a estimação dos parâmetros pelo algoritmo Metropolis–Hastings.

Seja $X_V = (X_1, \dots, X_p)$ um vetor aleatório p dimensional com distribuição normal multivariada, $X_V \sim N_p(0, \Sigma)$, em que $V = \{1, \dots, p\}$. Considere $G = (V, E)$ um grafo não direcionado e que \mathcal{G}_p é o conjunto de todos os grafos não direcionados com p vértices, que possui tamanho $2^{p(p-1)/2}$. A estrutura de G no modelo grafo gaussiano é determinada pela matriz de precisão

$K = \Sigma^{-1}$, em que K é restrita ao cone P_G de matrizes simétricas e positivas definidas. Se o elemento fora da diagonal $K_{ij} = 0$, então $(i, j) \notin E$. Quando $(i, j) \notin E$, então as variáveis aleatórias X_i e X_j são condicionalmente independentes dadas as variáveis restantes.

Considere que a matriz de precisão K tem distribuição *a priori* G-Wishart, $K \sim \text{GWis}(\delta, D, G)$, que é dada por

$$f(K|\delta, D, G) = \frac{1}{I_G(\delta, D)} \det(K)^{(\delta-2)/2} \exp \left\{ -\frac{1}{2} \langle K, D \rangle \right\},$$

com respeito a medida de Lebesgue em P_G (ROVERATO, 2002; ATAY-KAYIS; MASSAM, 2005; LETAC; MASSAM, 2007). $I_G(\delta, D)$ é a constante de normalização, se $\delta > 2$ e D é uma matriz positiva definida ($p \times p$), então $I_G(\delta, D) < \infty$ (DIACONIS; YLVISAKER, 1979). O termo $\langle A, B \rangle = \text{tr}(A^\top B)$ é o traço do produto interno. Segundo Muirhead (2009), se o grafo G é completo, então a distribuição G-Wishart, $\text{GWis}(\delta, D, G)$, é reduzida a distribuição Wishart, $\text{Wis}(\delta, D)$. E G é uma matriz binária simétrica ($p \times p$) da estrutura do grafo que K representa, em que se o elemento (i, j) fora da diagonal principal for 1, isso indica a presença da aresta (i, j) , caso contrário, a ausência da aresta.

Uma motivação para a utilização da G-Wishart como distribuição *a priori* de K é que se X_V apresenta distribuição normal multivariada, então a distribuição *a posteriori* de K é também G-Wishart. Seja $\mathcal{D} = \{x_V^{(1)}, \dots, x_V^{(n)}\}$ os dados observados de n amostras independentes tal que $X_V \sim N_p(0, K^{-1})$. Dado que $G \in \mathcal{G}_p$ e $K \in P_G$, se a distribuição *a priori* para K é dada por $\text{GWis}(\delta_0, D_0, G)$, logo a distribuição *a posteriori* para K é dada por $\text{GWis}(\delta, D, G)$, em que $\delta = n + \delta_0$, $D = U + D_0$ e $U = \sum_{i=1}^n x_V^{(i)} (x_V^{(i)})^\top$.

Na metodologia utilizada por Dobra, Lenkoski e Rodriguez (2011), a matriz de precisão K é trabalhada indiretamente, pois essa matriz é decomposta via Cholesky e resultando no produto de duas matrizes triangulares. Pode-se escrever $K \in P_G$ como $K = \Phi^\top \Phi$, em que Φ é uma matriz triangular superior, a qual pode ser decomposta como $\Phi = \Psi Q$, sendo que Q e Ψ são matrizes triangulares superiores e que Q é o resultado de uma decomposição de Cholesky para $D^{-1} = Q^\top Q$. Portanto, K pode ser decomposta em $K = Q^\top \Psi^\top \Psi Q$. Essa decomposição é usada com o intuito de simplificar a manipulação de K , ou seja, ao invés de trabalhar com uma matriz no seu estado natural, os autores trabalham com uma matriz triangular, que apresenta uma estrutura mais simples que a matriz de origem. Além disso, quando um elemento de Ψ é modificado, isso conduz uma zona de influência em cadeia de melhor controle em comparação com Φ e K . Por exemplo, se o elemento Ψ_{ij} é modificado, isso acarreta que todos os elementos da linha i tal que os índices das colunas são maiores que j são também modificados (ATAY-KAYIS; MASSAM, 2005, proposition 2, lemma 2).

De acordo com Dobra, Lenkoski e Rodriguez (2011), as restrições de zero nos elementos fora da diagonal de K associados ao grafo G induzem conjuntos bem definidos de elementos livres $\Phi^{v(G)} = \{\Phi_{ij} : (i, j) \in v(G)\}$ e $\Psi^{v(G)} = \{\Psi_{ij} : (i, j) \in v(G)\}$ das matrizes Φ e Ψ , em que $v(G) = v_=(G) \cup v_<(G)$, $v_=(G) = \{(i, i) : i \in V\}$ e $v_<(G) = \{(i, j) : i < j \text{ e } (i, j) \in E\}$. Veja

Atay-Kayis e Massam (2005, proposition 2, lemma 2) para mais detalhes sobre os elementos livres.

Seja $M^{v(G)}$ o conjunto de matrizes triangulares incompletas cujos elementos são indexados por $v(G)$ e os elementos da diagonal são estritamente positivos. Com isso, $\Phi^{v(G)}, \Psi^{v(G)} \in M^{v(G)}$. Conforme Atay-Kayis e Massam (2005, lemma 2), os elementos não livres de Ψ são determinados por operações de complementação e ficam em função dos elementos livres $\Psi^{v(G)}$.

O intuito das definições apresentadas nesta seção, é de fundamentar transformações de variáveis aleatórias com o objetivo de calcular a função densidade de probabilidade para Ψ . Assim, resta calcular os devidos jacobianos das transformações. Segundo Roverato (2002), o jacobiano da transformação de $K \rightarrow \Phi^{v(G)}$ é dado por

$$J\left(K \rightarrow \Phi^{v(G)}\right) = 2^p \prod_{i=1}^p \Phi_{ii}^{v_i^G+1},$$

em que $v_i^G = |\{j : j > i \text{ e } (i, j) \in E\}|$. De acordo com Atay-Kayis e Massam (2005), o jacobiano da transformação de $\Phi^{v(G)} \rightarrow \Psi^{v(G)}$ é dado por

$$J\left(\Phi^{v(G)} \rightarrow \Psi^{v(G)}\right) = \prod_{i=1}^p Q_{ii}^{d_i^G+1},$$

em que $d_i^G = |\{j : j < i \text{ e } (i, j) \in E\}|$. Conforme Dobra, Lenkoski e Rodriguez (2011), tem-se que $\det(K) = \prod_{i=1}^p \Phi_{ii}^2$ e $\Phi_{ii} = \Psi_{ii} Q_{ii}$. Com isso, a densidade de $\Psi^{v(G)}$ com respeito a medida de Lebesgue em $M^{v(G)}$ é dada por

$$f(\Psi^{v(G)} | \delta, D, G) = \frac{2^p}{I_G(\delta, D)} \prod_{i=1}^p Q_{ii}^{v_i^G+d_i^G+\delta} \prod_{i=1}^p \Psi_{ii}^{v_i^G+\delta-1} \exp\left\{-\frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p \Psi_{ij}^2\right\}.$$

Note que o termo $v_i^G + d_i^G$ representa o número de vizinhos do vértice i no grafo G .

ESTIMAÇÃO VIA METROPOLIS-HASTINGS

Neste capítulo, o método MCMC via Metropolis-Hastings é apresentado e definido. Esse método é aplicado ao modelo proposto por [Dobra, Lenkoski e Rodriguez \(2011\)](#) com o objetivo de estimar os parâmetros do modelo grafo gaussiano.

3.1 Definição

O algoritmo Metropolis-Hastings é um método de Monte Carlo em cadeias de Markov (MCMC) para obter amostras aleatórias a partir de uma distribuição de probabilidade da qual a amostragem direta é difícil. Esse método é geralmente usado para amostragem de distribuições multidimensionais, especialmente quando o número de dimensões é alto.

No algoritmo Metropolis-Hastings um valor é gerado a partir de uma distribuição auxiliar e é aceito com certa probabilidade. Este mecanismo de correção garante a convergência da cadeia para a distribuição de equilíbrio.

Suponha que a cadeia esteja no estado x e um valor x' é gerado de uma distribuição geradora de candidato $f(\cdot|x)$. Note que a geradora de candidato pode depender do estado atual da cadeia. O novo valor x' é aceito com probabilidade,

$$\alpha(x, x') = \min \left\{ 1, \frac{\pi(x')f(x|x')}{\pi(x)f(x'|x)} \right\},$$

em que $\pi(\cdot)$ é a distribuição que deseja-se simular valores. Se considerar que $\pi(\cdot)$ é uma distribuição *a posteriori*, então

$$\alpha(x, x') = \min \left\{ 1, \frac{p(y|x')p(x')f(x|x')}{p(y|x)p(x)f(x'|x)} \right\},$$

em que $\pi(x|y) \propto p(y|x)p(x)$, no qual $p(y|x)$ é a função de verossimilhança de $y|x$ e $p(x)$ é distribuição *a priori* de x .

3.2 Metropolis-Hastings no Modelo Grafo Gaussiano

Segundo [Dobra, Lenkoski e Rodriguez \(2011\)](#), o algoritmo Metropolis–Hastings proposto é uma nova versão que, em contraste ao trabalho de [Mitsakakis, Massam e Escobar \(2011\)](#), utiliza uma distribuição que depende do estado atual da cadeia e deixando inalterado todos os elementos livres em $\Psi^{v(G)}$, com exceção do elemento livre em teste de mudança.

Seja $K^{[s]} = Q^\top (\Psi^{[s]})^\top \Psi^{[s]} Q$ o estado atual da cadeia com $(\Psi^{[s]})^{v(G)} \in M^{v(G)}$. O próximo estado $K^{[s+1]} = Q^\top (\Psi^{[s+1]})^\top \Psi^{[s+1]} Q$ é obtido por uma perturbação sequencial dos elementos livres $(\Psi^{[s]})^{v(G)}$. Os elementos livres fora da diagonal $\Psi_{ij}^{[s]}$ são atualizados amostrando de uma distribuição normal, $\gamma' \sim N(\Psi_{ij}^{[s]}, \sigma_m^2)$. Entretanto, os elementos da diagonal $\Psi_{ii}^{[s]} > 0$ são atualizados amostrando de uma distribuição normal truncada inferiormente em zero, $\gamma \sim \text{TN}(\Psi_{ii}^{[s]}, \sigma_m^2, 0, \infty)$.

Para propor a mudança de um elemento livre da diagonal (i^*, i^*) é necessário considerar Ψ' tal que $\Psi'_{ij} = \Psi_{ij}^{[s]}$ para $(i, j) \in v(G) \setminus \{(i^*, i^*)\}$ e $\Psi'_{i^*i^*} = \gamma$. Assim, a cadeia de Markov se move para $K' = Q^\top (\Psi')^\top \Psi' Q$ com probabilidade de aceitação $\min\{R_m, 1\}$, em que

$$R_m = \frac{\phi(\Psi_{i^*i^*}^{[s]}/\sigma_m)}{\phi(\Psi'_{i^*i^*}/\sigma_m)} \left(\frac{\Psi'_{i^*i^*}}{\Psi_{i^*i^*}^{[s]}} \right)^{v_{i^*}^G + \delta - 1} R'_m.$$

O termo $\phi(\cdot)$ representa a função acumulada da normal padrão e

$$R'_m = \exp \left\{ -\frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p \left[(\Psi'_{ij})^2 - (\Psi_{ij}^{[s]})^2 \right] \right\}.$$

Entretanto, para propor a mudança de um elemento livre fora da diagonal (i^*, j^*) é necessário considerar Ψ' tal que $\Psi'_{ij} = \Psi_{ij}^{[s]}$ para $(i, j) \in v(G) \setminus \{(i^*, j^*)\}$ e $\Psi'_{i^*j^*} = \gamma'$. Assim, a cadeia de Markov se move para $K' = Q^\top (\Psi')^\top \Psi' Q$ com probabilidade de aceitação igual a $\min\{R'_m, 1\}$.

REVERSIBLE JUMP

Neste capítulo, o método *Reversible Jump* é apresentado e definido. Esse método é aplicado ao modelo proposto por [Dobra, Lenkoski e Rodriguez \(2011\)](#) com o objetivo de buscar a estrutura do modelo grafo gaussiano.

4.1 Definição

Segundo [Green \(1995\)](#), o *Reversible Jump* é uma extensão do algoritmo Metropolis-Hastings que dá, à cadeia de Markov do método MCMC, a possibilidade de saltar entre espaços paramétricos, o que significa de um modelo para o outro, mesmo que os espaços tenham dimensões diferentes, segundo um critério bem estabelecido, o de Metropolis-Hastings por exemplo, este garante a reversibilidade do processo, e conseqüentemente a existência da distribuição de equilíbrio.

Suponha que o estado atual da cadeia é (k, θ) , isto é, estamos no modelo k com parâmetros θ . Um novo modelo k' com parâmetros θ' é proposto com probabilidade $r_{k,k'}$. O ato de saltar entre dimensões no *Reversible Jump* constitui em incluir ou retirar parâmetros do modelo atual.

Assuma inicialmente que o modelo proposto tem dimensão maior, isto é, $n_{k'} > n_k$ e que $\theta' = g(\theta, u)$ para uma função determinística g e um vetor aleatório $u \sim q(u)$ com dimensão $n_{k'} - n_k$. Então, aceita-se (k', θ') com probabilidade $\alpha(k, k') = \min\{1, A\}$, em que

$$A = \frac{\pi(k', \theta')}{\pi(k, \theta)} \frac{r_{k',k}}{r_{k,k'} q(u)} \left| \frac{\partial g(\theta, u)}{\partial(\theta, u)} \right|.$$

Caso contrário, se o modelo proposto tenha dimensão menor, então aceita-se (k', θ') com probabilidade $\alpha(k, k') = \min\{1, A^{-1}\}$.

4.2 Reversible Jump no Modelo Grafo Gaussiano

O algoritmo *Reversible Jump* é utilizado para propor uma atualização na estrutura do grafo G , pois a ação de incluir ou retirar uma aresta de G é equivalente a modificar o espaço paramétrico do modelo atual para outro com dimensão diferente.

Sejam $(K^{[s]}, G^{[s]})$ e $(K^{[s+1]}, G^{[s+1]})$ os estados s e $s+1$ da cadeia, respectivamente, em que $K^{[s]} \in P_{G^{[s]}}$ e $K^{[s+1]} \in P_{G^{[s+1]}}$. Para prosseguir a $(K^{[s+1]}, G^{[s+1]})$ é necessário um processo de duas etapas. A primeira etapa consiste em selecionar um candidato G' para $G^{[s+1]}$ e a segunda etapa é atualizar a $K^{[s]}$ para $K^{[s+1]}$.

Sejam $\text{nbd}_p^+(G)$ o conjunto dos grafos que podem ser obtidos com a inclusão de uma aresta no grafo $G \in \mathcal{G}_p$ e $\text{nbd}_p^-(G)$ o conjunto dos grafos que podem ser obtidos com a retirada de uma aresta no grafo $G \in \mathcal{G}_p$. O conjunto de grafos $\text{nbd}_p(G) = \text{nbd}_p^+(G) \cup \text{nbd}_p^-(G)$ é denotado como vizinhança de G em \mathcal{G}_p . A interpretação de $\text{nbd}_p(G)$ é semelhante a definição de vizinhança na teoria dos grafos. Um grafo G_1 é vizinho de G_2 se a operação de inclusão ou retirada de alguma aresta em G_1 resultar em G_2 .

No trabalho de [Dobra, Lenkoski e Rodriguez \(2011\)](#), um candidato $G' \in \text{nbd}_p(G^{[s]})$ é amostrado da seguinte distribuição

$$f_1(G'|G^{[s]}) = \frac{1}{2}\text{Uni}\left(\text{nbd}_p^+(G^{[s]})\right) + \frac{1}{2}\text{Uni}\left(\text{nbd}_p^-(G^{[s]})\right),$$

em que $\text{Uni}(A)$ representa a distribuição uniforme discreta do conjunto A . Os autores utilizam f_1 como gerador de candidato, pois esta abordagem fornece probabilidade igual de G' pertencer a $\text{nbd}_p^+(G^{[s]})$ ou $\text{nbd}_p^-(G^{[s]})$. Nós propomos um outro gerador de candidato que é dado pela seguinte distribuição

$$f_2(G'|G^{[s]}) = \text{Uni}\left(\text{nbd}_p(G^{[s]})\right).$$

Na proposta f_2 , o objetivo é fornecer probabilidade igual para todos os possíveis candidatos a partir do estado atual $G^{[s]}$. Para as próximas propostas de gerador de candidato é necessário definir alguns conceitos e considerar dois casos: G' pertence a $\text{nbd}_p^-(G^{[s]})$ ou $\text{nbd}_p^+(G^{[s]})$.

O primeiro conceito necessário para construir o próximo gerador de candidato é da correlação parcial. Este tipo de correlação consiste em medir a associação entre duas variáveis aleatórias condicionada por um conjunto de variáveis aleatórias. O objetivo de condicionar a correlação é retirar qualquer possível efeito que o conjunto de variáveis aleatórias tenha sob a medição de associação entre as duas variáveis aleatórias de interesse. Com isso, a correlação parcial pode ser denotada por $\rho(X, Y|Z)$, que significa a correlação entre X e Y condicionado por Z e pode ser aplicado para representar arestas de um grafo.

O próximo passo é escolher uma forma de calcular a correlação parcial, pois existem vários modos de calcular este tipo de correlação e a forma de cálculo tem que se encaixar nas suposições e estruturas que usamos no trabalho. De acordo com [Whittaker \(1990, corollary](#)

5.8.2), a correlação parcial pode ser calculada a partir da inversa da matriz de covariâncias de X , que é a matriz K ,

$$\rho(X_i, X_j | X \setminus \{X_i, X_j\}) = \frac{-K_{ij}}{\sqrt{K_{ii}K_{jj}}}.$$

No caso de $G' \in \text{nb}_p^-(G^{[s]})$, deve-se calcular a correlação parcial sob $G^{[s]}$ para cada aresta (i^*, j^*) , $i^* < j^*$, que foi retirada de $G^{[s]}$ para construir o conjunto $\text{nb}_p^-(G^{[s]})$. Se $G' \in \text{nb}_p^+(G^{[s]})$, então deve-se calcular a correlação parcial sob G' para cada aresta (i^*, j^*) , $i^* < j^*$, que foi incluída em $G^{[s]}$ para construir o conjunto $\text{nb}_p^+(G^{[s]})$.

Para ambos os casos, $G' \in \text{nb}_p^-(G^{[s]})$ ou $G' \in \text{nb}_p^+(G^{[s]})$, todos os elementos da diagonal de K' são iguais aos elementos da diagonal de $K^{[s]}$. Se $(i, j) \in E^{[s]}$ ou $(i, j) \notin E^{[s]}$, tem-se que $K'_{ij} = K^{[s]}_{ij}$. Entretanto, no caso de $G' \in \text{nb}_p^+(G^{[s]})$, tem-se que $K'_{i^*j^*}$ é desconhecido.

Como o valor de $K'_{i^*j^*}$ é desconhecido para todo $G' \in \text{nb}_p^+(G^{[s]})$, tem-se que propor uma quantidade baseada nas informações dos dados e do estado atual da cadeia. A informação extraída diretamente dos dados é a inversa da matriz de covariâncias inicial $K^{[0]}$ e a informação do estado atual é $K^{[s]}$. Para cada $G' \in \text{nb}_p^+(G^{[s]})$ deve-se modificar através de operações de complementação (ATAY-KAYIS; MASSAM, 2005) a estrutura de $K^{[0]}$ para um K^* , em que K^* representa uma nova estrutura do grafo que seja igual a G' . A proposta é comparar as matrizes K^* e $K^{[s]}$ e solucionar a seguinte equação do primeiro grau

$$\frac{-K^*_{i^*j^*}}{\sqrt{K^*_{i^*i^*}K^*_{j^*j^*}}} = \frac{-K'_{i^*j^*}}{\sqrt{K^{[s]}_{i^*i^*}K^{[s]}_{j^*j^*}}},$$

com solução dada por

$$K'_{i^*j^*} = \frac{K^*_{i^*j^*} \sqrt{K^{[s]}_{i^*i^*}K^{[s]}_{j^*j^*}}}{\sqrt{K^*_{i^*i^*}K^*_{j^*j^*}}}.$$

Como é possível obter todos os termos necessários para calcular a correlação parcial correspondente a aresta $(i^*, j^*) \in E'$ de cada $G' \in \text{nb}_p^+(G^{[s]})$ a próxima proposta de gerador de candidato é dada pela distribuição

$$f_3(G' | G^{[s]}) = \frac{1}{2\text{pcs}_+^*} \left| \frac{-K'_{i^*j^*}}{\sqrt{K^{[s]}_{i^*i^*}K^{[s]}_{j^*j^*}}} \right| \mathbb{I}(G') \Big|_{\text{nb}_p^+(G^{[s]})} + \frac{1}{2\text{pcs}_-^*} \left[1 - \left| \frac{-K^{[s]}_{i^*j^*}}{\sqrt{K^{[s]}_{i^*i^*}K^{[s]}_{j^*j^*}}} \right| \right] \mathbb{I}(G') \Big|_{\text{nb}_p^-(G^{[s]})},$$

em que

$$\text{pcs}_+^* = \sum_{G' \in \text{nb}_p^+(G^{[s]})} \left| \frac{-K'_{i^*j^*}}{\sqrt{K^{[s]}_{i^*i^*}K^{[s]}_{j^*j^*}}} \right|$$

e

$$\text{pcs}_-^* = \sum_{G' \in \text{nb}_p^-(G^{[s]})} \left[1 - \left| \frac{-K^{[s]}_{i^*j^*}}{\sqrt{K^{[s]}_{i^*i^*}K^{[s]}_{j^*j^*}}} \right| \right].$$

Nesta proposta, a geradora de candidato f_3 fornece probabilidade igual de G' pertencer a $\text{nb}_p^+(G^{[s]})$ ou $\text{nb}_p^-(G^{[s]})$, entretando a distribuição dentro de cada conjunto não é uniforme. Em $\text{nb}_p^+(G^{[s]})$, quanto maior o valor absoluto da correlação parcial da aresta (i^*, j^*) ao ser incluída no grafo, maior é a probabilidade de G' que contém (i^*, j^*) . Todavia, em $\text{nb}_p^-(G^{[s]})$, quanto menor o valor absoluto da correlação parcial da aresta (i^*, j^*) ao ser retirada do grafo, maior é a probabilidade de G' que retirou (i^*, j^*) .

Outra proposta de gerador de candidato é dada pela seguinte distribuição

$$f_4(G'|G^{[s]}) = \frac{1}{\text{pcs}^*} \left| \frac{-K'_{i^*j^*}}{\sqrt{K_{i^*i^*}^{[s]}K_{j^*j^*}^{[s]}}} \right| \frac{\mathbb{I}(G')}{\text{nb}_p^+(G^{[s]})} + \frac{1}{\text{pcs}^*} \left[1 - \left| \frac{-K_{i^*j^*}^{[s]}}{\sqrt{K_{i^*i^*}^{[s]}K_{j^*j^*}^{[s]}}} \right| \right] \frac{\mathbb{I}(G')}{\text{nb}_p^-(G^{[s]})},$$

em que

$$\text{pcs}^* = \sum_{G' \in \text{nb}_p(G^{[s]})} \left| \frac{-K'_{i^*j^*}}{\sqrt{K_{i^*i^*}^{[s]}K_{j^*j^*}^{[s]}}} \right| \frac{\mathbb{I}(G')}{\text{nb}_p^+(G^{[s]})} + \left[1 - \left| \frac{-K_{i^*j^*}^{[s]}}{\sqrt{K_{i^*i^*}^{[s]}K_{j^*j^*}^{[s]}}} \right| \right] \frac{\mathbb{I}(G')}{\text{nb}_p^-(G^{[s]})}.$$

Nesta última proposta, a geradora de candidato f_4 considera que $G' \in \text{nb}_p(G^{[s]})$, em que a distribuição dos elementos desse conjunto não é uniforme e a definição das probabilidades para cada G' segue a mesma construção que f_3 . A diferença entre estas duas propostas de geradora de candidato é que f_4 considera todas as alternativas de salto no estado atual da cadeia e tendo sempre a possibilidade de fazer a melhor escolha de salto dimensional.

Se o candidato amostrado $G' \in \text{nb}_p^+(G^{[s]})$, então deve-se propor incluir a aresta (i^*, j^*) , $i^* < j^*$, em $G^{[s]}$. Como vértice i^* tem somente um vizinho adicionado em G' , portanto $d_{i^*}^{G'} = d_{i^*}^{G^{[s]}}$, $d_{j^*}^{G'} = d_{j^*}^{G^{[s]}} + 1$, $v_{i^*}^{G'} = v_{i^*}^{G^{[s]}} + 1$, $v_{j^*}^{G'} = v_{j^*}^{G^{[s]}}$ e $v(G') = v(G^{[s]}) \cup \{(i^*, j^*)\}$. Considere a decomposição do estado atual de $K^{[s]} = Q^\top (\Psi^{[s]})^\top \Psi^{[s]} Q$ com $(\Psi^{[s]})^{v(G^{[s]})} \in M^{v(G^{[s]})}$. Uma matriz triangular superior Ψ' é definida tal que $\Psi'_{ij} = \Psi_{ij}^{[s]}$ para $(i, j) \in v(G^{[s]})$. Para os casos f_1 e f_2 , amostramos $\gamma'' \sim N(\Psi_{i^*j^*}^{[s]}, \sigma_g^2)$ e estabelecemos $\Psi'_{i^*j^*} = \gamma''$. No caso f_3 e f_4 , pegamos o valor de $K'_{i^*j^*}$ e através de operações de complementação obtemos $\Psi'_{i^*j^*}$. Os outros elementos de Ψ' são determinados por $(\Psi')^{v(G')}$ através de operações de complementação. Com isso, tem-se que $K' = Q^\top (\Psi')^\top \Psi' Q$ com $(\Psi')^{v(G')} \in M^{v(G')}$ e um candidato para o estado $s + 1$ da cadeia é (K', G') , em que $K' \in P_{G'}$.

Com a definição do candidato (K', G') para o estado $s + 1$ da cadeia, o próximo passo é estabelecer a probabilidade de aceitação do salto dimensional. Se ocorrer o movimento de $(K^{[s]}, G^{[s]})$ para (K', G') , a dimensão do espaço paramétrico é aumentado em 1. Desde que $(\Psi')^{v(G^{[s]})} = (\Psi^{[s]})^{v(G^{[s]})}$, o jacobiano da transformação de $\left((\Psi^{[s]})^{v(G^{[s]})}, \gamma'' \right)$ para $(\Psi')^{v(G')}$ é igual a 1. Além disso, $\Psi^{[s]}$ e Ψ' possuem os mesmos elementos na diagonal, consequentemente $\det(K') = \det(K^{[s]}) = \prod_{i=1}^p (Q_{ii} \Psi_{ii}^{[s]})^2$. Pela definição de [Green \(1995\)](#), a probabilidade da cadeia

de Markov mover para (K', G') é $\min\{R_g^+, 1\}$, em que R_g^+ é dada por

$$R_g^+ = \sigma_g \sqrt{2\pi} Q_{i^*i^*} Q_{j^*j^*} \Psi_{i^*i^*}^{[s]} \frac{I_{G^{[s]}}(\delta_0, D_0)}{I_{G'}(\delta_0, D_0)} \frac{P(G')}{P(G^{[s]})} \frac{f_l(G^{[s]}|G')}{f_l(G'|G^{[s]})} \\ \times \exp \left\{ -\frac{1}{2} \left[\langle K' - K^{[s]}, U + D_0 \rangle - \frac{(\Psi'_{i^*j^*} - \Psi_{i^*j^*}^{[s]})^2}{\sigma_g^2} \right] \right\},$$

em que f_l representa a geradora de candidato escolhida tal que $l \in \{1, 2, 3, 4\}$.

Se o candidato amostrado $G' \in \text{nbd}_p^-(G^{[s]})$, então deve-se propor retirar a aresta (i^*, j^*) , $i^* < j^*$, em $G^{[s]}$. Como vértice i^* tem somente um vizinho removido em G' , portanto $d_{i^*}^{G'} = d_{i^*}^{G^{[s]}}$, $d_{j^*}^{G'} = d_{j^*}^{G^{[s]}} - 1$, $v_{i^*}^{G'} = v_{i^*}^{G^{[s]}} - 1$, $v_{j^*}^{G'} = v_{j^*}^{G^{[s]}}$ e $v(G') = v(G^{[s]}) \setminus \{(i^*, j^*)\}$. Considere a decomposição do estado atual de $K^{[s]} = Q^\top (\Psi^{[s]})^\top \Psi^{[s]} Q$ com $(\Psi^{[s]})^{v(G^{[s]})} \in M^{v(G^{[s]})}$. Uma matriz triangular superior Ψ' é definida tal que $\Psi'_{ij} = \Psi_{ij}^{[s]}$ para $(i, j) \in v(G')$. O elemento livre $\Psi'_{i^*j^*}$ torna-se um elemento não livre em Ψ' . Os outros elementos de Ψ' são determinados por operações de complementação. Com isso, tem-se que $K' = Q^\top (\Psi')^\top \Psi' Q$ com $(\Psi')^{v(G')} \in M^{v(G')}$ e um candidato para o estado $s + 1$ da cadeia é (K', G') , em que $K' \in P_{G'}$.

Com a definição do candidato (K', G') para o estado $s + 1$ da cadeia, o próximo passo é estabelecer a probabilidade de aceitação do salto dimensional. Se ocorrer o movimento de $(K^{[s]}, G^{[s]})$ para (K', G') , a dimensão do espaço paramétrico é diminuída em 1. Por [Green \(1995\)](#), a probabilidade da cadeia de Markov mover para (K', G') é $\min\{R_g^-, 1\}$, em que R_g^- é dada por

$$R_g^- = \left(\sigma_g \sqrt{2\pi} Q_{i^*i^*} Q_{j^*j^*} \Psi_{i^*i^*}^{[s]} \right)^{-1} \frac{I_{G^{[s]}}(\delta_0, D_0)}{I_{G'}(\delta_0, D_0)} \frac{P(G')}{P(G^{[s]})} \frac{f_l(G^{[s]}|G')}{f_l(G'|G^{[s]})} \\ \times \exp \left\{ -\frac{1}{2} \left[\langle K' - K^{[s]}, U + D_0 \rangle + \frac{(\Psi'_{i^*j^*} - \Psi_{i^*j^*}^{[s]})^2}{\sigma_g^2} \right] \right\},$$

em que f_l representa a geradora de candidato escolhida tal que $l \in \{1, 2, 3, 4\}$.

SIMULAÇÃO

Neste capítulo, o método de simulação de dados provenientes de um modelo grafo gaussiano e critérios para avaliar e comparar resultados de uma simulação são apresentados. Os resultados dos cenários simulados são relatados e interpretados.

5.1 Simulação de Dados

Um dos aspectos importantes do estudo de simulação é a capacidade de simular dados com características pré especificadas. No nosso caso, o interesse é simular dados de um grafo não direcionado G com estrutura conhecida e que a distribuição conjunta dos vértices seja uma normal multivariada. Um método para simular G é proposto por [Wang e Li \(2012\)](#). Os autores propõem, primeiramente, construir a estrutura de G através da matriz K , em que $K_{ii} = 1 \forall i \in V$, $K_{ij} = K_{ji} = 0$ se $(i, j) \notin E$ e $K_{ij} = K_{ji} = -\rho(X_i, X_j | X \setminus \{X_i, X_j\})$ se $(i, j) \in E$. A seguir, tem-se um exemplo de K quando o grafo possui três vértices, $V = \{1, 2, 3\}$, e duas arestas, $E = \{(1, 2), (2, 3)\}$, em que $\rho(1, 2|3) = 0,5$ e $\rho(2, 3|1) = -0,4$,

$$K = \begin{bmatrix} K_{11} & K_{12} & K_{13} \\ K_{12} & K_{22} & K_{23} \\ K_{13} & K_{23} & K_{33} \end{bmatrix} = \begin{bmatrix} 1 & -0,5 & 0 \\ -0,5 & 1 & 0,4 \\ 0 & 0,4 & 1 \end{bmatrix}.$$

Após definir a estrutura de G , tem-se que definir a matriz $\Sigma = cK^{-1}$, em que $c \in (0, \infty)$ é uma constante que controla a variabilidade da amostra gerada. Quanto maior o valor de c maior será a variabilidade dos dados simulados e quanto menor o valor de c menor será a variabilidade dos dados simulados. Para a continuação do exemplo acima, considere que $c = 100$ e, portanto, Σ é dada por

$$\Sigma = 100 \times \begin{bmatrix} 1,42 & 0,85 & -0,34 \\ 0,85 & 1,69 & -0,68 \\ -0,34 & -0,68 & 1,27 \end{bmatrix} = \begin{bmatrix} 142 & 85 & -34 \\ 85 & 169 & -68 \\ -34 & -68 & 127 \end{bmatrix}.$$

Depois de definir K e Σ , a próxima etapa é simular os dados com tamanho amostral n . No nosso caso, o interesse é uma distribuição normal multivariada com $N_{|V|}(0, \Sigma)$, em que 0 é o vetor de médias zeros e Σ é a matriz de covariâncias. Para simular os dados, primeiramente, é necessário fixar uma semente, para que os dados sejam replicáveis, e, por fim, deve-se utilizar algum método de simulação da normal multivariada para simular n realizações de $N_{|V|}(0, \Sigma)$.

5.2 Critério de Avaliação e Comparação

Outro aspecto importante do estudo de simulação é a capacidade de avaliar e comparar os diversos cenários simulados. O critério pode ser de dois tipos: quantitativo ou gráfico. O tipo quantitativo consiste em construir um indicador quantitativa que avalia e compara sob cada resultado do método de estudo nos cenários simulados. O tipo gráfico corresponde em construir uma ferramenta gráfica com o mesmo objetivo.

5.2.1 Critério Quantitativo

Nesta subseção, nós propomos seis tipos de medidas quantitativas para serem usadas como critério de avaliação e comparação dos resultados das simulações.

5.2.1.1 Contagem de Permanência das Arestas

A contagem de permanência da aresta (CPA_{ij}) consiste em fazer a contabilidade do número de vezes que a aresta (i, j) estava presente no modelo escolhido ao final de cada interação do algoritmo *Reversible Jump*. CPA_{ij} é uma medida discreta que pertence ao conjunto $\{0, 1, 2, \dots, m\}$, em que m é o número total de interações computadas até o final do algoritmo *Reversible Jump*. Se $CPA_{ij} = m$, significa que todos os modelos visitados continham a aresta (i, j) . Entretanto, se $CPA_{ij} = 0$, significa que todos os modelos visitados não continham a aresta (i, j) .

5.2.1.2 Probabilidade a Posteriori de Permanência das Arestas

A probabilidade *a posteriori* de permanência da aresta ($PPPA_{ij}$) consiste em calcular uma probabilidade sob a contabilidade do número de vezes que a aresta (i, j) estava presente no modelo escolhido ao final de cada interação do algoritmo *Reversible Jump*. Assim, tem-se que

$$PPPA_{ij} = \frac{CPA_{ij}}{m},$$

em que $PPPA_{ij} \in [0, 1]$.

5.2.1.3 Índice de Erro

Como no estudo de simulação temos o conhecimento da estrutura do grafo $G = (V, E)$ que foi utilizado para gerar os dados simulados do modelo grafo gaussiano. Então, nós podemos

calcular para cada aresta (i, j) o quanto erramos ao captar a presença ou ausência dessa aresta no modelo. Com isso, podemos definir o índice de erro da aresta (i, j) como

$$\text{IEA}_{ij} = \begin{cases} 1 - \text{PPPA}_{ij}, & \text{se } (i, j) \in E \\ -\text{PPPA}_{ij}, & \text{se } (i, j) \notin E \end{cases}.$$

Note que $\text{IEA}_{ij} \in (-1, 1)$. Se IEA_{ij} é igual ou próximo do valor 0, significa que não houve erro do algoritmo captar a presença ou ausência da aresta (i, j) . Se IEA_{ij} é igual ou próximo do valor 1, significa que houve erro do algoritmo captar a presença da aresta (i, j) . Se IEA_{ij} é igual ou próximo do valor -1, significa que houve erro do algoritmo captar a ausência da aresta (i, j) .

5.2.1.4 Soma Absoluta dos Erros

A soma absoluta dos erros (SAE) consiste em condensar todas as informações das arestas referentes aos seus índices de erro em uma única medida. Assim, tem-se que

$$\text{SAE} = \sum_{(i,j):i<j} |\text{IEA}_{ij}|.$$

Note que $\text{SAE} \in [0, \binom{|V|}{2}]$. Se $\text{SAE} = 0$, significa que o algoritmo acertou completamente a estrutura do grafo. Todavia, se $\text{SAE} = \binom{|V|}{2}$, significa que o algoritmo errou completamente a estrutura do grafo.

Outra alternativa, a soma absoluta dos erros no intervalo unitário (SAEIU) consiste em uma transformação de SAE em uma escala fixa. Assim, tem-se que

$$\text{SAEIU} = \left(\binom{|V|}{2} \right)^{-1} \text{SAE}.$$

Note que $\text{SAEIU} \in [0, 1]$. Se $\text{SAEIU} = 0$, significa que o algoritmo acertou completamente a estrutura do grafo. Todavia, se $\text{SAEIU} = 1$, significa que o algoritmo errou completamente a estrutura do grafo.

5.2.1.5 Soma Quadrática dos Erros

A soma quadrática dos erros (SQE) consiste na mesma ideia lógica de desenvolvimento da medida SAE. Assim, tem-se que

$$\text{SQE} = \sum_{(i,j):i<j} (\text{IEA}_{ij})^2.$$

Note que $\text{SQE} \in [0, \binom{|V|}{2}]$. Se $\text{SQE} = 0$, significa que o algoritmo acertou completamente a estrutura do grafo. Todavia, se $\text{SQE} = \binom{|V|}{2}$, significa que o algoritmo errou completamente a estrutura do grafo.

Outra alternativa, a soma quadrática dos erros no intervalo unitário (SQEIU) consiste em uma transformação de SQE em uma escala fixa. Assim, tem-se que

$$\text{SQEIU} = \binom{|V|}{2}^{-1} \text{SQE}.$$

Note que $\text{SQEIU} \in [0, 1]$. Se $\text{SQEIU} = 0$, significa que o algoritmo acertou completamente a estrutura do grafo. Todavia, se $\text{SQEIU} = 1$, significa que o algoritmo errou completamente a estrutura do grafo.

5.2.1.6 Índices de Semelhança

O índice de semelhança (IS) consiste em calcular o grau de semelhança entre dois modelos, em que um dos modelos é considerado como referência. Sejam $G' = (V, E')$ e $G'' = (V, E'')$ dois grafos com estruturas diferentes. Assuma que G' é a referência principal. Com isso, tem-se que

$$\text{IS}(G''|G') = \sum_{(i,j) \in E''} \mathbb{I}[(i,j) \in (E'' \cap E')] + \mathbb{I}[(i,j) \in (E'' \cup E')^c].$$

Note que $\text{IS} \in \left[0, \binom{|V|}{2}\right]$. Se $\text{IS} = 0$, significa que G'' não tem semelhança com G' . Entretanto, se $\text{IS} = \binom{|V|}{2}$, significa que G'' é igual a G' .

Outra alternativa, o índice de semelhança no intervalo unitário (ISIU) consiste em uma transformação de IS em uma escala fixa. Assim, tem-se que

$$\text{ISIU}(G''|G') = \binom{|V|}{2}^{-1} \text{IS}(G''|G').$$

Note que $\text{ISIU} \in [0, 1]$. Se $\text{ISIU} = 0$, significa que G'' não tem semelhança com G' . Entretanto, se $\text{ISIU} = 1$, significa que G'' é igual a G' .

5.2.2 Critério Gráfico

Nesta seção, nós propormos dois tipos de gráficos para serem utilizadas como critério de avaliação e comparação dos resultados das simulações. Na seção 5.2.2.1, o critério apresentado também pode ser usado em dados não simulados, pois seu intuito é avaliar a convergência do modelo.

5.2.2.1 Gráfico de Linhas para os Determinantes

Quando se trata de modelos grafos, quando o número de vértices no grafo cresce, a quantidade de parâmetros também cresce rapidamente. Isso torna difícil de avaliar a convergência do modelo, pois isso devemos analisar conjuntamente a convergência de todos os parâmetros. Com isso, a necessidade de resumir todas as informações em um único indicador, se torna uma

alternativa simples e viável no contexto de grafo. Assim, o objetivo é utilizar o determinante como uma medida de resumo para todos os elementos de K . O gráfico de linhas para os determinantes (GLD) é usado para observar a convergência do modelo grafo gaussiano para dados simulados e reais. Se for observada alguma tendência no GLD, pode-se supor que o modelo não chegou a uma convergência, que tem como consequência descredibilizar as estimativas dos parâmetros e o resultado do algoritmo *Reversible Jump*. Entretanto, caso não seja observada uma tendência no GLD e exista um valor constante no qual as variações fiquem entorno, pode ser um indicativo gráfico da convergência do modelo.

5.2.2.2 Gráfico de Avaliação das Probabilidades a Posteriori

No contexto do estudo de simulações, observa-se que na literatura existe a necessidade de um critério gráfico que possa avaliar o desempenho do algoritmo de encontrar a verdadeira estrutura do grafo que foi usada para simular os dados. Com isso, nós propomos a construção de um gráfico que possa avaliar o comportamento dos modelos mais visitados e comparar com o número de visitas do verdadeiro modelo.

Primeiramente, o comportamento dos modelos mais visitados são observados por um gráfico de linha contínua, que representa a ordenação decrescente das probabilidades *a posteriori* dos modelos. Esta linha por construção tem comportamento decrescente, em que quanto menos brusco for o decaimento da linha, maior é o número de modelos visitados pelo algoritmo *Reversible Jump*. Analogamente, quanto mais brusco for o decaimento da linha, menor é o número de modelos visitados pelo algoritmo.

O ranque da probabilidade *a posteriori* do verdadeiro modelo é observado por uma reta horizontal, que corta a linha contínua decrescente na altura em que o verdadeiro modelo se posiciona. Quanto mais elevado se dispor a reta horizontal, maior é o número de visitas que o verdadeiro modelo recebeu durante o algoritmo *Reversible Jump*.

O cenário ideal que podemos observar, é quando a reta horizontal situa-se na maior elevação (indicando que o verdadeiro modelo possui a maior probabilidade *a posteriori*) e que a linha contínua decrescente tenha um decaimento brusco (indicando que o algoritmo visitou poucos modelos).

5.3 Resultados das Simulações

Nesta seção, os resultados das simulações são apresentados e interpretados. Primeiramente, observou-se que os algoritmos ainda não conseguiram alcançar a convergência, sendo necessário um maior número de iterações. Com isso, o objetivo da seção é analisar o comportamento dos algoritmos ao aumentar o número de iterações. As simulações foram processadas nos seguintes cenários: $m \in \{100.000, 200.000, 300.000\}$ e $n \in \{100, 250, 500\}$, em que m é o número total de iterações computadas até o final do algoritmo *Reversible Jump* e n é o tamanho

amostral simulado. Entre as iterações do *Reversible Jump* foram feitas 5 iterações do Metropolis-Hastings (upgrade = 5) e 500 iterações do Metropolis-Hastings antes da primeira iteração do RJ (warm-up = 100, pois warm-up \times upgrade = 500). Além disso, quatro variações do algoritmo *Reversible Jump* para o modelo grafo gaussiano foram avaliadas nos cenários mencionados. O primeiro algoritmo é referente aos autores [Dobra, Lenkoski e Rodriguez \(2011\)](#), em que a distribuição do gerador de candidato adotado é f_1 . Os demais algoritmos são consideradas nossas propostas 1, 2 e 3, nas quais foram adotadas as seguintes distribuições f_2 , f_3 e f_4 para o gerador de candidato, respectivamente. Por último, os hiperparâmetros $\sigma_m^2 = \sigma_g^2 = 10$, a semente 12061991 e o modelo 544 foi usado como base para simular os dados, que possui a estrutura apresentada na Figura 1 e Tabela 1.

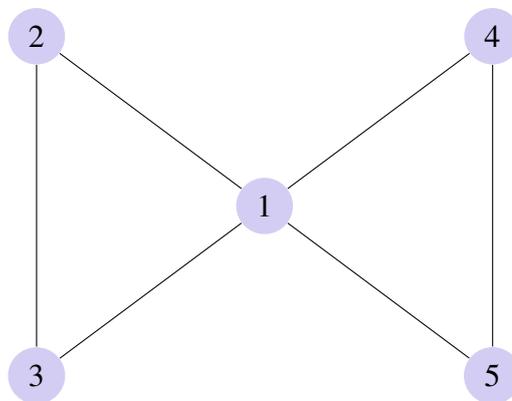


Figura 1 – Estrutura do grafo referente ao modelo 544 utilizado para simular os dados. Fonte: Autor.

Tabela 1 – Correlações parciais entre os vértices.

Vértice	2	3	4	5
1	-0,5	-0,5	-0,5	-0,5
2	-	-0,5	0,0	0,0
3	-	-	0,0	0,0
4	-	-	-	-0,5

Fonte: Autor.

5.3.1 Resultados no cenário com $m = 100.000$ e $n = 100$

Nesta subseção, os resultados das simulações para o cenário $m = 100.000$ e $n = 100$ são apresentados e interpretados para as quatro variações do algoritmo *Reversible Jump*.

5.3.1.1 Algoritmo de Dobra et al. ($m = 100.000$ e $n = 100$)

Neste cenário, o modelo 542 foi o mais visitado, obtendo um total de 12.438 visitas. Entretanto, o modelo 544 obteve, somente, um total de 147 visitas. O $IS(542|544) = 9$ e $ISIU(542|544) = 0,90$. Com isso, tem-se que o modelo 542 possui uma semelhança de 90% com o modelo 544 que pode ser observada na Figura 2.

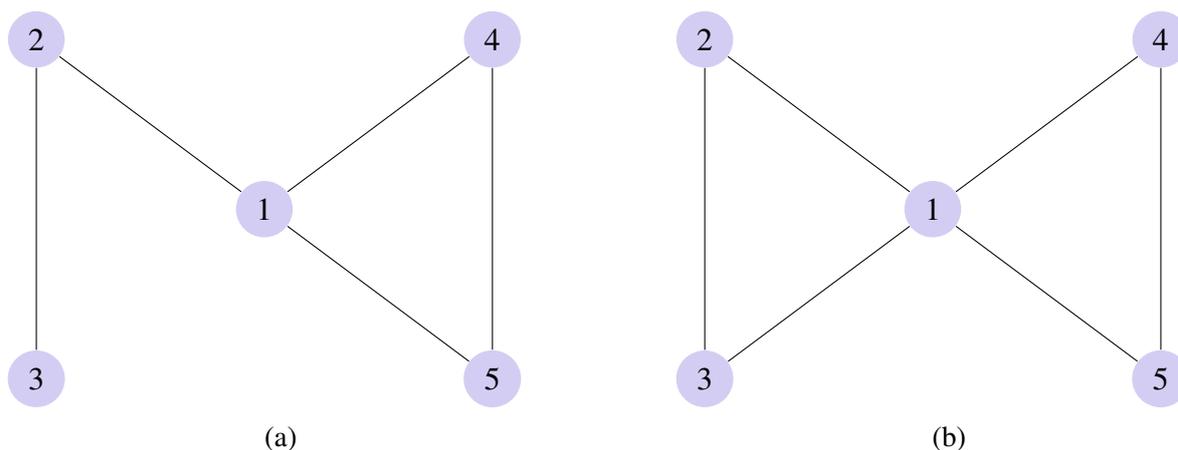


Figura 2 – (a) Grafo do modelo 542; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 2, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,4), (1,5), (2,3) e (4,5). Com uma moderada probabilidade, ele visitou grafos que tinham as arestas (1,2), (3,4) e (3,5). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (1,3), (2,4) e (2,5).

Tabela 2 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 100.000$ e $n = 100$.

Vértices	1	2	3	4	5
1	-	54.480	3.061	100.000	100.000
2	0,545	-	70.760	119	93
3	0,031	0,708	-	44.181	34.711
4	1,000	0,001	0,442	-	85.780
5	1,000	0,001	0,347	0,858	-

Fonte: Autor.

Na Tabela 3, observa-se que o algoritmo acertou em captar a presença das arestas (1,4) e (1,5), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a presença das arestas (1,2), (2,3) e (4,5), e a ausência das arestas (3,4) e (3,5). Todavia, o algoritmo errou em não captar a presença da aresta (1,3).

Tabela 3 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 100.000$ e $n = 100$.

Vértices	2	3	4	5
1	0,455	0,969	0,000	0,000
2	-	0,292	-0,001	-0,001
3	-	-	-0,442	-0,347
4	-	-	-	0,142

Fonte: Autor.

5.3.1.2 Algoritmo da Proposta 1 ($m = 100.000$ e $n = 100$)

Neste cenário, o modelo 542 foi o mais visitado, obtendo um total de 16.754 visitas. Entretanto, o modelo 544 obteve, somente, um total de 819 visitas. O $IS(542|544) = 9$ e $ISIU(542|544) = 0,90$. Com isso, tem-se que o modelo 542 possui uma semelhança de 90% com o modelo 544 que pode ser observada na Figura 3.

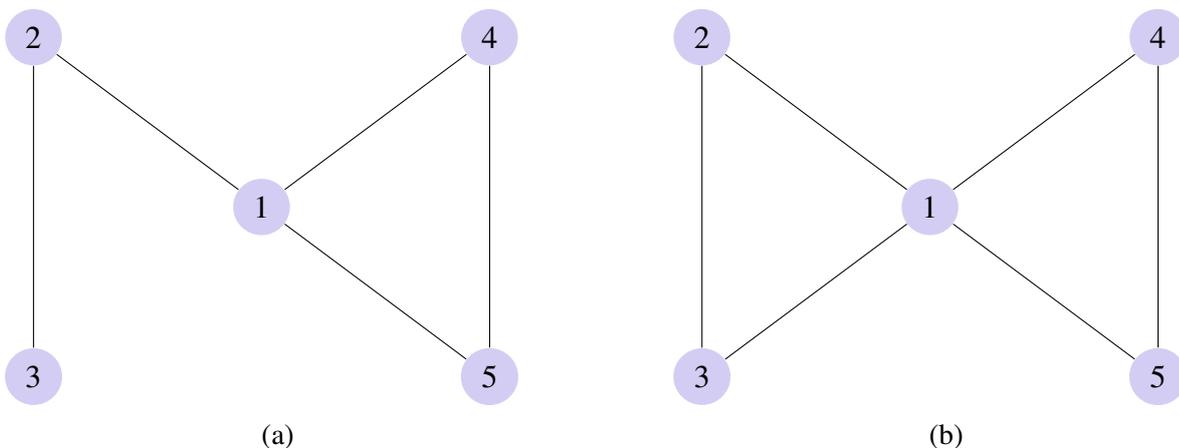


Figura 3 – (a) Grafo do modelo 542; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 4, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,4) e (1,5). Com uma moderada probabilidade, ele visitou grafos que tinham as arestas (1,2), (2,3) e (4,5). Com uma probabilidade baixa, ele visitou grafos que tinham as arestas (3,4) e (3,5). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (1,3), (2,4) e (2,5).

Na Tabela 5, observa-se que o algoritmo acertou em captar a presença das arestas (1,4) e (1,5), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a presença das arestas (1,2), (2,3) e (4,5), e a ausência das arestas (3,4) e (3,5). Todavia, o algoritmo errou em não captar a presença da aresta (1,3).

Tabela 4 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 100.000$ e $n = 100$.

Vértices	1	2	3	4	5
1	-	55.776	5.970	100.000	100.000
2	0,558	-	61.137	109	29
3	0,060	0,611	-	17.029	10.855
4	1,000	0,001	0,170	-	68.063
5	1,000	0,000	0,109	0,681	-

Fonte: Autor.

Tabela 5 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 100.000$ e $n = 100$.

Vértices	2	3	4	5
1	0,442	0,940	0,000	0,000
2	-	0,389	-0,001	0,000
3	-	-	-0,170	-0,109
4	-	-	-	0,319

Fonte: Autor.

5.3.1.3 Algoritmo da Proposta 2 ($m = 100.000$ e $n = 100$)

Neste cenário, o modelo 286 foi o mais visitado, obtendo um total de 21.077 visitas. Entretanto, o modelo 544 obteve, somente, um total de 690 visitas. O $IS(286|544) = 7$ e $ISIU(286|544) = 0,70$. Com isso, tem-se que o modelo 286 possui uma semelhança de 70% com o modelo 544 que pode ser observada na Figura 4.

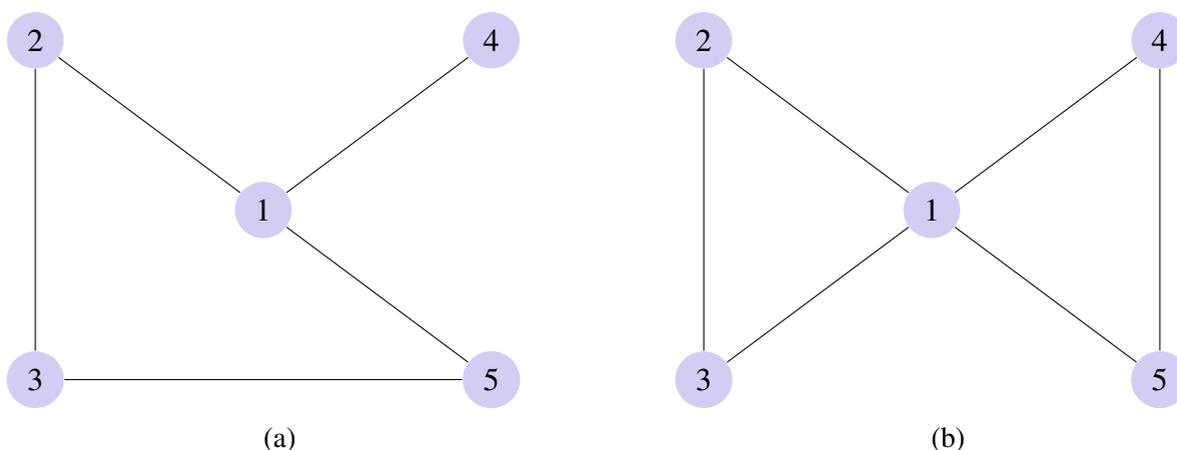


Figura 4 – (a) Grafo do modelo 286; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 6, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,4), (1,5) e (2,3). Com uma moderada probabilidade, ele visitou grafos que tinham a aresta (3,5). Com uma probabilidade baixa, ele visitou grafos que tinham as arestas (1,3), (3,4) e (4,5). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (2,4) e (2,5).

Tabela 6 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 100.000$ e $n = 100$.

Vértices	1	2	3	4	5
1	-	79.020	24.499	100.000	100.000
2	0,790	-	84.590	284	121
3	0,245	0,846	-	23.538	52.329
4	1,000	0,003	0,235	-	10.127
5	1,000	0,001	0,523	0,101	-

Fonte: Autor.

Na Tabela 7, observa-se que o algoritmo acertou em captar a presença das arestas (1,4) e (1,5), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a presença das arestas (1,2) e (2,3), e a ausência da aresta (3,4). Todavia, o algoritmo errou em não captar a presença das arestas (1,3) e (4,5), e a ausência da aresta (3,5).

Tabela 7 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 100.000$ e $n = 100$.

Vértices	2	3	4	5
1	0,210	0,755	0,000	0,000
2	-	0,154	-0,003	-0,001
3	-	-	-0,235	-0,523
4	-	-	-	0,899

Fonte: Autor.

5.3.1.4 Algoritmo da Proposta 3 ($m = 100.000$ e $n = 100$)

Neste cenário, o modelo 286 foi o mais visitado, obtendo um total de 15.578 visitas. Entretanto, o modelo 544 obteve, somente, um total de 1.222 visitas. O $IS(286|544) = 7$ e $ISIU(286|544) = 0,70$. Com isso, tem-se que o modelo 286 possui uma semelhança de 70% com o modelo 544 que pode ser observada na Figura 5.

Na Tabela 8, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,4), (1,5) e (2,3). Com uma moderada probabilidade, ele visitou grafos que tinham as arestas (1,3) e (3,5). Com uma probabilidade baixa, ele visitou

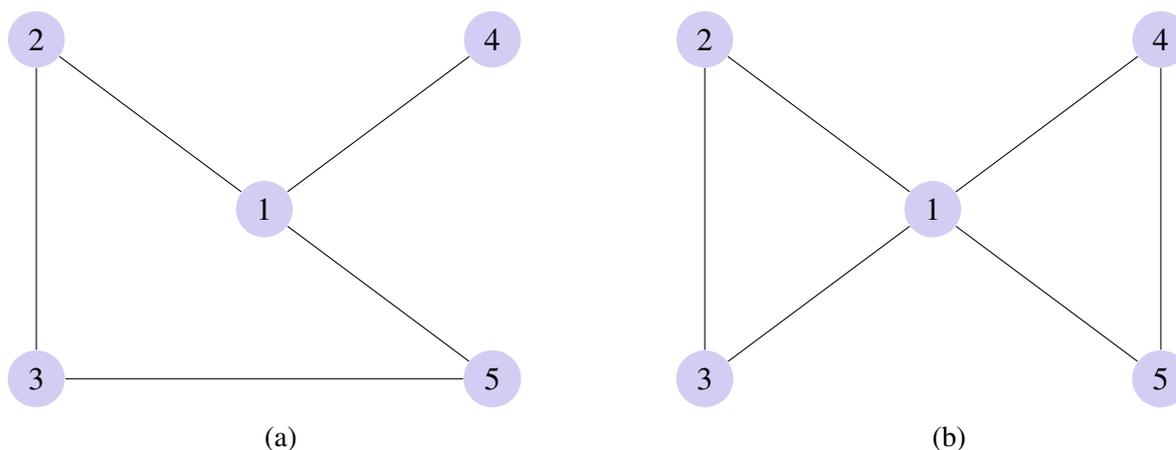


Figura 5 – (a) Grafo do modelo 286; (b) Grafo do modelo 544. Fonte: Autor.

grafos que tinham a aresta (3,4). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (2,4), (2,5) e (4,5).

Tabela 8 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 100.000$ e $n = 100$.

Vértices	1	2	3	4	5
1	-	76.441	36.948	100.000	100.000
2	0,764	-	72.487	446	29
3	0,369	0,725	-	11.628	43.276
4	1,000	0,004	0,116	-	6.025
5	1,000	0,000	0,433	0,060	-

Fonte: Autor.

Na Tabela 9, observa-se que o algoritmo acertou em captar a presença das arestas (1,4) e (1,5), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a presença das arestas (1,2) e (2,3), e a ausência da aresta (3,4). Todavia, o algoritmo errou em não captar a presença das arestas (1,3) e (4,5), e a ausência da aresta (3,5).

Tabela 9 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 100.000$ e $n = 100$.

Vértices	2	3	4	5
1	0,236	0,631	0,000	0,000
2	-	0,275	-0,004	0,000
3	-	-	-0,116	-0,433
4	-	-	-	0,940

Fonte: Autor.

5.3.1.5 Comparação dos Algoritmos ($m = 100.000$ e $n = 100$)

Na Figura 6, observa-se que não existem tendências nos gráficos de todas as variações do algoritmo *Reversible Jump*. Isso é um indício de que houve convergência nos algoritmos.

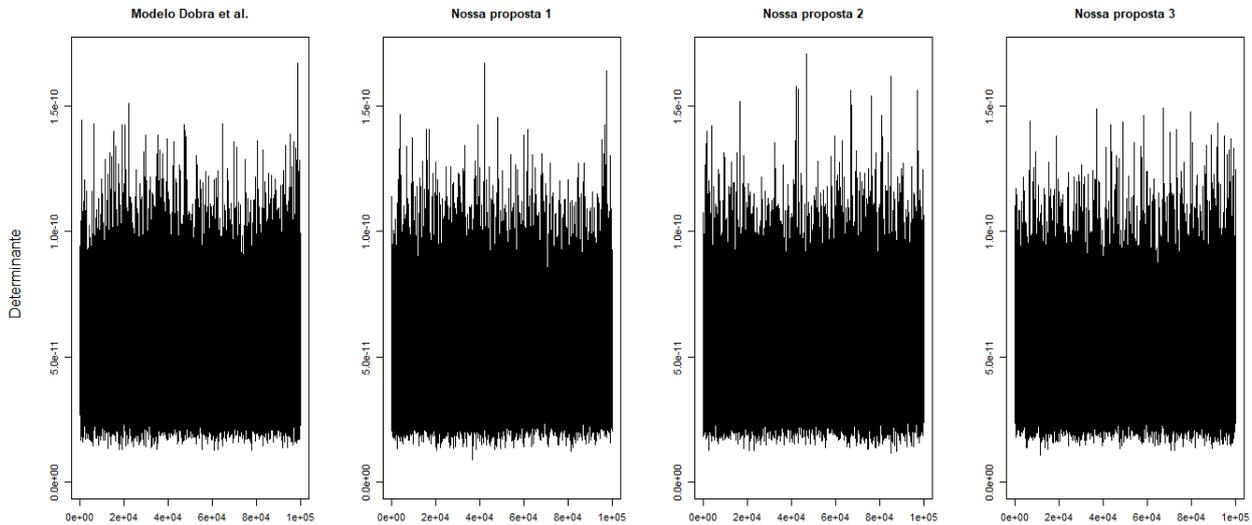


Figura 6 – Gráficos dos determinantes para avaliar a convergência dos algoritmos considerando $n = 100$, $m = 100.000$, warm-up = 100 e upgrade = 5. Fonte: Autor.

Na Figura 7, observa-se um decaimento brusco em todas as linhas decrescentes dos gráficos e que menos de 100 modelos foram visitados durante o processamento dos algoritmos. Além disso, o verdadeiro modelo 544 obteve baixos valores de probabilidade *a posteriori* em todas as variações. Entretanto, o algoritmo da proposta 3 teve um desempenho levemente superior aos algoritmos das propostas 1 e 2, e um desempenho melhor que o algoritmo de Dobra et al.

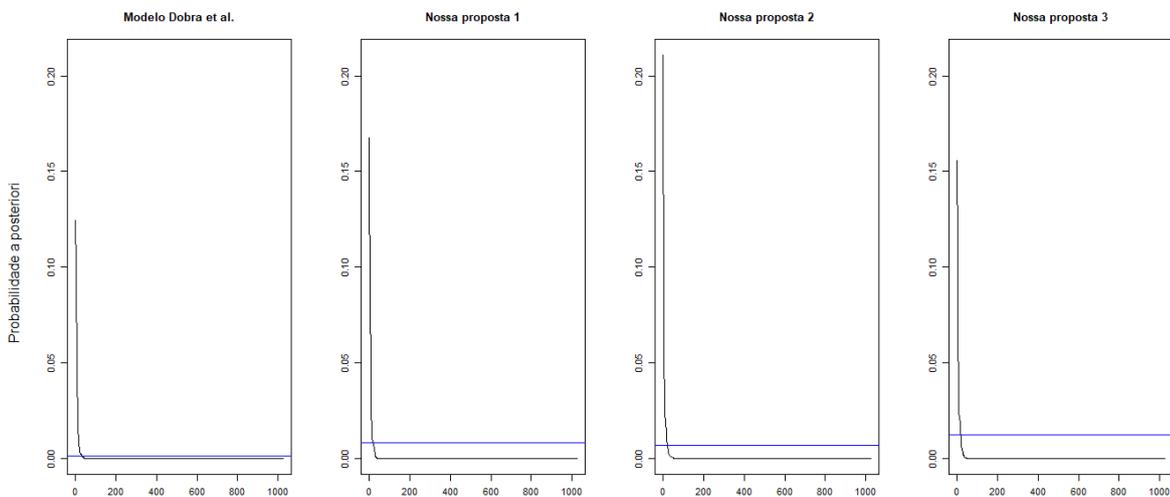


Figura 7 – Gráficos de avaliação das probabilidades *a posteriori* considerando $n = 100$, $m = 100.000$, warm-up = 100 e upgrade = 5. Fonte: Autor.

Na Tabela 10, observa-se que o algoritmo da proposta 1 obteve os melhores resultados em comparação às outras variações, pois teve os menores valores em todas as medidas SAE, SAEIU, SQE e SQEIU. Isso indica que o algoritmo da proposta 1 cometeu menos erros ao captar as presenças e ausências das arestas do verdadeiro modelo 544, em relação aos outros algoritmos. Considerando a Tabela 10 e as Figuras 6 e 7, pode-se concluir que o algoritmo da proposta 1 obteve o melhor desempenho para o cenário $m = 100.000$ e $n = 100$.

Tabela 10 – A soma absoluta dos erros (SAE), a soma absoluta dos erros no intervalo unitário (SAEIU), a soma quadrática dos erros (SQE) e a soma quadrática dos erros no intervalo unitário (SQEIU), para o cenário $m = 100.000$ e $n = 100$.

Algoritmo	SAE	SAEIU	SQE	SQEIU
Dobra et al.	2,650	0,265	1,568	0,157
Proposta 1	2,371	0,237	1,374	0,137
Proposta 2	2,780	0,278	1,775	0,177
Proposta 3	2,635	0,263	1,613	0,161

Fonte: Autor.

5.3.2 Resultados no cenário com $m = 100.000$ e $n = 250$

Nesta subseção, os resultados das simulações para o cenário $m = 100.000$ e $n = 250$ são apresentados e interpretados para as quatro variações do algoritmo *Reversible Jump*.

5.3.2.1 Algoritmo de Dobra et al. ($m = 100.000$ e $n = 250$)

Neste cenário, o modelo 32 foi o mais visitado, obtendo um total de 75.321 visitas. Entretanto, o modelo 544 obteve, somente, um total de 581 visitas. O $IS(32|544) = 9$ e $ISIU(32|544) = 0,90$. Com isso, tem-se que o modelo 32 possui uma semelhança de 90% com o modelo 544 que pode ser observada na Figura 8.

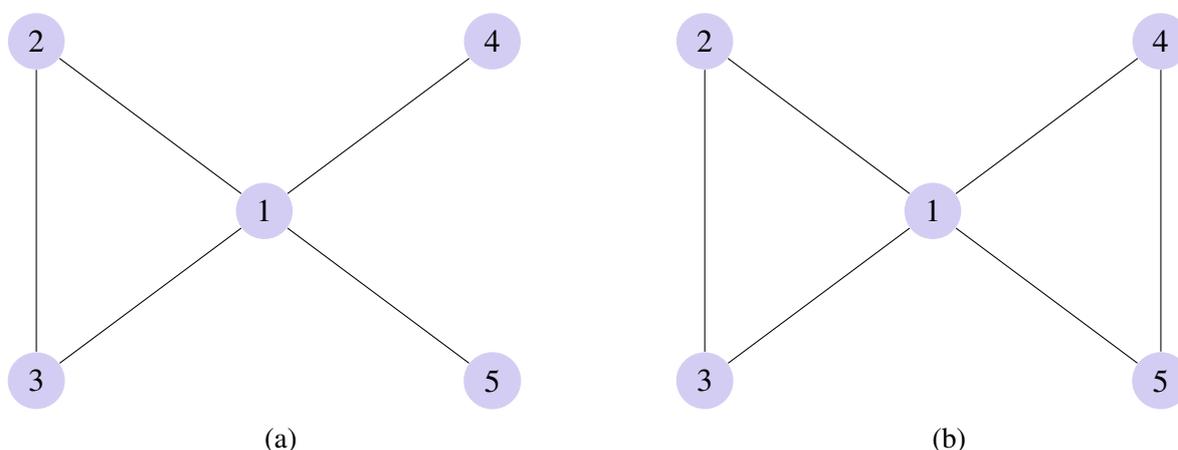


Figura 8 – (a) Grafo do modelo 32; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 11, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,3), (1,4), (1,5) e (2,3). Com uma probabilidade baixa, ele visitou grafos que tinham as arestas (3,4), (3,5) e (4,5). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (2,4) e (2,5).

Tabela 11 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 100.000$ e $n = 250$.

Vértices	1	2	3	4	5
1	-	100.000	100.000	100.000	100.000
2	1,000	-	100.000	112	86
3	1,000	1,000	-	23.376	14.579
4	1,000	0,001	0,234	-	24.091
5	1,000	0,001	0,146	0,241	-

Fonte: Autor.

Na Tabela 12, observa-se que o algoritmo acertou em captar a presença das arestas (1,2), (1,3), (1,4), (1,5) e (2,3), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a ausência das arestas (3,4) e (3,5). Todavia, o algoritmo errou em não captar a presença da aresta (4,5).

Tabela 12 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 100.000$ e $n = 250$.

Vértices	2	3	4	5
1	0,000	0,000	0,000	0,000
2	-	0,000	-0,001	-0,001
3	-	-	-0,234	-0,146
4	-	-	-	0,759

Fonte: Autor.

5.3.2.2 Algoritmo da Proposta 1 ($m = 100.000$ e $n = 250$)

Neste cenário, o modelo 32 foi o mais visitado, obtendo um total de 34.380 visitas. Entretanto, o modelo 544 obteve, somente, um total de 23.465 visitas. O $IS(32|544) = 9$ e $ISIU(32|544) = 0,90$. Com isso, tem-se que o modelo 32 possui uma semelhança de 90% com o modelo 544 que pode ser observada na Figura 9.

Na Tabela 13, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,3), (1,4), (1,5) e (2,3). Com uma probabilidade baixa, ele visitou grafos que tinham as arestas (3,4), (3,5) e (4,5). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (2,4) e (2,5).

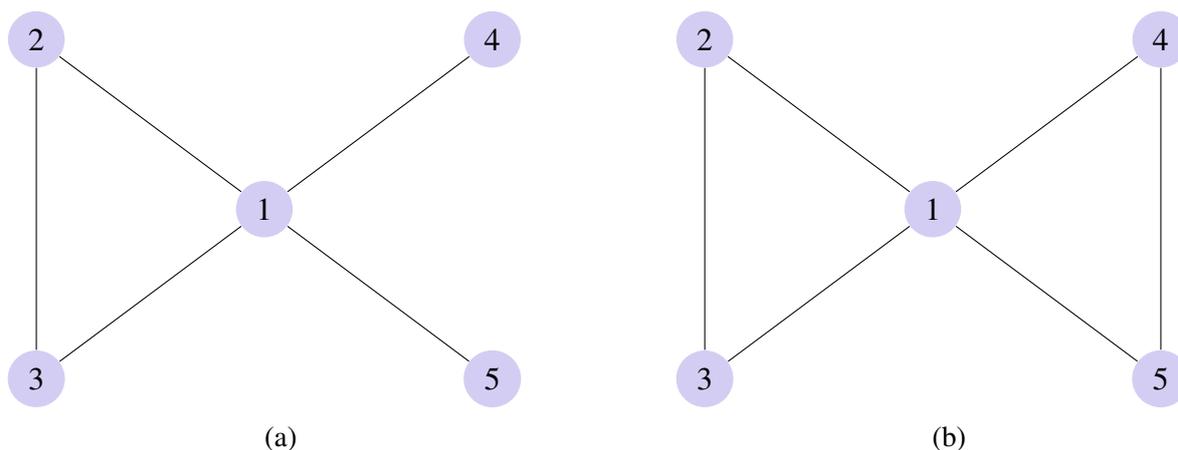


Figura 9 – (a) Grafo do modelo 32; (b) Grafo do modelo 544. Fonte: Autor.

Tabela 13 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 100.000$ e $n = 250$.

Vértices	1	2	3	4	5
1	-	100.000	100.000	100.000	100.000
2	1,000	-	100.000	28	29
3	1,000	1,000	-	26.036	24.604
4	1,000	0,000	0,260	-	32.581
5	1,000	0,000	0,246	0,326	-

Fonte: Autor.

Na Tabela 14, observa-se que o algoritmo acertou em captar a presença das arestas (1,2), (1,3), (1,4), (1,5) e (2,3), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a ausência das arestas (3,4) e (3,5). Todavia, o algoritmo errou em não captar a presença da aresta (4,5).

Tabela 14 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 100.000$ e $n = 250$.

Vértices	2	3	4	5
1	0,000	0,000	0,000	0,000
2	-	0,000	0,000	0,000
3	-	-	-0,260	-0,246
4	-	-	-	0,674

Fonte: Autor.

5.3.2.3 Algoritmo da Proposta 2 ($m = 100.000$ e $n = 250$)

Neste cenário, o modelo 288 foi o mais visitado, obtendo um total de 45.862 visitas. Entretanto, o modelo 544 obteve, somente, um total de 4.413 visitas. O $IS(288|544) = 8$ e $ISIU(288|544) = 0,80$. Com isso, tem-se que o modelo 288 possui uma semelhança de 80% com o modelo 544 que pode ser observada na Figura 10.

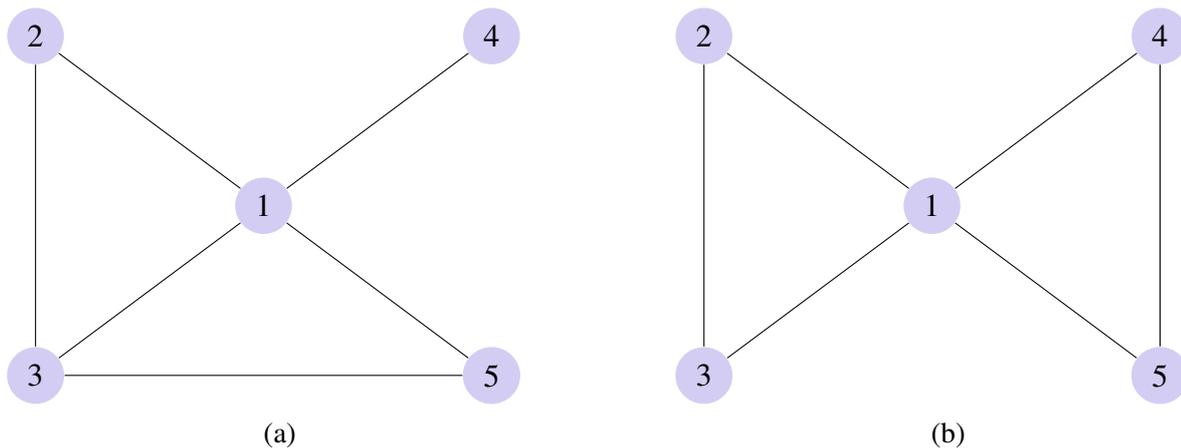


Figura 10 – (a) Grafo do modelo 288; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 15, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,3), (1,4), (1,5) e (2,3). Com uma moderada probabilidade, ele visitou grafos que tinham a aresta (3,5). Com uma probabilidade baixa, ele visitou grafos que tinham a aresta (3,4). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (2,4), (2,5) e (4,5).

Tabela 15 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 100.000$ e $n = 250$.

Vértices	1	2	3	4	5
1	-	100.000	100.000	100.000	100.000
2	1,000	-	100.000	357	65
3	1,000	1,000	-	12.444	55.939
4	1,000	0,004	0,124	-	5.637
5	1,000	0,001	0,559	0,056	-

Fonte: Autor.

Na Tabela 16, observa-se que o algoritmo acertou em captar a presença das arestas (1,2), (1,3), (1,4), (1,5) e (2,3), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a ausência da aresta (3,4). Todavia, o algoritmo errou em não captar a presença das arestas (4,5), e a ausência da aresta (3,5).

Tabela 16 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 100.000$ e $n = 250$.

Vértices	2	3	4	5
1	0,000	0,000	0,000	0,000
2	-	0,000	-0,004	-0,001
3	-	-	-0,124	-0,559
4	-	-	-	0,944

Fonte: Autor.

5.3.2.4 Algoritmo da Proposta 3 ($m = 100.000$ e $n = 250$)

Neste cenário, o modelo 32 foi o mais visitado, obtendo um total de 71.959 visitas. Entretanto, o modelo 544 obteve, somente, um total de 86 visitas. O $IS(32|544) = 9$ e $ISIU(32|544) = 0,90$. Com isso, tem-se que o modelo 32 possui uma semelhança de 90% com o modelo 544 que pode ser observada na Figura 11.

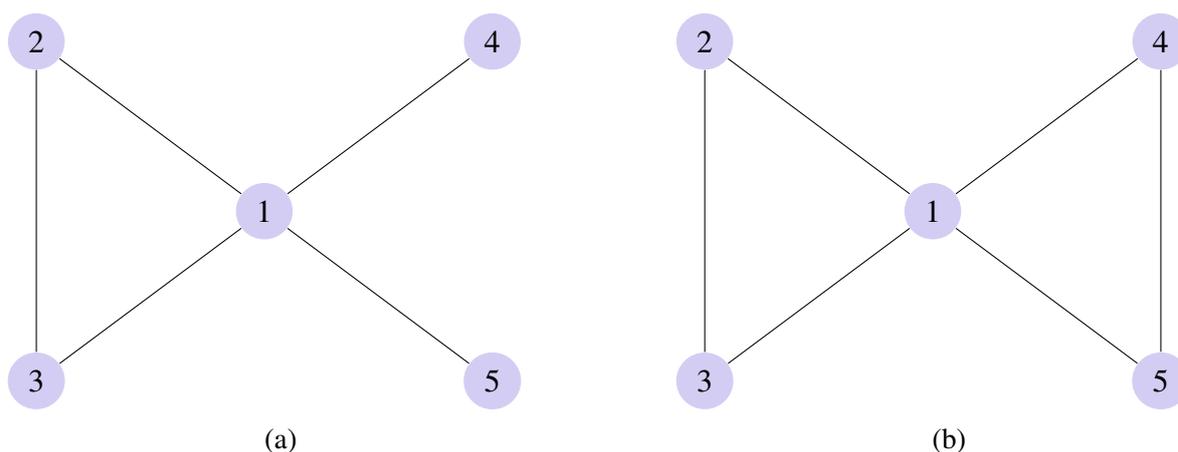


Figura 11 – (a) Grafo do modelo 32; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 17, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas $(1,2)$, $(1,3)$, $(1,4)$, $(1,5)$ e $(2,3)$. Com uma probabilidade baixa, ele visitou grafos que tinham a aresta $(3,5)$. E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas $(2,4)$, $(2,5)$, $(3,4)$ e $(4,5)$.

Na Tabela 18, observa-se que o algoritmo acertou em captar a presença das arestas $(1,2)$, $(1,3)$, $(1,4)$, $(1,5)$ e $(2,3)$, e a ausência das arestas $(2,4)$, $(2,5)$ e $(3,4)$. O algoritmo também conseguiu parcialmente captar a ausência da aresta $(3,5)$. Todavia, o algoritmo errou em não captar a presença da aresta $(4,5)$.

Tabela 17 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 100.000$ e $n = 250$.

Vértices	1	2	3	4	5
1	-	100.000	100.000	100.000	100.000
2	1,000	-	100.000	164	9
3	1,000	1,000	-	32	27.817
4	1,000	0,002	0,000	-	107
5	1,000	0,000	0,278	0,001	-

Fonte: Autor.

Tabela 18 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 100.000$ e $n = 250$.

Vértices	2	3	4	5
1	0,000	0,000	0,000	0,000
2	-	0,000	-0,002	0,000
3	-	-	0,000	-0,278
4	-	-	-	0,999

Fonte: Autor.

5.3.2.5 Comparação dos Algoritmos ($m = 100.000$ e $n = 250$)

Na Figura 12, observa-se que não existem tendências nos gráficos de todas as variações do algoritmo *Reversible Jump*. Isso é um indício de que houve convergência nos algoritmos.

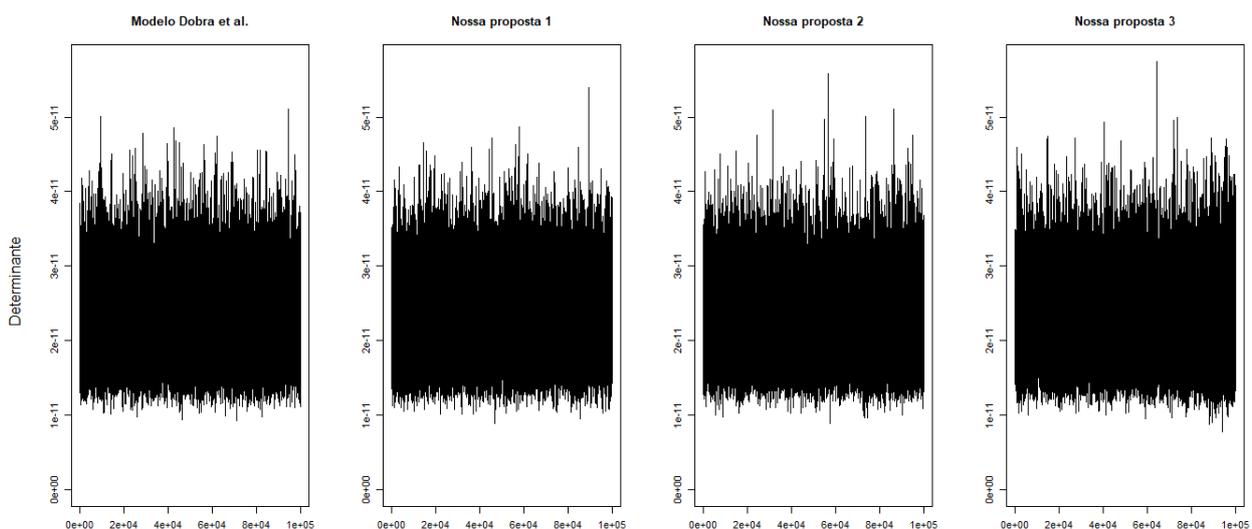


Figura 12 – Gráficos dos determinantes para avaliar a convergência dos algoritmos considerando $n = 250$, $m = 100.000$, warm-up = 100 e upgrade = 5. Fonte: Autor.

Na Figura 13, observa-se um decaimento brusco em todas as linhas decrescentes dos

gráficos e que menos de 100 modelos foram visitados durante o processamento dos algoritmos. Além disso, o verdadeiro modelo 544 obteve baixos valores de probabilidade *a posteriori* nos algoritmos de Dobra et al. e da proposta 3. Todavia, o algoritmo da proposta 1 obteve o maior valor de probabilidade *a posteriori* para o modelo 544. Com isso, o algoritmo da proposta 1 teve um desempenho superior aos outros algoritmos.

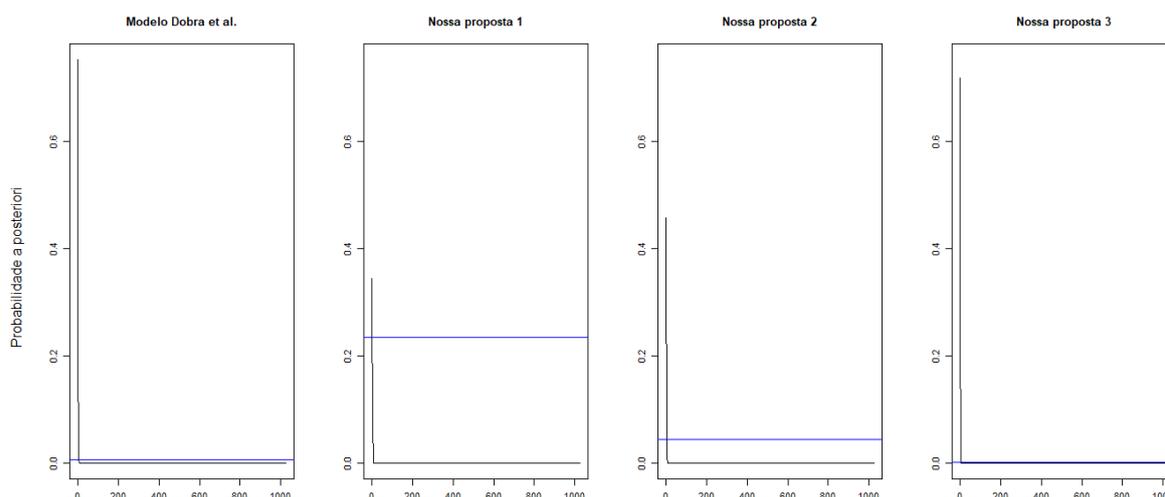


Figura 13 – Gráficos de avaliação das probabilidades *a posteriori* considerando $n = 250$, $m = 100.000$, $\text{warm-up} = 100$ e $\text{upgrade} = 5$. Fonte: Autor.

Na Tabela 19, observa-se que o algoritmo de Dobra et al. obteve os melhores resultados de SAE e SAEIU em comparação as outras variações, pois teve os menores valores. Entretanto, o algoritmo da proposta 1 obteve os melhores resultados de SQE e SQEIU em comparação as outras variações. Isso indica que os algoritmos do Dobra et al. e da proposta 1 cometeram menos erros ao captar as presenças e ausências das arestas do verdadeiro modelo 544, em relação ao outros algoritmos. Considerando a Tabela 19 e as Figuras 12 e 13, pode-se concluir que o algoritmo da proposta 1 obteve o melhor desempenho para o cenário $m = 100.000$ e $n = 250$.

Tabela 19 – A soma absoluta dos erros (SAE), a soma absoluta dos erros no intervalo unitário (SAEIU), a soma quadrática dos erros (SQE) e a soma quadrática dos erros no intervalo unitário (SQEIU), para o cenário $m = 100.000$ e $n = 250$.

Algoritmo	SAE	SAEIU	SQE	SQEIU
Dobra et al.	1,141	0,114	0,652	0,065
Proposta 1	1,181	0,118	0,583	0,058
Proposta 2	1,632	0,163	1,219	0,122
Proposta 3	1,279	0,128	1,075	0,108

Fonte: Autor.

5.3.3 Resultados no cenário com $m = 100.000$ e $n = 500$

Nesta subsecção, os resultados das simulações para o cenário $m = 100.000$ e $n = 500$ são apresentados e interpretados para as quatro variações do algoritmo *Reversible Jump*.

5.3.3.1 Algoritmo de Dobra et al. ($m = 100.000$ e $n = 500$)

Neste cenário, o modelo 928 foi o mais visitado, obtendo um total de 27.030 visitas. Entretanto, o modelo 544 obteve, somente, um total de 13.204 visitas. O $IS(928|544) = 8$ e $ISIU(928|544) = 0,80$. Com isso, tem-se que o modelo 928 possui uma semelhança de 80% com o modelo 544 que pode ser observada na Figura 14.

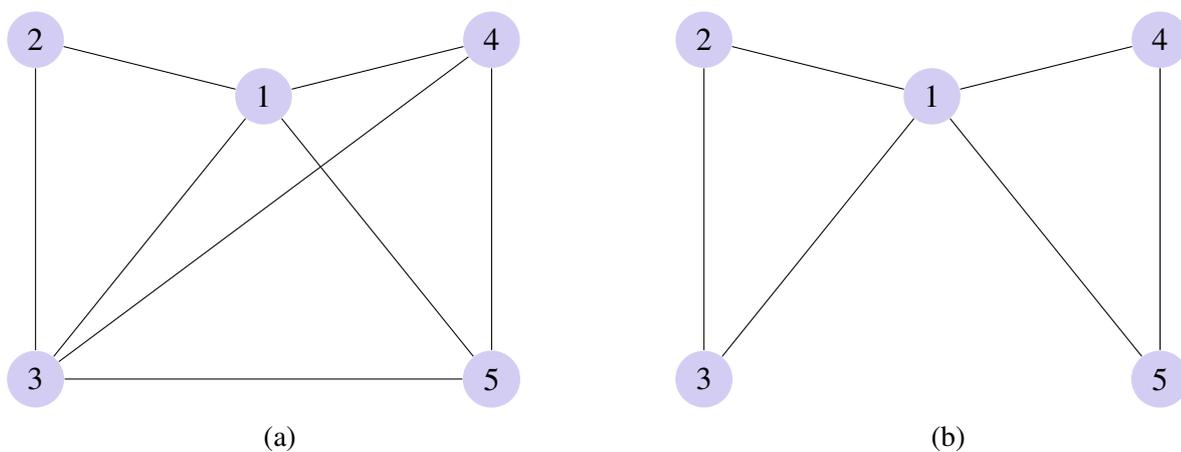


Figura 14 – (a) Grafo do modelo 928; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 20, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas $(1, 2)$, $(1, 3)$, $(1, 4)$, $(1, 5)$, $(2, 3)$ e $(4, 5)$. Com uma moderada probabilidade, ele visitou grafos que tinham as arestas $(3, 4)$ e $(3, 5)$. E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas $(2, 4)$ e $(2, 5)$.

Tabela 20 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 100.000$ e $n = 500$.

Vértices	1	2	3	4	5
1	-	100.000	100.000	100.000	100.000
2	1,000	-	100.000	68	19
3	1,000	1,000	-	63.431	50.531
4	1,000	0,001	0,634	-	73.596
5	1,000	0,000	0,505	0,736	-

Fonte: Autor.

Na Tabela 21, observa-se que o algoritmo acertou em captar a presença das arestas $(1,2)$, $(1,3)$, $(1,4)$, $(1,5)$ e $(2,3)$, e a ausência das arestas $(2,4)$ e $(2,5)$. O algoritmo também conseguiu parcialmente captar a presença da aresta $(4,5)$. Todavia, o algoritmo errou em não captar a ausência das arestas $(3,4)$ e $(3,5)$.

Tabela 21 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 100.000$ e $n = 500$.

Vértices	2	3	4	5
1	0,000	0,000	0,000	0,000
2	-	0,000	-0,001	0,000
3	-	-	-0,634	-0,505
4	-	-	-	0,264

Fonte: Autor.

5.3.3.2 Algoritmo da Proposta 1 ($m = 100.000$ e $n = 500$)

Neste cenário, o modelo 160 foi o mais visitado, obtendo um total de 26.867 visitas. Entretanto, o modelo 544 obteve, somente, um total de 25.701 visitas. O $IS(160|544) = 8$ e $ISIU(160|544) = 0,80$. Com isso, tem-se que o modelo 160 possui uma semelhança de 80% com o modelo 544 que pode ser observada na Figura 15.

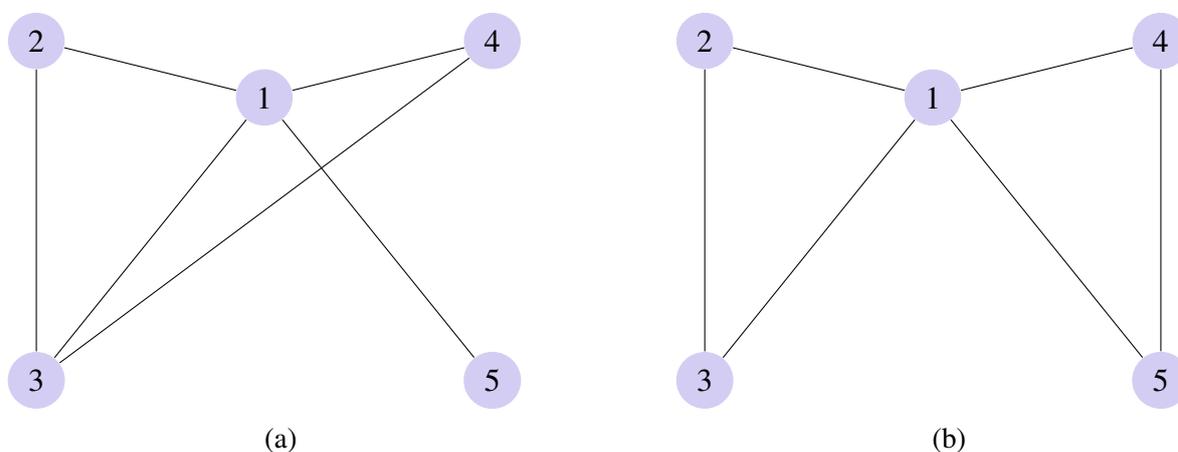


Figura 15 – (a) Grafo do modelo 160; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 22, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas $(1,2)$, $(1,3)$, $(1,4)$, $(1,5)$ e $(2,3)$. Com uma moderada probabilidade, ele visitou grafos que tinham a aresta $(3,4)$. Com uma probabilidade baixa, ele visitou grafos que tinham as arestas $(3,5)$ e $(4,5)$. E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas $(2,4)$ e $(2,5)$.

Tabela 22 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 100.000$ e $n = 500$.

Vértices	1	2	3	4	5
1	-	100.000	100.000	100.000	100.000
2	1,000	-	100.000	13	64
3	1,000	1,000	-	41.323	23.802
4	1,000	0,000	0,413	-	26.981
5	1,000	0,001	0,238	0,270	-

Fonte: Autor.

Na Tabela 23, observa-se que o algoritmo acertou em captar a presença das arestas (1,2), (1,3), (1,4), (1,5) e (2,3), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a ausência da aresta (3,5). Todavia, o algoritmo errou em não captar a presença da aresta (4,5) e a ausência da aresta (3,4).

Tabela 23 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 100.000$ e $n = 500$.

Vértices	2	3	4	5
1	0,000	0,000	0,000	0,000
2	-	0,000	0,000	-0,001
3	-	-	-0,413	-0,238
4	-	-	-	0,730

Fonte: Autor.

5.3.3.3 Algoritmo da Proposta 2 ($m = 100.000$ e $n = 500$)

Neste cenário, o modelo 32 foi o mais visitado, obtendo um total de 53.332 visitas. Entretanto, o modelo 544 obteve, somente, um total de 161 visitas. O $IS(32|544) = 9$ e $ISIU(32|544) = 0,90$. Com isso, tem-se que o modelo 32 possui uma semelhança de 90% com o modelo 544 que pode ser observada na Figura 16.

Na Tabela 24, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,3), (1,4), (1,5) e (2,3). Com uma moderada probabilidade, ele visitou grafos que tinham a aresta (3,5). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (2,4), (2,5), (3,4) e (4,5).

Na Tabela 25, observa-se que o algoritmo acertou em captar a presença das arestas (1,2), (1,3), (1,4), (1,5) e (2,3), e a ausência das arestas (2,4), (2,5) e (3,4). O algoritmo também conseguiu parcialmente captar a ausência da aresta (3,5). Todavia, o algoritmo errou em não captar a presença das arestas (4,5).

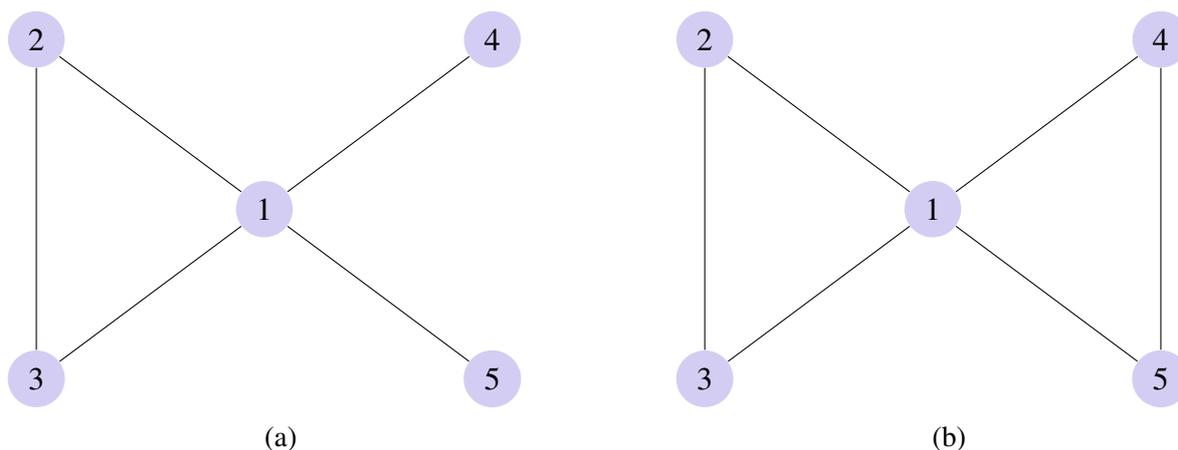


Figura 16 – (a) Grafo do modelo 32; (b) Grafo do modelo 544. Fonte: Autor.

Tabela 24 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 100.000$ e $n = 500$.

Vértices	1	2	3	4	5
1	-	100.000	100.000	100.000	100.000
2	1,000	-	100.000	230	48
3	1,000	1,000	-	2.580	45.179
4	1,000	0,002	0,026	-	1.673
5	1,000	0,000	0,452	0,017	-

Fonte: Autor.

Tabela 25 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 100.000$ e $n = 500$.

Vértices	2	3	4	5
1	0,000	0,000	0,000	0,000
2	-	0,000	-0,002	0,000
3	-	-	-0,026	-0,452
4	-	-	-	0,983

Fonte: Autor.

5.3.3.4 Algoritmo da Proposta 3 ($m = 100.000$ e $n = 500$)

Neste cenário, o modelo 32 foi o mais visitado, obtendo um total de 74.620 visitas. Entretanto, o modelo 544 obteve, somente, um total de 10.599 visitas. O $IS(32|544) = 9$ e $ISIU(32|544) = 0,90$. Com isso, tem-se que o modelo 32 possui uma semelhança de 90% com o modelo 544 que pode ser observada na Figura 17.

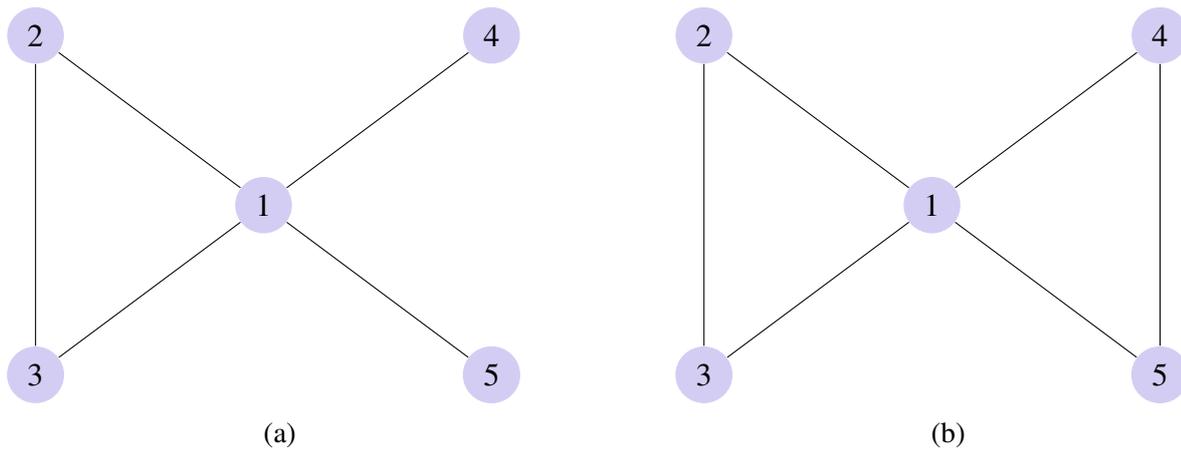


Figura 17 – (a) Grafo do modelo 32; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 26, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,3), (1,4), (1,5) e (2,3). Com uma probabilidade baixa, ele visitou grafos que tinham a aresta (3,5) e (4,5). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (2,4), (2,5) e (3,4).

Tabela 26 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 100.000$ e $n = 500$.

Vértices	1	2	3	4	5
1	-	100.000	100.000	100.000	100.000
2	1,000	-	100.000	175	83
3	1,000	1,000	-	280	14.313
4	1,000	0,002	0,003	-	11.026
5	1,000	0,001	0,143	0,110	-

Fonte: Autor.

Na Tabela 27, observa-se que o algoritmo acertou em captar a presença das arestas (1,2), (1,3), (1,4), (1,5) e (2,3), e a ausência das arestas (2,4), (2,5) e (3,4). O algoritmo também conseguiu parcialmente captar a ausência da aresta (3,5). Todavia, o algoritmo errou em não captar a presença da aresta (4,5).

5.3.3.5 Comparação dos Algoritmos ($m = 100.000$ e $n = 500$)

Na Figura 18, observa-se que não existem tendências nos gráficos de todas as variações do algoritmo *Reversible Jump*. Isso é um indício de que houve convergência nos algoritmos.

Na Figura 19, observa-se um decaimento brusco em todas as linhas decrescentes dos gráficos. Pela Figura, temos que menos de 100 modelos foram visitados durante o processamento dos algoritmos. Além disso, o verdadeiro modelo 544 obteve baixo valor de probabilidade *a posteriori* no algoritmo da proposta 2. Todavia, o algoritmo da proposta 1 obteve o maior valor de

Tabela 27 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 100.000$ e $n = 500$.

Vértices	2	3	4	5
1	0,000	0,000	0,000	0,000
2	-	0,000	-0,002	-0,001
3	-	-	-0,003	-0,143
4	-	-	-	0,890

Fonte: Autor.

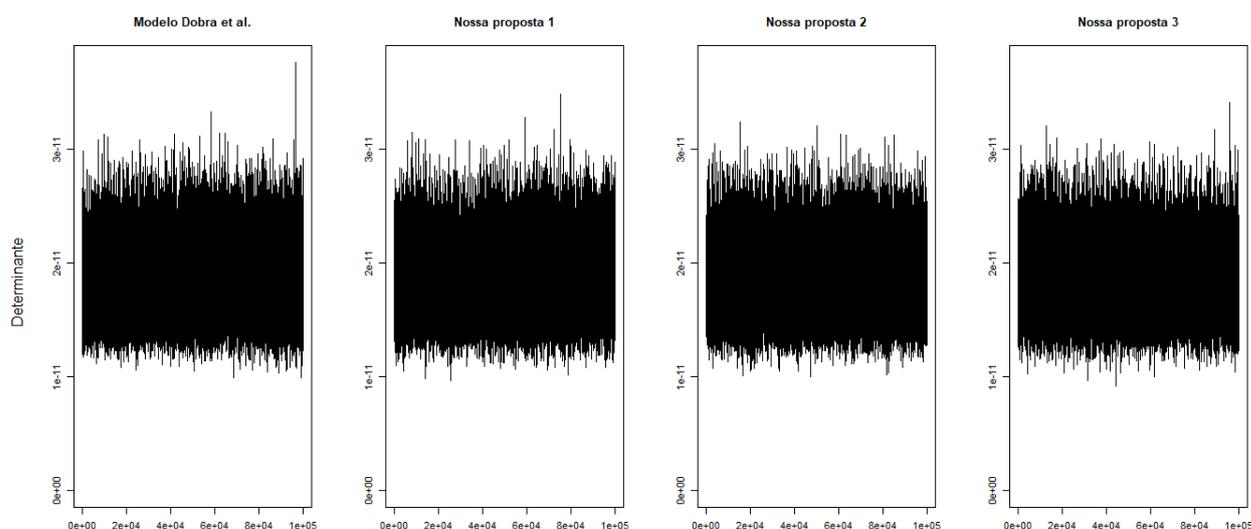


Figura 18 – Gráficos dos determinantes para avaliar a convergência dos algoritmos considerando $n = 500$, $m = 100.000$, warm-up = 100 e upgrade = 5. Fonte: Autor.

probabilidade *a posteriori* para o modelo 544, tendo em sequência os algoritmos de Dobra et al. e da proposta 3. Com isso, a da proposta 1 teve um desempenho superior aos outros algoritmos.

Na Tabela 28, observa-se que o algoritmo da proposta 3 obteve os melhores resultados de SAE e SAEIU em comparação as outras variações, pois teve os menores valores. Entretanto, o algoritmo do modelo Dobra et al. obteve os melhores resultados de SQE e SQEIU em comparação as outras variações, mas seu resultado foi levemente superior a da proposta 1. Isso indica que os algoritmo de Dobra et al. e a da proposta 3 cometeram menos erros ao captar as presenças e ausências das arestas do verdadeiro modelo 544, em relação ao outros algoritmos. Considerando a Tabela 28 e as Figuras 18 e 19, pode-se concluir que a da proposta 1 obteve o melhor desempenho para o cenário $m = 100.000$ e $n = 500$.

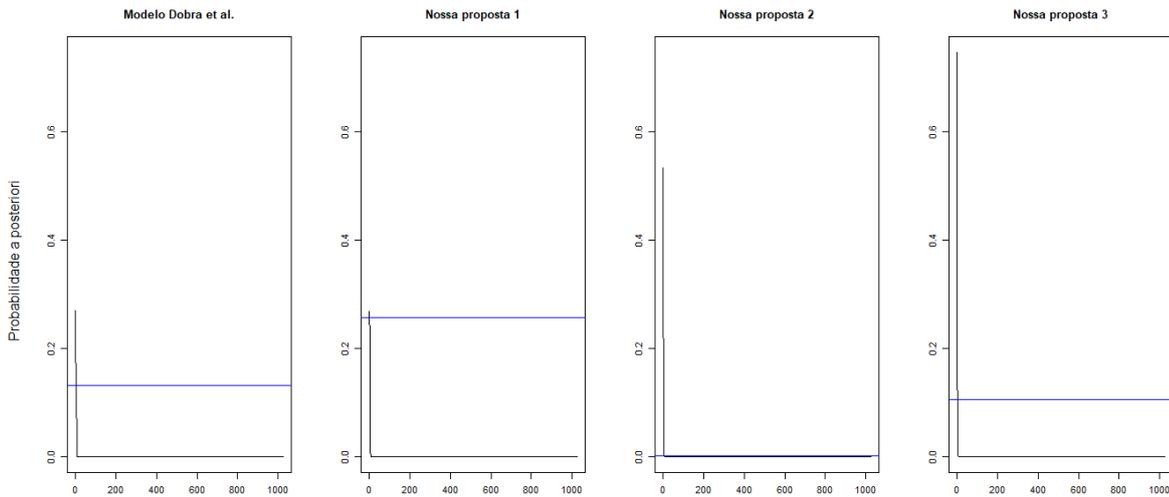


Figura 19 – Gráficos de avaliação das probabilidades *a posteriori* considerando $n = 500$, $m = 100.000$, $\text{warm-up} = 100$ e $\text{upgrade} = 5$. Fonte: Autor.

Tabela 28 – A soma absoluta dos erros (SAE), a soma absoluta dos erros no intervalo unitário (SAEIU), a soma quadrática dos erros (SQE) e a soma quadrática dos erros no intervalo unitário (SQEIU), para o cenário $m = 100.000$ e $n = 500$.

Algoritmo	SAE	SAEIU	SQE	SQEIU
Dobra et al.	1,405	0,140	0,727	0,073
Proposta 1	1,382	0,138	0,761	0,076
Proposta 2	1,464	0,146	1,172	0,117
Proposta 3	1,038	0,104	0,812	0,081

Fonte: Autor.

5.3.4 Resultados no cenário com $m = 200.000$ e $n = 100$

Nesta subseção, os resultados das simulações para o cenário $m = 200.000$ e $n = 100$ são apresentados e interpretados para as quatro variações do algoritmo *Reversible Jump*.

5.3.4.1 Algoritmo de Dobra et al. ($m = 200.000$ e $n = 100$)

Neste cenário, o modelo 525 foi o mais visitado, obtendo um total de 28.861 visitas. Entretanto, o modelo 544 obteve, somente, um total de 172 visitas. O $\text{IS}(525|544) = 7$ e $\text{ISIU}(525|544) = 0,70$. Com isso, tem-se que o modelo 525 possui uma semelhança de 70% com o modelo 544 que pode ser observada na Figura 20.

Na Tabela 29, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,4), (1,5) e (4,5). Com uma moderada probabilidade, ele visitou grafos que tinham as arestas (1,2), (2,3), (3,4) e (3,5). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (1,3), (2,4) e (2,5).

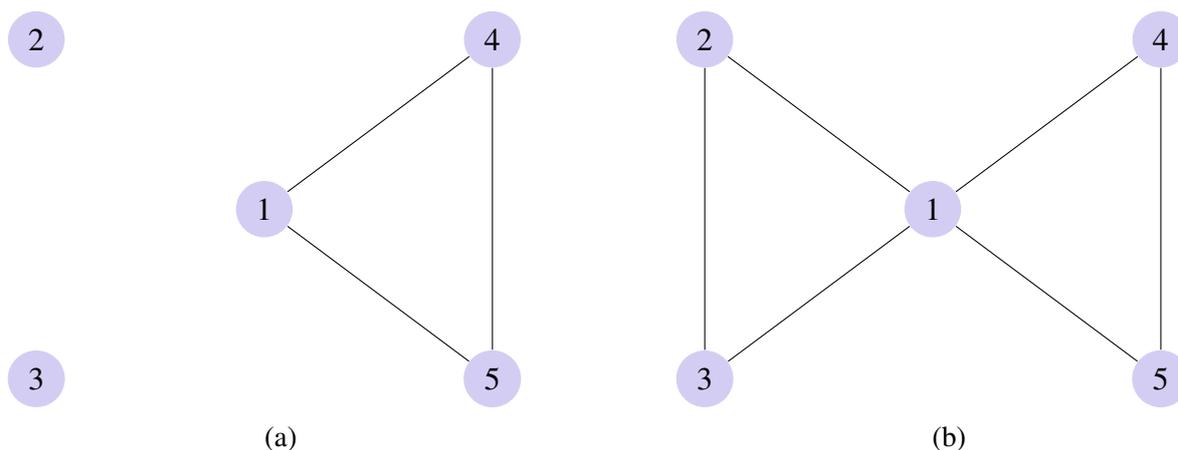


Figura 20 – (a) Grafo do modelo 525; (b) Grafo do modelo 544. Fonte: Autor.

Tabela 29 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 200.000$ e $n = 100$.

Vértices	1	2	3	4	5
1	-	95.692	7.208	200.000	200.000
2	0,478	-	84.018	189	171
3	0,036	0,420	-	89.407	65.438
4	1,000	0,001	0,447	-	185.780
5	1,000	0,001	0,327	0,929	-

Fonte: Autor.

Na Tabela 30, observa-se que o algoritmo acertou em captar a presença das arestas (1,4), (1,5) e (4,5), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a ausência das arestas (3,4) e (3,5). Todavia, o algoritmo errou em não captar a presença das arestas (1,2), (1,3) e (2,3).

Tabela 30 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 200.000$ e $n = 100$.

Vértices	2	3	4	5
1	0,522	0,964	0,000	0,000
2	-	0,580	-0,001	-0,001
3	-	-	-0,447	-0,327
4	-	-	-	0,071

Fonte: Autor.

5.3.4.2 Algoritmo da Proposta 1 ($m = 200.000$ e $n = 100$)

Neste cenário, o modelo 542 foi o mais visitado, obtendo um total de 30.911 visitas. Entretanto, o modelo 544 obteve, somente, um total de 1.851 visitas. O $IS(542|544) = 9$ e $ISIU(542|544) = 0,90$. Com isso, tem-se que o modelo 542 possui uma semelhança de 90% com o modelo 544 que pode ser observada na Figura 21.

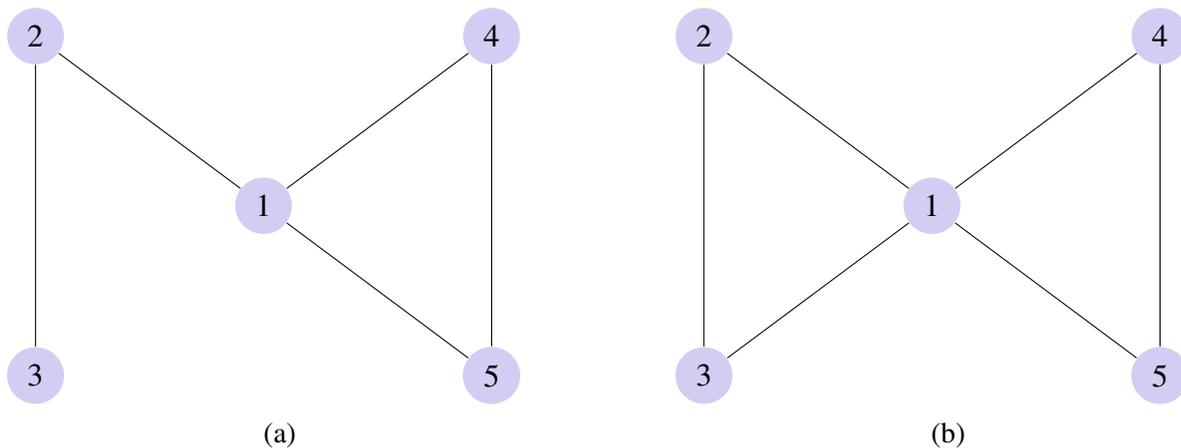


Figura 21 – (a) Grafo do modelo 542; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 31, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,4) e (1,5). Com uma moderada probabilidade, ele visitou grafos que tinham as arestas (1,2), (2,3) e (4,5). Com uma probabilidade baixa, ele visitou grafos que tinham as arestas (3,4) e (3,5). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (1,3), (2,4) e (2,5).

Tabela 31 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 200.000$ e $n = 100$.

Vértices	1	2	3	4	5
1	-	117.798	15.373	200.000	200.000
2	0,589	-	130.076	130	83
3	0,077	0,650	-	33.998	23.011
4	1,000	0,001	0,170	-	107.773
5	1,000	0,000	0,115	0,539	-

Fonte: Autor.

Na Tabela 32, observa-se que o algoritmo acertou em captar a presença das arestas (1,4) e (1,5), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a presença das arestas (1,2), (2,3) e (4,5), e a ausência da aresta (3,4) e (3,5). Todavia, o algoritmo errou em não captar a presença da aresta (1,3).

Tabela 32 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 200.000$ e $n = 100$.

Vértices	2	3	4	5
1	0,411	0,923	0,000	0,000
2	-	0,350	-0,001	0,000
3	-	-	-0,170	-0,115
4	-	-	-	0,461

Fonte: Autor.

5.3.4.3 Algoritmo da Proposta 2 ($m = 200.000$ e $n = 100$)

Neste cenário, o modelo 286 foi o mais visitado, obtendo um total de 52.547 visitas. Entretanto, o modelo 544 obteve, somente, um total de 161 visitas. O $IS(286|544) = 7$ e $ISIU(286|544) = 0,70$. Com isso, tem-se que o modelo 286 possui uma semelhança de 70% com o modelo 544 que pode ser observada na Figura 22.

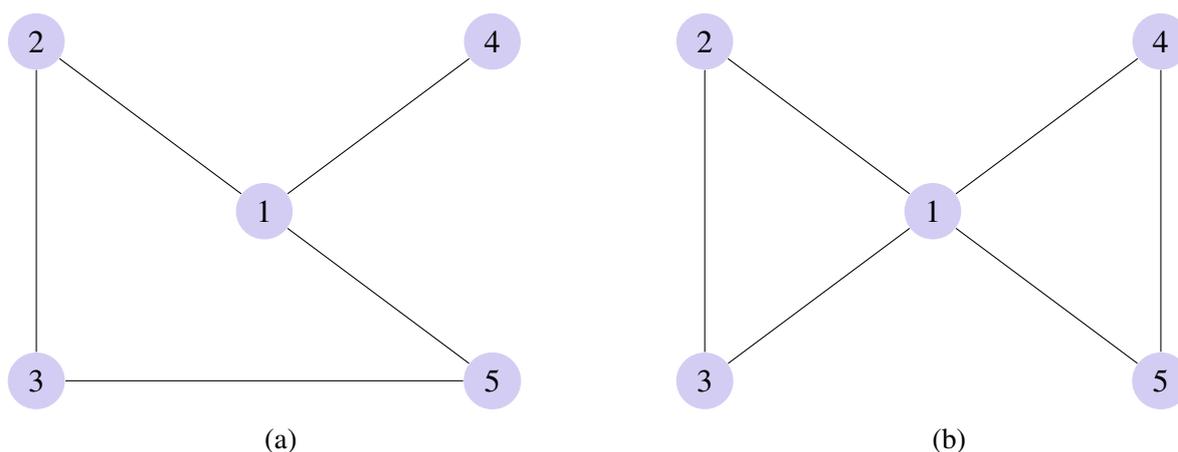


Figura 22 – (a) Grafo do modelo 286; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 33, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,4), (1,5) e (2,3). Com uma moderada probabilidade, ele visitou grafos que tinham a aresta (3,5). Com uma probabilidade baixa, ele visitou grafos que tinham as arestas (1,3) e (3,4). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (2,4), (2,5) e (4,5).

Na Tabela 34, observa-se que o algoritmo acertou em captar a presença das arestas (1,4) e (1,5), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a presença das arestas (1,2) e (2,3), e a ausência da aresta (3,4). Todavia, o algoritmo errou em não captar a presença das arestas (1,3) e (4,5), e a ausência da aresta (3,5).

Tabela 33 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 200.000$ e $n = 100$.

Vértices	1	2	3	4	5
1	-	153.725	46.329	200.000	200.000
2	0,769	-	167.795	884	148
3	0,232	0,839	-	42.805	113.822
4	1,000	0,004	0,214	-	13.988
5	1,000	0,001	0,569	0,070	-

Fonte: Autor.

Tabela 34 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 200.000$ e $n = 100$.

Vértices	2	3	4	5
1	0,231	0,768	0,000	0,000
2	-	0,161	-0,004	-0,001
3	-	-	-0,214	-0,569
4	-	-	-	0,930

Fonte: Autor.

5.3.4.4 Algoritmo da Proposta 3 ($m = 200.000$ e $n = 100$)

Neste cenário, o modelo 286 foi o mais visitado, obtendo um total de 30.172 visitas. Entretanto, o modelo 544 obteve, somente, um total de 3.700 visitas. O $IS(286|544) = 7$ e $ISIU(286|544) = 0,70$. Com isso, tem-se que o modelo 286 possui uma semelhança de 70% com o modelo 544 que pode ser observada na Figura 23.

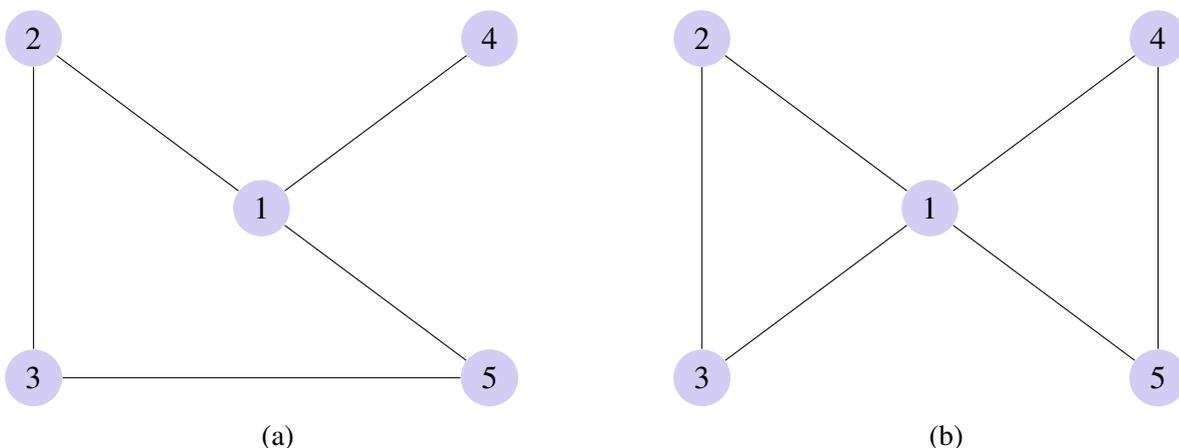


Figura 23 – (a) Grafo do modelo 286; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 35, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,4), (1,5) e (2,3). Com uma moderada probabilidade, ele

visitou grafos que tinham as arestas (1,3) e (3,5). Com uma probabilidade baixa, ele visitou grafos que tinham a aresta (3,4) e (4,5). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (2,4) e (2,5).

Tabela 35 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 200.000$ e $n = 100$.

Vértices	1	2	3	4	5
1	-	167.948	69.280	200.000	186.298
2	0,840	-	152.505	934	58
3	0,346	0,763	-	33.332	90.884
4	1,000	0,005	0,167	-	20.646
5	0,931	0,000	0,454	0,103	-

Fonte: Autor.

Na Tabela 36, observa-se que o algoritmo acertou em captar a presença das arestas (1,4) e (1,5), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a presença das arestas (1,2) e (2,3), e a ausência da aresta (3,4). Todavia, o algoritmo errou em não captar a presença da aresta (1,3) e (4,5), e a ausência da aresta (3,5).

Tabela 36 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 200.000$ e $n = 100$.

Vértices	2	3	4	5
1	0,160	0,654	0,000	0,069
2	-	0,237	-0,005	0,000
3	-	-	-0,167	-0,454
4	-	-	-	0,897

Fonte: Autor.

5.3.4.5 Comparação dos Algoritmos ($m = 200.000$ e $n = 100$)

Na Figura 24, observa-se que não existem tendências nos gráficos de todas as variações do algoritmo *Reversible Jump*. Isso é um indício de que houve convergência nos algoritmos. Entretanto, o algoritmo de Dobra et al. apresentou um comportamento diferente que os outros.

Na Figura 25, observa-se um decaimento brusco em todas as linhas decrescentes dos gráficos. Pela Figura, temos que menos de 100 modelos foram visitados durante o processamento dos algoritmos. Além disso, o verdadeiro modelo 544 obteve baixo valor de probabilidade *a posteriori* em todos os algoritmo. Todavia, o algoritmo da proposta 3 obteve o maior valor de probabilidade *a posteriori* para o modelo 544, tendo em sequência o algoritmo da proposta 1. Com isso, o algoritmo da proposta 3 teve um desempenho levemente superior aos outros algoritmos.

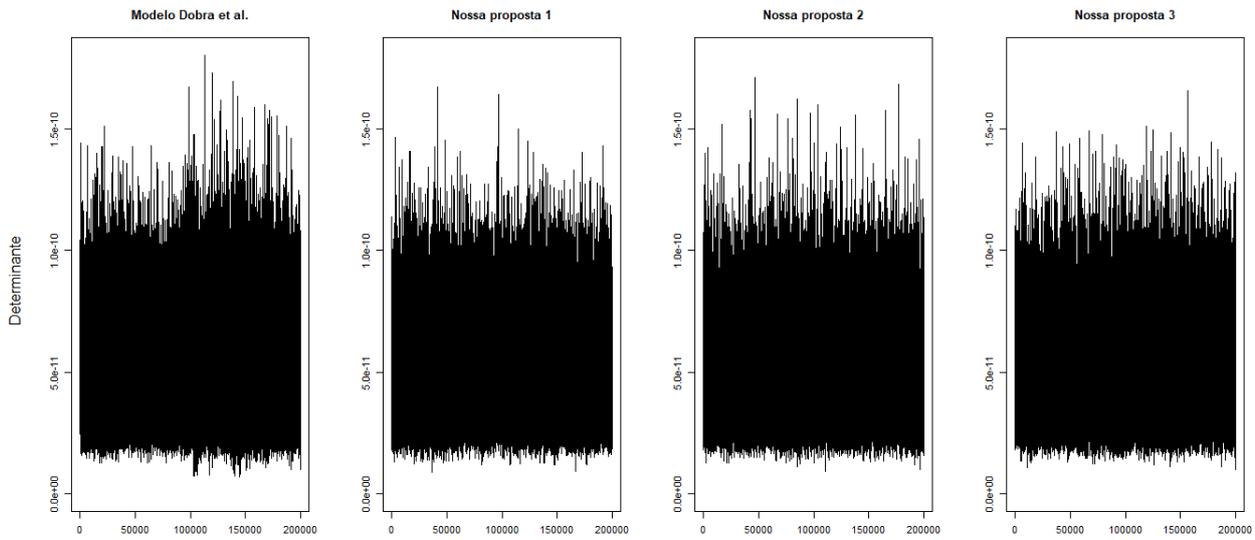


Figura 24 – Gráficos dos determinantes para avaliar a convergência dos algoritmos considerando $n = 100$, $m = 200.000$, warm-up = 100 e upgrade = 5. Fonte: Autor.

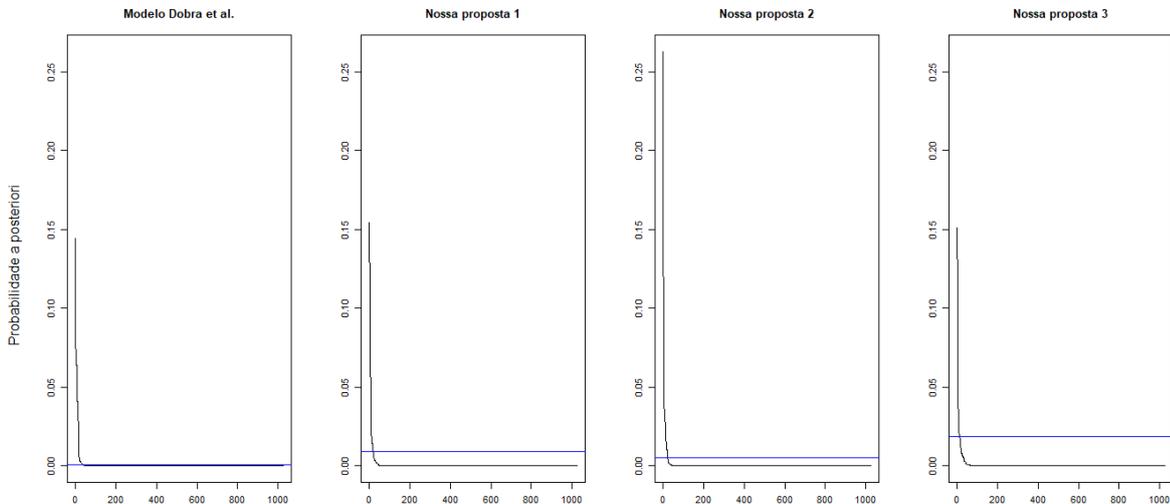


Figura 25 – Gráficos de avaliação das probabilidades *a posteriori* considerando $n = 100$, $m = 200.000$, warm-up = 100 e upgrade = 5. Fonte: Autor.

Na Tabela 37, observa-se que o algoritmo da proposta 1 obteve os melhores resultados de SAE, SAEIU, SQE e SQEIU em comparação as outras variações, pois teve os menores valores. Isso indica que o algoritmo da proposta 1 cometeu menos erros ao captar as presenças e ausências das arestas do verdadeiro modelo 544, em relação ao outros algoritmos. Considerando a Tabela 37 e as Figuras 24 e 25, pode-se concluir que o algoritmo da proposta 1 obteve o melhor desempenho para o cenário $m = 200.000$ e $n = 100$.

Tabela 37 – A soma absoluta dos erros (SAE), a soma absoluta dos erros no intervalo unitário (SAEIU), a soma quadrática dos erros (SQE) e a soma quadrática dos erros no intervalo unitário (SQEIU), para o cenário $m = 200.000$ e $n = 100$.

Algoritmo	SAE	SAEIU	SQE	SQEIU
Dobra et al.	2,913	0,291	1,849	0,185
Proposta 1	2,431	0,243	1,398	0,140
Proposta 2	2,879	0,288	1,905	0,190
Proposta 3	2,643	0,264	1,552	0,155

Fonte: Autor.

5.3.5 Resultados no cenário com $m = 200.000$ e $n = 250$

Nesta subseção, os resultados das simulações para o cenário $m = 200.000$ e $n = 250$ são apresentados e interpretados para as quatro variações do algoritmo *Reversible Jump*.

5.3.5.1 Algoritmo de Dobra et al. ($m = 200.000$ e $n = 250$)

Neste cenário, o modelo 32 foi o mais visitado, obtendo um total de 175.212 visitas. Entretanto, o modelo 544 obteve, somente, um total de 581 visitas. O $IS(32|544) = 9$ e $ISIU(32|544) = 0,90$. Com isso, tem-se que o modelo 32 possui uma semelhança de 90% com o modelo 544 que pode ser observada na Figura 26.

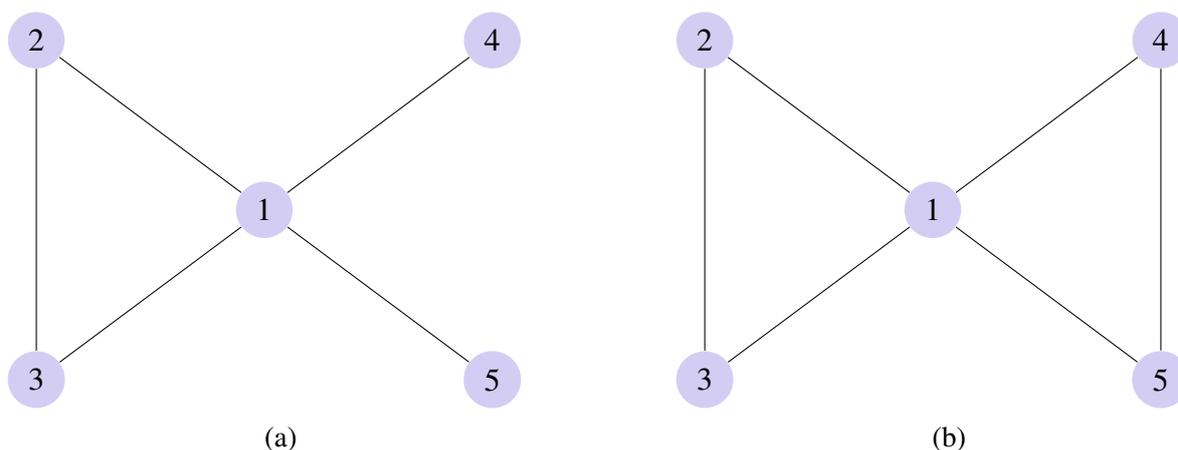


Figura 26 – (a) Grafo do modelo 32; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 38, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas $(1,2)$, $(1,3)$, $(1,4)$, $(1,5)$ e $(2,3)$. Com uma probabilidade baixa, ele visitou grafos que tinham as arestas $(3,4)$ e $(4,5)$. E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas $(2,4)$, $(2,5)$ e $(3,5)$.

Na Tabela 39, observa-se que o algoritmo acertou em captar a presença das arestas $(1,2)$, $(1,3)$, $(1,4)$, $(1,5)$ e $(2,3)$, e a ausência das arestas $(2,4)$ e $(2,5)$. O algoritmo também

Tabela 38 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 200.000$ e $n = 250$.

Vértices	1	2	3	4	5
1	-	200.000	200.000	200.000	200.000
2	1,000	-	200.000	129	117
3	1,000	1,000	-	23.437	14.579
4	1,000	0,001	0,117	-	24.091
5	1,000	0,001	0,073	0,120	-

Fonte: Autor.

conseguiu parcialmente captar a ausência das arestas (3,4) e (3,5). Todavia, o algoritmo errou em não captar a presença da aresta (4,5).

Tabela 39 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 200.000$ e $n = 250$.

Vértices	2	3	4	5
1	0,000	0,000	0,000	0,000
2	-	0,000	-0,001	-0,001
3	-	-	-0,117	-0,073
4	-	-	-	0,880

Fonte: Autor.

5.3.5.2 Algoritmo da Proposta 1 ($m = 200.000$ e $n = 250$)

Neste cenário, o modelo 672 foi o mais visitado, obtendo um total de 40.389 visitas. Entretanto, o modelo 544 obteve, somente, um total de 29.283 visitas. O $IS(672|544) = 9$ e $ISIU(672|544) = 0,90$. Com isso, tem-se que o modelo 672 possui uma semelhança de 90% com o modelo 544 que pode ser observada na Figura 27.

Na Tabela 40, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,3), (1,4), (1,5) e (2,3). Com uma moderada probabilidade, ele visitou grafos que tinham as arestas (3,4) e (4,5). Com uma probabilidade baixa, ele visitou grafos que tinham a aresta (3,5). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (2,4) e (2,5).

Na Tabela 41, observa-se que o algoritmo acertou em captar a presença das arestas (1,2), (1,3), (1,4), (1,5) e (2,3), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a ausência da aresta (3,5). Todavia, o algoritmo errou em não captar a presença da aresta (4,5) e a ausência da aresta (3,4).

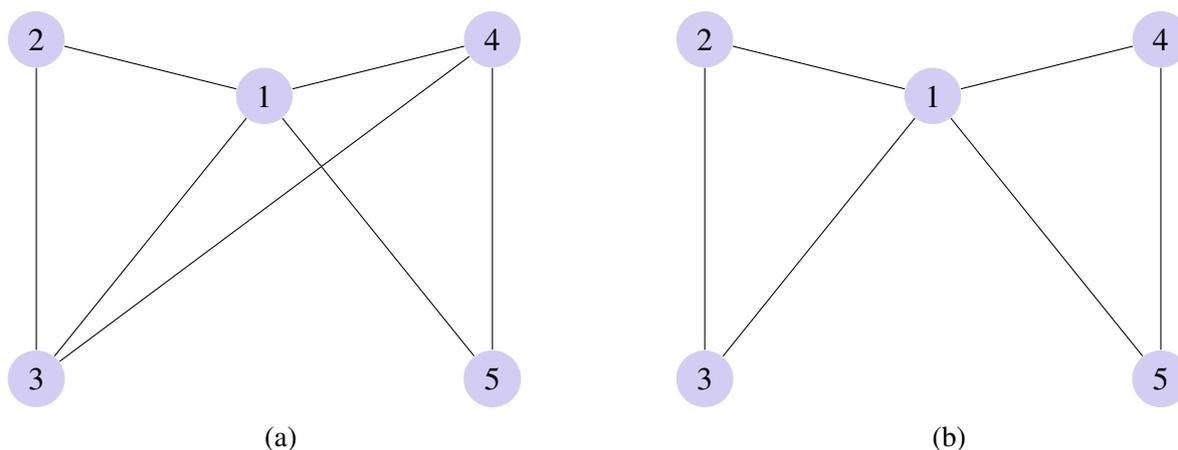


Figura 27 – (a) Grafo do modelo 672; (b) Grafo do modelo 544. Fonte: Autor.

Tabela 40 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 200.000$ e $n = 250$.

Vértices	1	2	3	4	5
1	-	200.000	200.000	200.000	200.000
2	1,000	-	200.000	46	61
3	1,000	1,000	-	112.763	57.635
4	1,000	0,000	0,564	-	97.436
5	1,000	0,000	0,288	0,487	-

Fonte: Autor.

Tabela 41 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 200.000$ e $n = 250$.

Vértices	2	3	4	5
1	0,000	0,000	0,000	0,000
2	-	0,000	0,000	0,000
3	-	-	-0,564	-0,288
4	-	-	-	0,513

Fonte: Autor.

5.3.5.3 Algoritmo da Proposta 2 ($m = 200.000$ e $n = 250$)

Neste cenário, o modelo 288 foi o mais visitado, obtendo um total de 50.438 visitas. Entretanto, o modelo 544 obteve, somente, um total de 35.926 visitas. O $IS(288|544) = 8$ e $ISIU(288|544) = 0,80$. Com isso, tem-se que o modelo 288 possui uma semelhança de 80% com o modelo 544 que pode ser observada na Figura 28.

Na Tabela 42, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,3), (1,4), (1,5) e (2,3). Com uma moderada probabilidade, ele visitou grafos que tinham as arestas (3,4) e (3,5). Com uma probabilidade

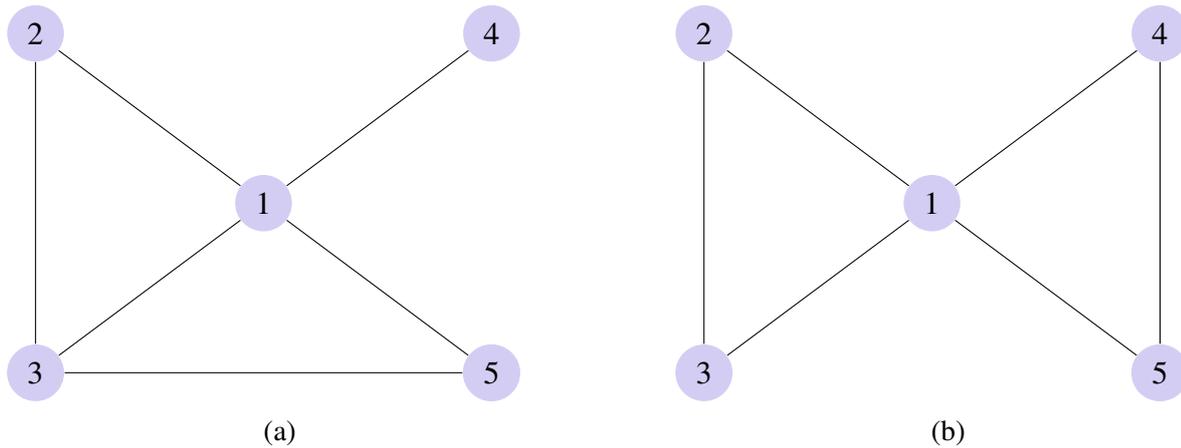


Figura 28 – (a) Grafo do modelo 288; (b) Grafo do modelo 544. Fonte: Autor.

baixa, ele visitou grafos que tinham a aresta (4,5). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (2,4) e (2,5).

Tabela 42 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 200.000$ e $n = 250$.

Vértices	1	2	3	4	5
1	-	200.000	200.000	200.000	200.000
2	1,000	-	200.000	528	162
3	1,000	1,000	-	68.899	101.094
4	1,000	0,003	0,344	-	37.397
5	1,000	0,001	0,505	0,187	-

Fonte: Autor.

Na Tabela 43, observa-se que o algoritmo acertou em captar a presença das arestas (1,2), (1,3), (1,4), (1,5) e (2,3), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a ausência da aresta (3,4). Todavia, o algoritmo errou em não captar a presença das arestas (4,5) e a ausência da aresta (3,5).

Tabela 43 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 200.000$ e $n = 250$.

Vértices	2	3	4	5
1	0,000	0,000	0,000	0,000
2	-	0,000	-0,003	-0,001
3	-	-	-0,344	-0,505
4	-	-	-	0,813

Fonte: Autor.

5.3.5.4 Algoritmo da Proposta 3 ($m = 200.000$ e $n = 250$)

Neste cenário, o modelo 32 foi o mais visitado, obtendo um total de 141.789 visitas. Entretanto, o modelo 544 obteve, somente, um total de 86 visitas. O $IS(32|544) = 9$ e $ISIU(32|544) = 0,90$. Com isso, tem-se que o modelo 32 possui uma semelhança de 90% com o modelo 544 que pode ser observada na Figura 29.

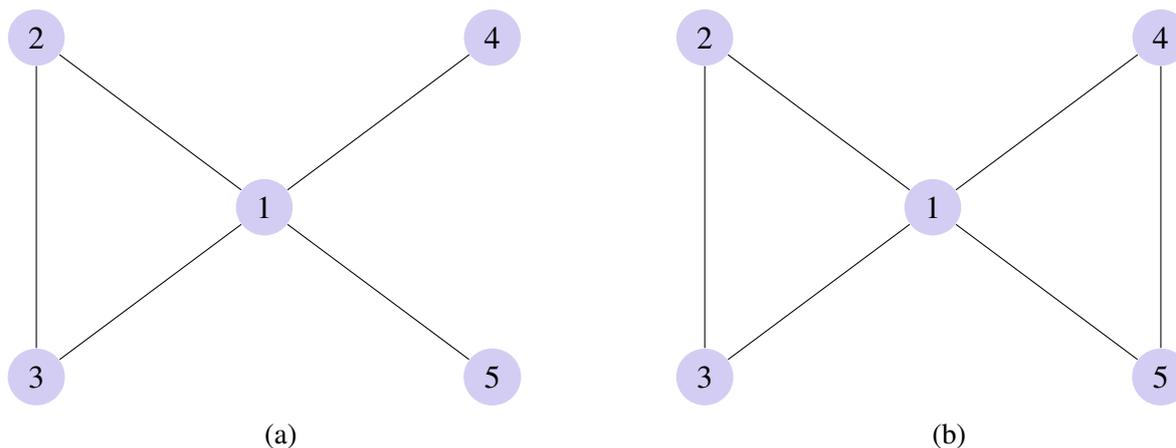


Figura 29 – (a) Grafo do modelo 32; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 44, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,3), (1,4), (1,5) e (2,3). Com uma probabilidade baixa, ele visitou grafos que tinham a aresta (3,5). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (2,4), (2,5), (3,4) e (4,5).

Tabela 44 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 200.000$ e $n = 250$.

Vértices	1	2	3	4	5
1	-	200.000	200.000	200.000	200.000
2	1,000	-	200.000	408	9
3	1,000	1,000	-	133	57.726
4	1,000	0,002	0,001	-	107
5	1,000	0,000	0,289	0,001	-

Fonte: Autor.

Na Tabela 45, observa-se que o algoritmo acertou em captar a presença das arestas (1,2), (1,3), (1,4), (1,5) e (2,3), e a ausência das arestas (2,4), (2,5) e (3,4). O algoritmo também conseguiu parcialmente captar a ausência da aresta (3,5). Todavia, o algoritmo errou em não captar a presença da aresta (4,5).

Tabela 45 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 200.000$ e $n = 250$.

Vértices	2	3	4	5
1	0,000	0,000	0,000	0,000
2	-	0,000	-0,002	0,000
3	-	-	-0,001	-0,289
4	-	-	-	0,999

Fonte: Autor.

5.3.5.5 Comparação dos Algoritmos ($m = 200.000$ e $n = 250$)

Na Figura 30, observa-se que não existem tendências nos gráficos de todas as variações do algoritmo *Reversible Jump*. Isso é um indício de que houve convergência nos algoritmos. Entretanto, o algoritmo da proposta 3 apresentou um comportamento diferente que os outros.

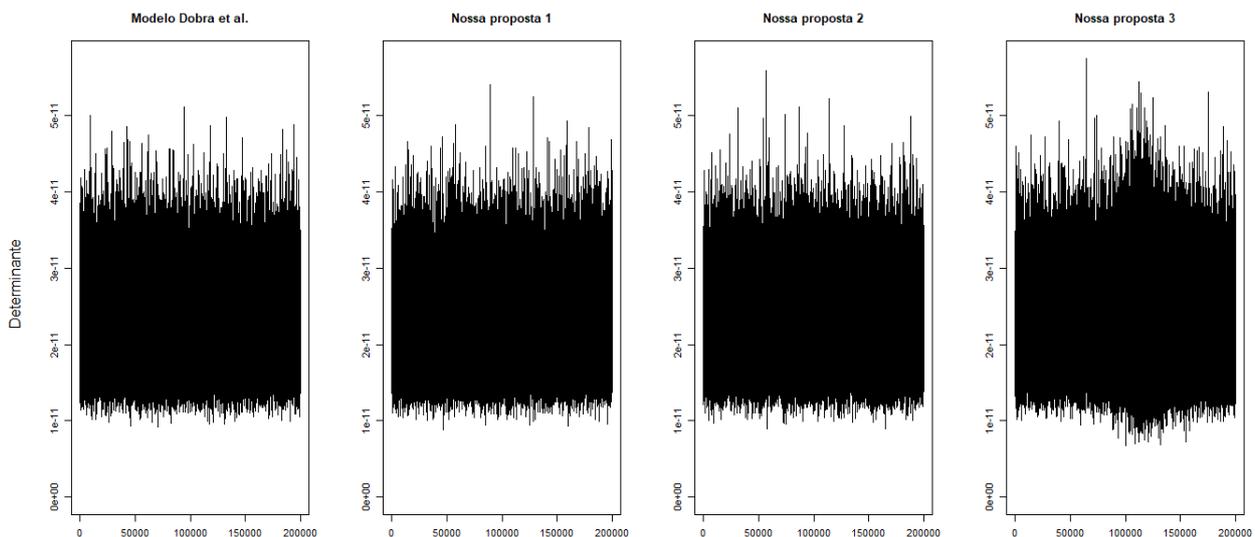


Figura 30 – Gráficos dos determinantes para avaliar a convergência dos algoritmos considerando $n = 250$, $m = 200.000$, warm-up = 100 e upgrade = 5. Fonte: Autor.

Na Figura 31, observa-se um decaimento brusco em todas as linhas decrescentes dos gráficos. Pela Figura, temos que menos de 100 modelos foram visitados durante o processamento dos algoritmos. Além disso, o verdadeiro modelo 544 obteve baixo valor de probabilidade *a posteriori* nos algoritmos de Dobra et al. e da proposta 3. Todavia, o algoritmo da proposta 2 obteve o maior valor de probabilidade *a posteriori* para o modelo 544, tendo em sequência o algoritmo da proposta 1. Com isso, o algoritmo da proposta 2 teve um desempenho levemente superior ao algoritmo da proposta 1, e obteve um desempenho superior aos outros dois algoritmos.

Na Tabela 46, observa-se que o algoritmo de Dobra et al. obteve os melhores resultados de SAE e SAEIU em comparação as outras variações, pois teve os menores valores. Entretanto, o algoritmo da proposta 1 obteve os melhores resultados de SQE e SQEIU. Isso indica que

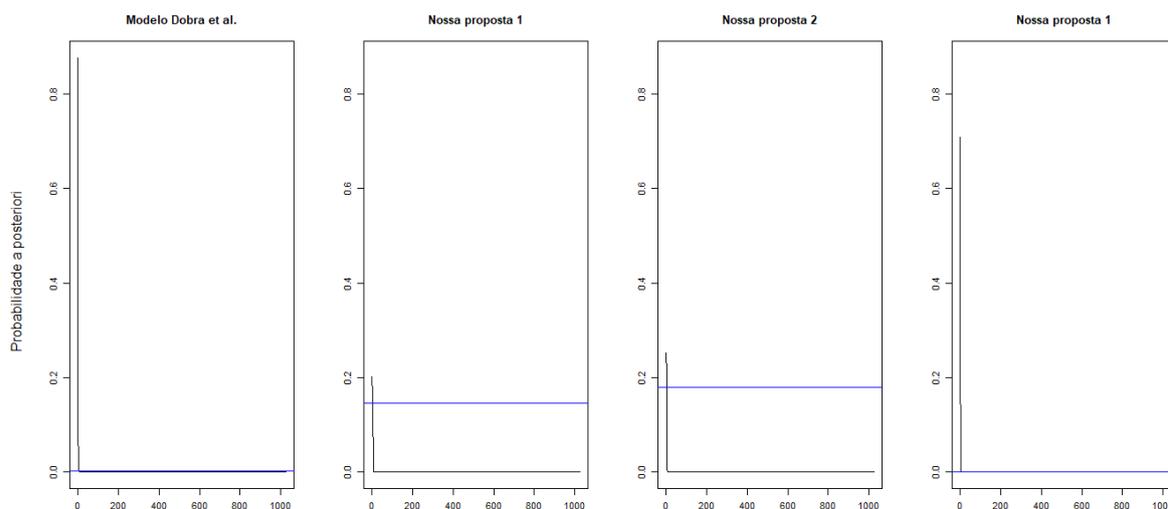


Figura 31 – Gráficos de avaliação das probabilidades *a posteriori* considerando $n = 250$, $m = 200.000$, $\text{warm-up} = 100$ e $\text{upgrade} = 5$. Fonte: Autor.

os algoritmos do Dobra et al. e da proposta 1 cometeu menos erros ao captar as presenças e ausências das arestas do verdadeiro modelo 544, em relação aos outros algoritmos. Considerando a Tabela 46 e as Figuras 30 e 31, pode-se concluir que o algoritmo da proposta 1 obteve o melhor desempenho para o cenário $m = 200.000$ e $n = 250$.

Tabela 46 – A soma absoluta dos erros (SAE), a soma absoluta dos erros no intervalo unitário (SAEIU), a soma quadrática dos erros (SQE) e a soma quadrática dos erros no intervalo unitário (SQEIU), para o cenário $m = 200.000$ e $n = 250$.

Algoritmo	SAE	SAEIU	SQE	SQEIU
Dobra et al.	1,071	0,107	0,793	0,079
Proposta 1	1,365	0,137	0,664	0,066
Proposta 2	1,666	0,167	1,035	0,104
Proposta 3	1,291	0,129	1,082	0,108

Fonte: Autor.

5.3.6 Resultados no cenário com $m = 200.000$ e $n = 500$

Nesta subseção, os resultados das simulações para o cenário $m = 200.000$ e $n = 500$ são apresentados e interpretados para as quatro variações do algoritmo *Reversible Jump*.

5.3.6.1 Algoritmo de Dobra et al. ($m = 200.000$ e $n = 500$)

Neste cenário, o modelo 928 foi o mais visitado, obtendo um total de 62.100 visitas. Entretanto, o modelo 544 obteve, somente, um total de 33.127 visitas. O $\text{IS}(928|544) = 8$ e

$ISIU(928|544) = 0,80$. Com isso, tem-se que o modelo 928 possui uma semelhança de 80% com o modelo 544 que pode ser observada na Figura 32.

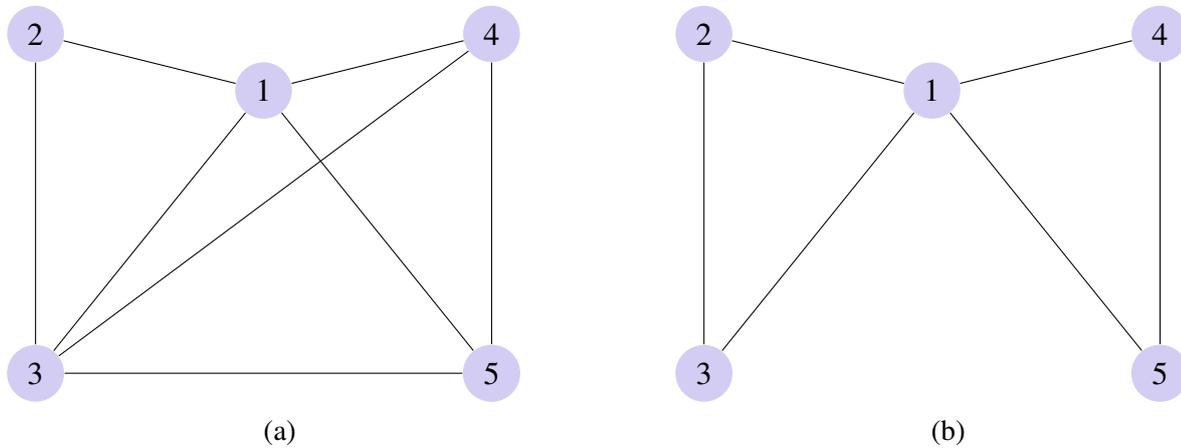


Figura 32 – (a) Grafo do modelo 928; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 47, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,3), (1,4), (1,5), (2,3) e (4,5). Com uma moderada probabilidade, ele visitou grafos que tinham as arestas (3,4) e (3,5). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (2,4) e (2,5).

Tabela 47 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 200.000$ e $n = 500$.

Vértices	1	2	3	4	5
1	-	200.000	200.000	200.000	200.000
2	1,000	-	200.000	68	19
3	1,000	1,000	-	125.483	103.626
4	1,000	0,000	0,627	-	173.596
5	1,000	0,000	0,518	0,868	-

Fonte: Autor.

Na Tabela 48, observa-se que o algoritmo acertou em captar a presença das arestas (1,2), (1,3), (1,4), (1,5) e (2,3), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a presença da aresta (4,5). Todavia, o algoritmo errou em não captar a ausência das arestas (3,4) e (3,5).

Tabela 48 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 200.000$ e $n = 500$.

Vértices	2	3	4	5
1	0,000	0,000	0,000	0,000
2	-	0,000	0,000	0,000
3	-	-	-0,627	-0,518
4	-	-	-	0,132

Fonte: Autor.

5.3.6.2 Algoritmo da Proposta 1 ($m = 200.000$ e $n = 500$)

Neste cenário, o modelo 160 foi o mais visitado, obtendo um total de 56.643 visitas. Entretanto, o modelo 544 obteve, somente, um total de 29.611 visitas. O $IS(160|544) = 8$ e $ISIU(160|544) = 0,80$. Com isso, tem-se que o modelo 160 possui uma semelhança de 80% com o modelo 544 que pode ser observada na Figura 33.

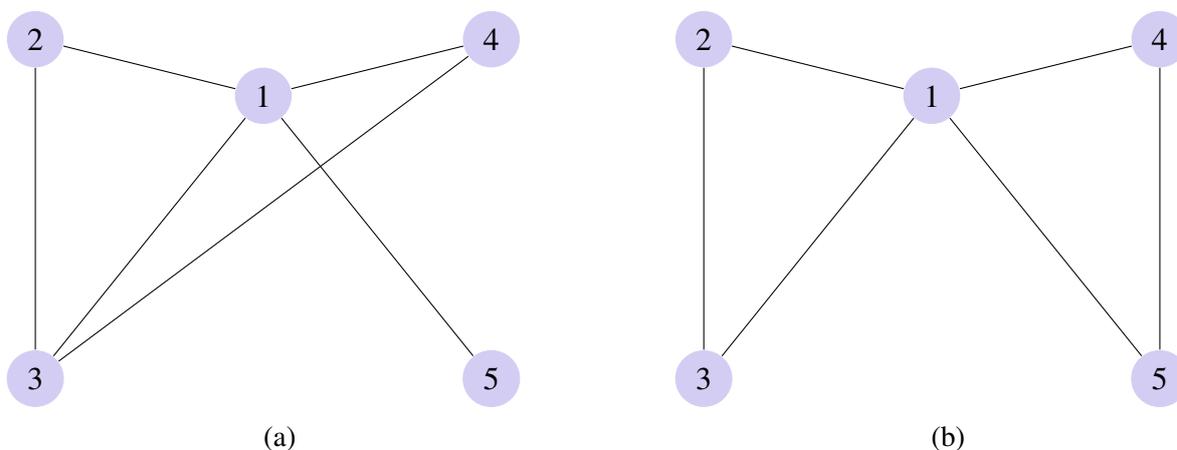


Figura 33 – (a) Grafo do modelo 160; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 49, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,3), (1,4), (1,5) e (2,3). Com uma moderada probabilidade, ele visitou grafos que tinham a aresta (3,4). Com uma probabilidade baixa, ele visitou grafos que tinham as arestas (3,5) e (4,5). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (2,4) e (2,5).

Na Tabela 50, observa-se que o algoritmo acertou em captar a presença das arestas (1,2), (1,3), (1,4), (1,5) e (2,3), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a ausência da aresta (3,5). Todavia, o algoritmo errou em não captar a presença da aresta (4,5) e a ausência da aresta (3,4).

Tabela 49 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 200.000$ e $n = 500$.

Vértices	1	2	3	4	5
1	-	200.000	200.000	200.000	200.000
2	1,000	-	200.000	13	75
3	1,000	1,000	-	93.877	56.125
4	1,000	0,000	0,469	-	40.685
5	1,000	0,000	0,281	0,203	-

Fonte: Autor.

Tabela 50 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 200.000$ e $n = 500$.

Vértices	2	3	4	5
1	0,000	0,000	0,000	0,000
2	-	0,000	0,000	0,000
3	-	-	-0,469	-0,281
4	-	-	-	0,797

Fonte: Autor.

5.3.6.3 Algoritmo da Proposta 2 ($m = 200.000$ e $n = 500$)

Neste cenário, o modelo 32 foi o mais visitado, obtendo um total de 105.342 visitas. Entretanto, o modelo 544 obteve, somente, um total de 161 visitas. O $IS(32|544) = 9$ e $ISIU(32|544) = 0,90$. Com isso, tem-se que o modelo 32 possui uma semelhança de 90% com o modelo 544 que pode ser observada na Figura 34.

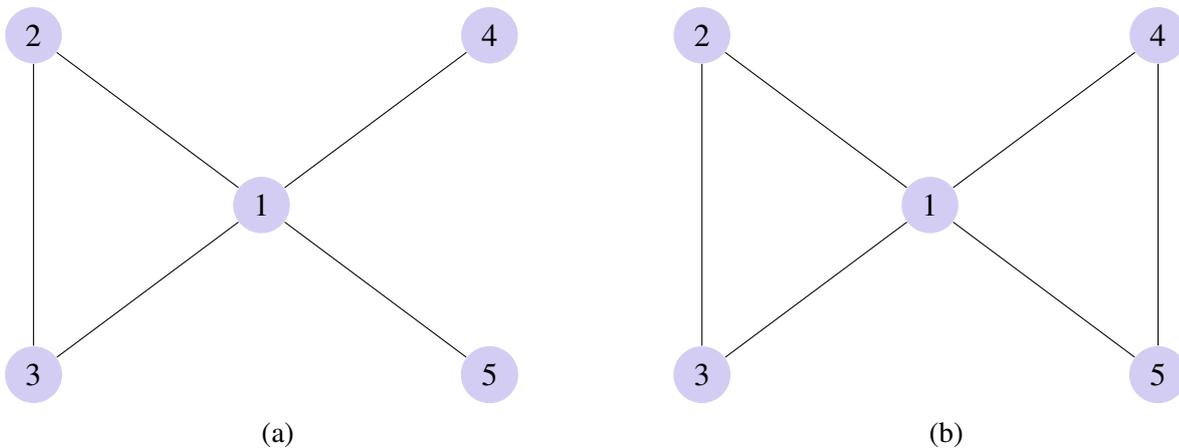


Figura 34 – (a) Grafo do modelo 32; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 51, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,3), (1,4), (1,5) e (2,3). Com uma moderada probabilidade, ele visitou grafos que tinham a aresta (3,5). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (2,4), (2,5), (3,4) e (4,5).

Tabela 51 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 200.000$ e $n = 500$.

Vértices	1	2	3	4	5
1	-	200.000	200.000	200.000	200.000
2	1,000	-	200.000	355	118
3	1,000	1,000	-	4.667	91.892
4	1,000	0,002	0,023	-	1.673
5	1,000	0,001	0,459	0,008	-

Fonte: Autor.

Na Tabela 52, observa-se que o algoritmo acertou em captar a presença das arestas (1,2), (1,3), (1,4), (1,5) e (2,3), e a ausência das arestas (2,4), (2,5) e (3,4). O algoritmo também conseguiu parcialmente captar a ausência da aresta (3,5). Todavia, o algoritmo errou em não captar a presença da aresta (4,5).

Tabela 52 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 200.000$ e $n = 500$.

Vértices	2	3	4	5
1	0,000	0,000	0,000	0,000
2	-	0,000	-0,002	-0,001
3	-	-	-0,023	-0,459
4	-	-	-	0,992

Fonte: Autor.

5.3.6.4 Algoritmo da Proposta 3 ($m = 200.000$ e $n = 500$)

Neste cenário, o modelo 32 foi o mais visitado, obtendo um total de 88.290 visitas. Entretanto, o modelo 544 obteve, somente, um total de 80.521 visitas. O $IS(32|544) = 9$ e $ISIU(32|544) = 0,90$. Com isso, tem-se que o modelo 32 possui uma semelhança de 90% com o modelo 544 que pode ser observada na Figura 35.

Na Tabela 53, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,3), (1,4), (1,5) e (2,3). Com uma moderada probabilidade, ele visitou grafos que tinham a aresta (4,5). Com uma probabilidade baixa, ele visitou grafos que tinham a aresta (3,5). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (2,4), (2,5) e (3,4).

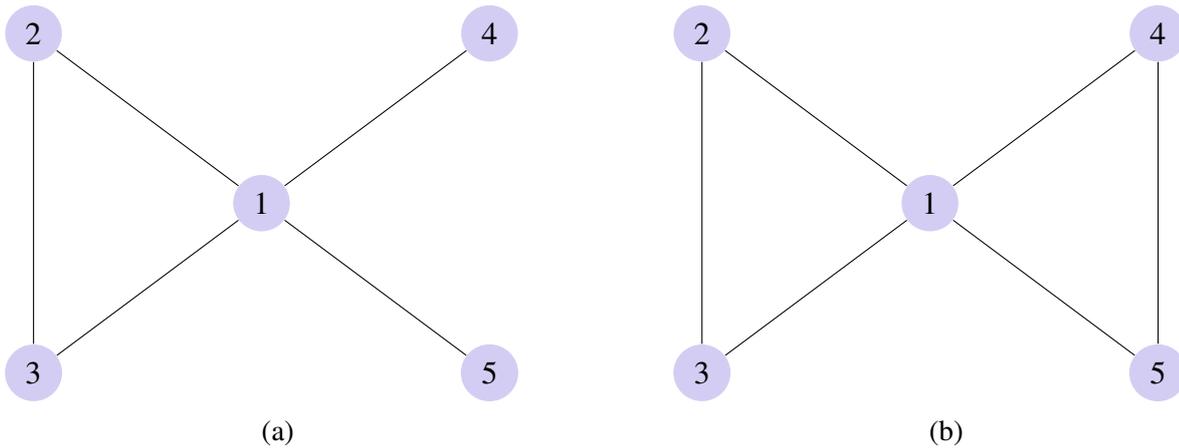


Figura 35 – (a) Grafo do modelo 32; (b) Grafo do modelo 544. Fonte: Autor.

Tabela 53 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 200.000$ e $n = 500$.

Vértices	1	2	3	4	5
1	-	200.000	200.000	200.000	200.000
2	1,000	-	200.000	357	103
3	1,000	1,000	-	484	30.422
4	1,000	0,002	0,002	-	96.987
5	1,000	0,001	0,152	0,485	-

Fonte: Autor.

Na Tabela 54, observa-se que o algoritmo acertou em captar a presença das arestas (1, 2), (1, 3), (1, 4), (1, 5) e (2, 3), e a ausência das arestas (2, 4), (2, 5) e (3, 4). O algoritmo também conseguiu parcialmente captar a ausência da aresta (3, 5). Todavia, o algoritmo errou em não captar a presença da aresta (4, 5).

Tabela 54 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 200.000$ e $n = 500$.

Vértices	2	3	4	5
1	0,000	0,000	0,000	0,000
2	-	0,000	-0,002	-0,001
3	-	-	-0,002	-0,152
4	-	-	-	0,515

Fonte: Autor.

5.3.6.5 Comparação dos Algoritmos ($m = 200.000$ e $n = 500$)

Na Figura 36, observa-se que não existem tendências nos gráficos de todas as variações do algoritmo *Reversible Jump*. Isso é um indício de que houve convergência nos algoritmos.

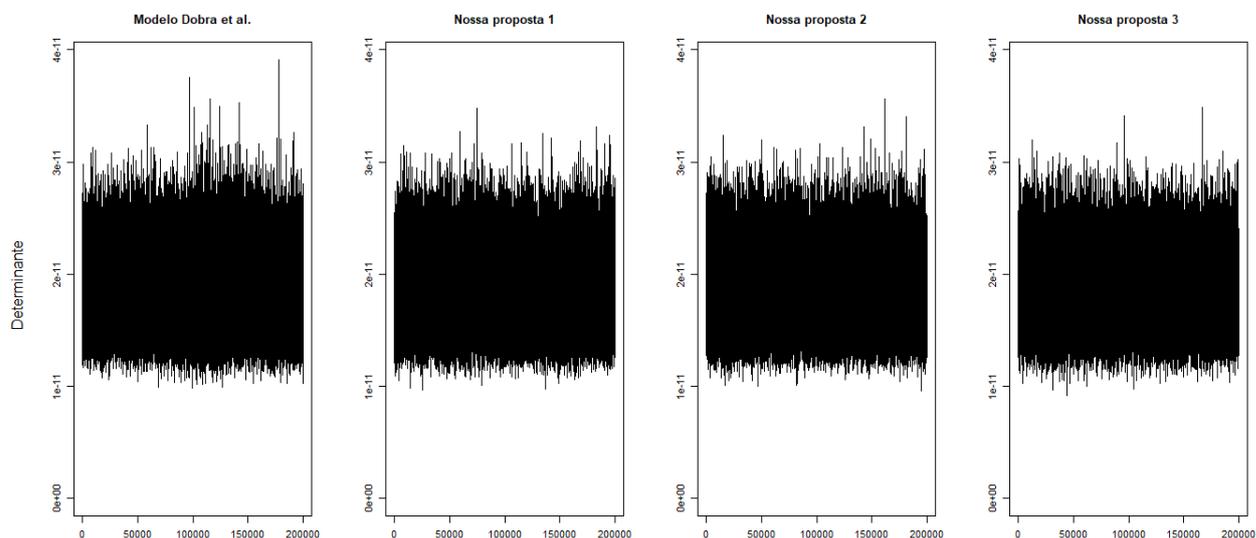


Figura 36 – Gráficos dos determinantes para avaliar a convergência dos algoritmos considerando $n = 500$, $m = 200.000$, warm-up = 100 e upgrade = 5. Fonte: Autor.

Na Figura 37, observa-se um decaimento brusco em todas as linhas decrescentes dos gráficos. Pela Figura, temos que menos de 100 modelos foram visitados durante o processamento dos algoritmos. Além disso, o verdadeiro modelo 544 obteve baixo valor de probabilidade *a posteriori* no algoritmo da proposta 2. Todavia, o algoritmo da proposta 3 obteve o maior valor de probabilidade *a posteriori* para o modelo 544, tendo em sequência os algoritmos do Dobra et al. e da proposta 1. Com isso, o algoritmo da proposta 3 teve um desempenho superior aos outros algoritmos.

Na Tabela 55, observa-se que o algoritmo da proposta 3 obteve os melhores resultados de SAE, SAEIU, SQE e SQEIU em comparação as outras variações, pois teve os menores valores. Isso indica que o algoritmo da proposta 3 cometeu menos erros ao captar as presenças e ausências das arestas do verdadeiro modelo 544, em relação aos outros algoritmos. Considerando a Tabela 55 e as Figuras 36 e 37, pode-se concluir que o algoritmo da proposta 3 obteve o melhor desempenho para o cenário $m = 200.000$ e $n = 250$.

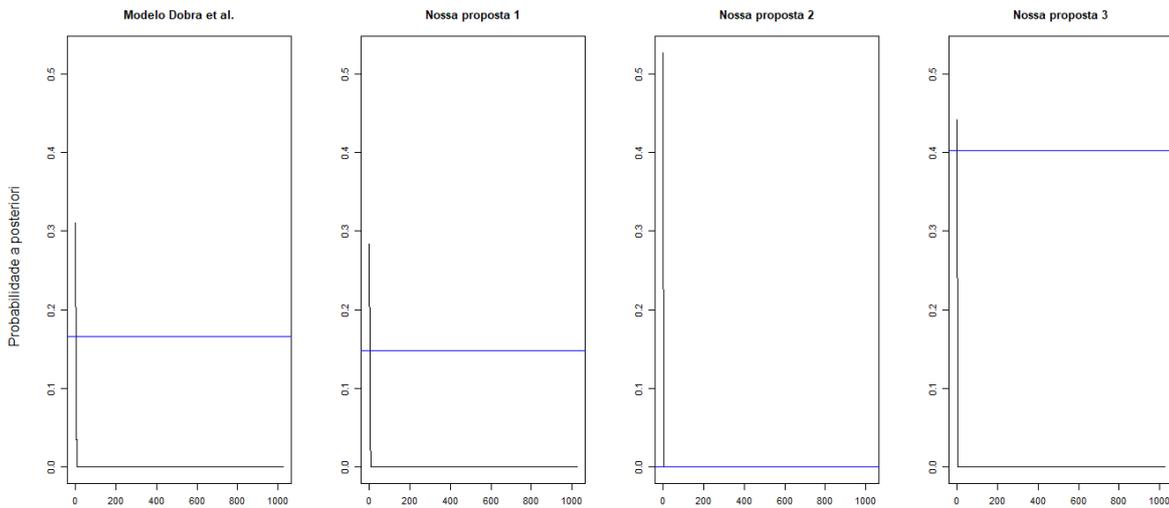


Figura 37 – Gráficos de avaliação das probabilidades *a posteriori* considerando $n = 500$, $m = 200.000$, $\text{warm-up} = 100$ e $\text{upgrade} = 5$. Fonte: Autor.

Tabela 55 – A soma absoluta dos erros (SAE), a soma absoluta dos erros no intervalo unitário (SAEIU), a soma quadrática dos erros (SQE) e a soma quadrática dos erros no intervalo unitário (SQEIU), para o cenário $m = 200.000$ e $n = 500$.

Algoritmo	SAE	SAEIU	SQE	SQEIU
Dobra et al.	1,278	0,128	0,680	0,068
Proposta 1	1,547	0,155	0,934	0,093
Proposta 2	1,477	0,148	1,195	0,119
Proposta 3	0,672	0,067	0,288	0,029

Fonte: Autor.

5.3.7 Resultados no cenário com $m = 300.000$ e $n = 100$

Nesta subseção, os resultados das simulações para o cenário $m = 300.000$ e $n = 100$ são apresentados e interpretados para as quatro variações do algoritmo *Reversible Jump*.

5.3.7.1 Algoritmo de Dobra et al. ($m = 300.000$ e $n = 100$)

Neste cenário, o modelo 525 foi o mais visitado, obtendo um total de 51.852 visitas. Entretanto, o modelo 544 obteve, somente, um total de 172 visitas. O $IS(525|544) = 7$ e $ISIU(525|544) = 0,70$. Com isso, tem-se que o modelo 525 possui uma semelhança de 70% com o modelo 544 que pode ser observada na Figura 38.

Na Tabela 56, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,4), (1,5) e (4,5). Com uma moderada probabilidade, ele visitou grafos que tinham as arestas (1,2), (2,3), (3,4) e (3,5). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (1,3), (2,4) e (2,5).

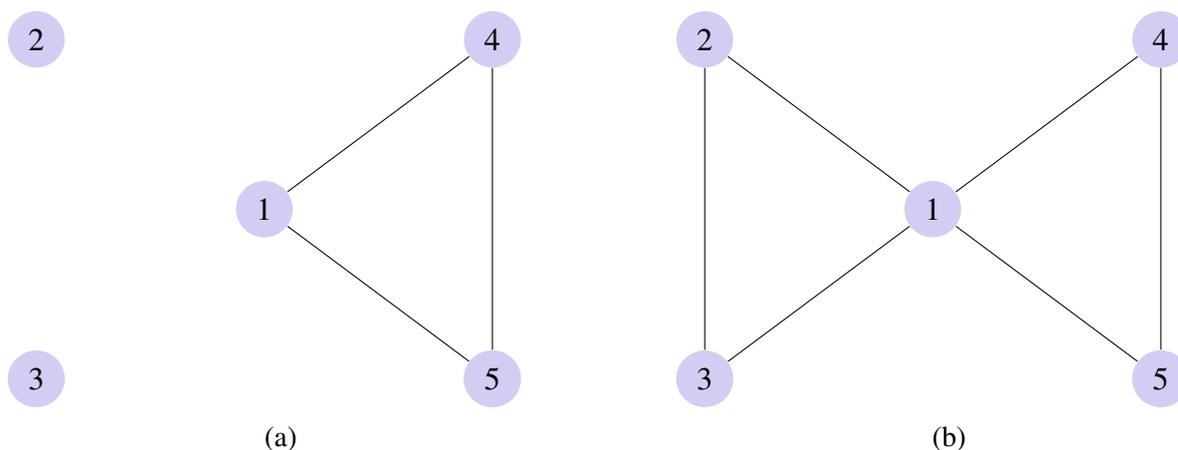


Figura 38 – (a) Grafo do modelo 525; (b) Grafo do modelo 544. Fonte: Autor.

Tabela 56 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 300.000$ e $n = 100$.

Vértices	1	2	3	4	5
1	-	125.749	111.976	300.000	300.000
2	0,419	-	118.304	252	253
3	0,040	0,394	-	131.973	95.962
4	1,000	0,001	0,440	-	285.780
5	1,000	0,001	0,320	0,953	-

Fonte: Autor.

Na Tabela 57, observa-se que o algoritmo acertou em captar a presença das arestas (1,4), (1,5) e (4,5), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a presença das arestas (1,2) e (2,3), e a ausência das arestas (3,4) e (3,5). Todavia, o algoritmo errou em não captar a presença da aresta (1,3).

Tabela 57 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 300.000$ e $n = 100$.

Vértices	2	3	4	5
1	0,581	0,960	0,000	0,000
2	-	0,606	-0,001	-0,001
3	-	-	-0,440	-0,320
4	-	-	-	0,047

Fonte: Autor.

5.3.7.2 Algoritmo da Proposta 1 ($m = 300.000$ e $n = 100$)

Neste cenário, o modelo 541 foi o mais visitado, obtendo um total de 45.334 visitas. Entretanto, o modelo 544 obteve, somente, um total de 1.851 visitas. O $IS(525|544) = 8$ e $ISIU(541|544) = 0,80$. Com isso, tem-se que o modelo 541 possui uma semelhança de 80% com o modelo 544 que pode ser observada na Figura 39.

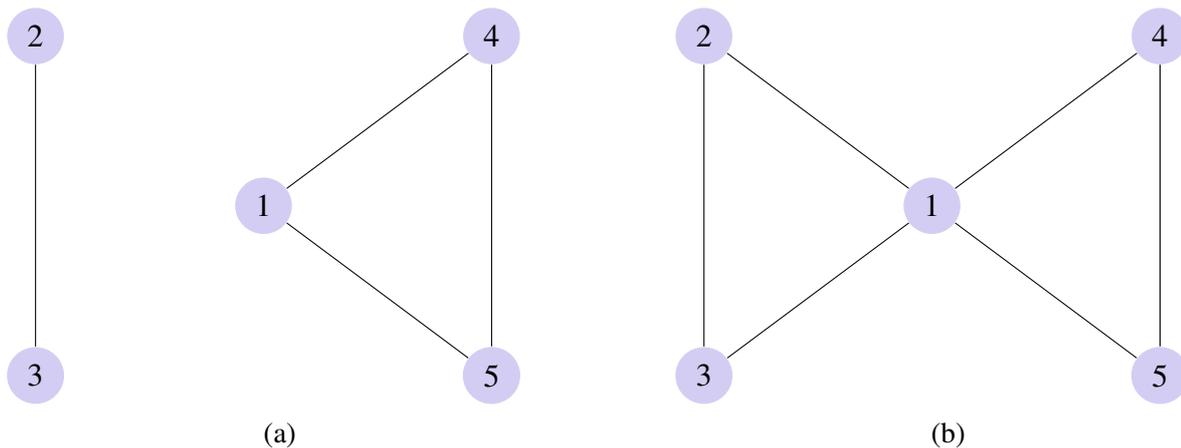


Figura 39 – (a) Grafo do modelo 541; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 58, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,4), (1,5) e (2,3). Com uma moderada probabilidade, ele visitou grafos que tinham as arestas (1,2) e (4,5). Com uma probabilidade baixa, ele visitou grafos que tinham as arestas (3,4) e (3,5). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (1,3), (2,4) e (2,5).

Tabela 58 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 300.000$ e $n = 100$.

Vértices	1	2	3	4	5
1	-	125.482	20.816	300.000	300.000
2	0,418	-	216.650	163	85
3	0,069	0,722	-	80.650	55.543
4	1,000	0,001	0,269	-	147.903
5	1,000	0,000	0,185	0,493	-

Fonte: Autor.

Na Tabela 59, observa-se que o algoritmo acertou em captar a presença das arestas (1,4) e (1,5), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a presença das arestas (1,2), (2,3) e (4,5), e a ausência das arestas (3,4) e (3,5). Todavia, o algoritmo errou em não captar a presença da aresta (1,3).

Tabela 59 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 300.000$ e $n = 100$.

Vértices	2	3	4	5
1	0,582	0,931	0,000	0,000
2	-	0,278	-0,001	0,000
3	-	-	-0,269	-0,185
4	-	-	-	0,507

Fonte: Autor.

5.3.7.3 Algoritmo da Proposta 2 ($m = 300.000$ e $n = 100$)

Neste cenário, o modelo 286 foi o mais visitado, obtendo um total de 88.611 visitas. Entretanto, o modelo 544 obteve, somente, um total de 1.097 visitas. O $IS(286|544) = 7$ e $ISIU(286|544) = 0,70$. Com isso, tem-se que o modelo 286 possui uma semelhança de 70% com o modelo 544 que pode ser observada na Figura 40.

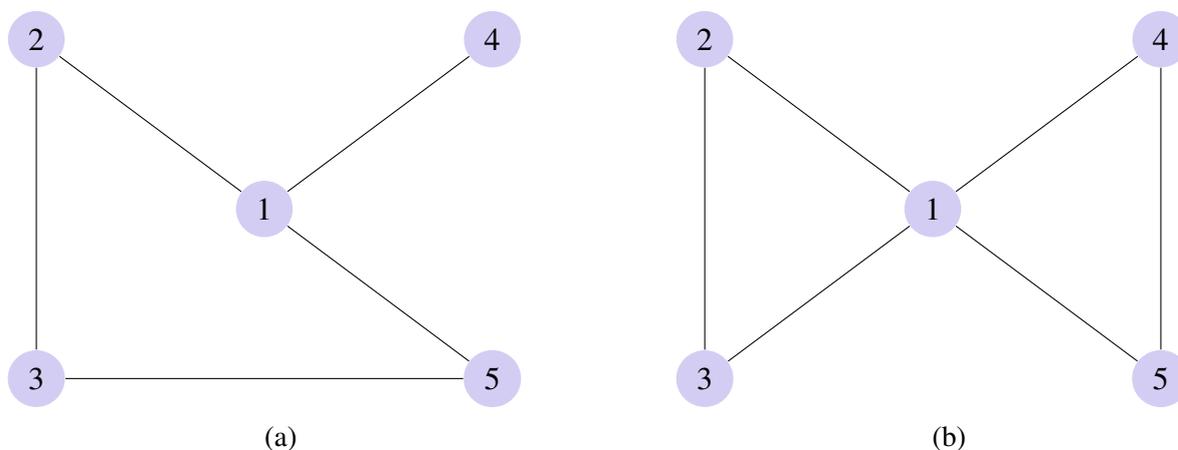


Figura 40 – (a) Grafo do modelo 286; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 60, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,4), (1,5) e (2,3). Com uma moderada probabilidade, ele visitou grafos que tinham a aresta (3,5). Com uma probabilidade baixa, ele visitou grafos que tinham as arestas (1,3) e (3,4). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (2,4), (2,5) e (4,5).

Na Tabela 61, observa-se que o algoritmo acertou em captar a presença das arestas (1,4) e (1,5), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a presença das arestas (1,2) e (2,3), e a ausência das arestas (3,4) e (3,5). Todavia, o algoritmo errou em não captar a presença das arestas (1,3) e (4,5).

Tabela 60 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 300.000$ e $n = 100$.

Vértices	1	2	3	4	5
1	-	233.219	68.833	299.885	300.000
2	0,777	-	255.130	1.219	173
3	0,229	0,850	-	51.975	175.208
4	1,000	0,004	0,173	-	17.948
5	1,000	0,001	0,584	0,060	-

Fonte: Autor.

Tabela 61 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 300.000$ e $n = 100$.

Vértices	2	3	4	5
1	0,223	0,771	0,000	0,000
2	-	0,150	-0,004	-0,001
3	-	-	-0,173	-0,584
4	-	-	-	0,940

Fonte: Autor.

5.3.7.4 Algoritmo da Proposta 3 ($m = 300.000$ e $n = 100$)

Neste cenário, o modelo 286 foi o mais visitado, obtendo um total de 60.319 visitas. Entretanto, o modelo 544 obteve, somente, um total de 3.700 visitas. O $IS(286|544) = 7$ e $ISIU(286|544) = 0,70$. Com isso, tem-se que o modelo 286 possui uma semelhança de 70% com o modelo 544 que pode ser observada na Figura 41.

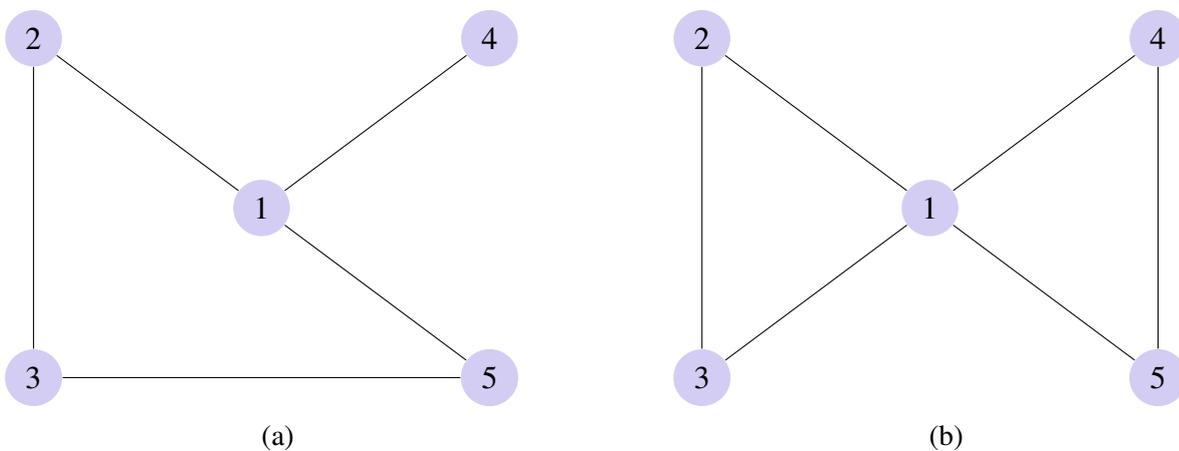


Figura 41 – (a) Grafo do modelo 286; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 62, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,4), (1,5) e (2,3). Com uma moderada probabilidade, ele visitou grafos que tinham a aresta (3,5). Com uma probabilidade baixa, ele visitou grafos que tinham as arestas (1,3) e (3,4). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (2,4), (2,5) e (4,5).

Tabela 62 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 300.000$ e $n = 100$.

Vértices	1	2	3	4	5
1	-	251.603	96.714	300.000	285.158
2	0,839	-	245.820	1.245	63
3	0,322	0,819	-	38.179	136.965
4	1,000	0,004	0,127	-	21.480
5	0,951	0,000	0,457	0,072	-

Fonte: Autor.

Na Tabela 63, observa-se que o algoritmo acertou em captar a presença das arestas (1,4) e (1,5), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a presença das arestas (1,2) e (2,3), e a ausência das arestas (3,4) e (3,5). Todavia, o algoritmo errou em não captar a presença das arestas (1,3) e (4,5).

Tabela 63 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 300.000$ e $n = 100$.

Vértices	2	3	4	5
1	0,161	0,678	0,000	0,049
2	-	0,181	-0,004	0,000
3	-	-	-0,127	-0,457
4	-	-	-	0,928

Fonte: Autor.

5.3.7.5 Comparação dos Algoritmos ($m = 300.000$ e $n = 100$)

Na Figura 42, observa-se que não existem tendências nos gráficos de todas as variações do algoritmo *Reversible Jump*. Isso é um indício de que houve convergência nos algoritmos.

Na Figura 43, observa-se que um decaimento brusco em todas as linhas decrescentes dos gráficos. Isso indica que menos de 100 modelos foram visitados durante o processamento dos algoritmos. Além disso, o verdadeiro modelo 544 obteve baixos valores de probabilidade *a posteriori* em todas as variações. Entretanto, o algoritmo da proposta 3 teve um desempenho levemente superior aos outros algoritmos.

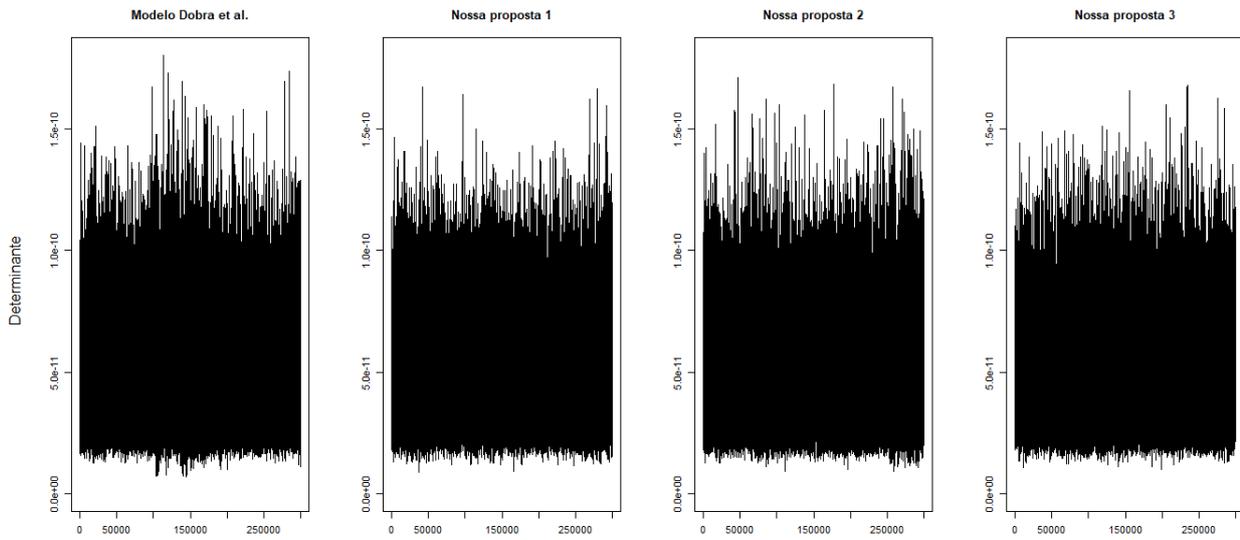


Figura 42 – Gráficos dos determinantes para avaliar a convergência dos algoritmos considerando $n = 100$, $m = 300.000$, warm-up = 100 e upgrade = 5. Fonte: Autor.

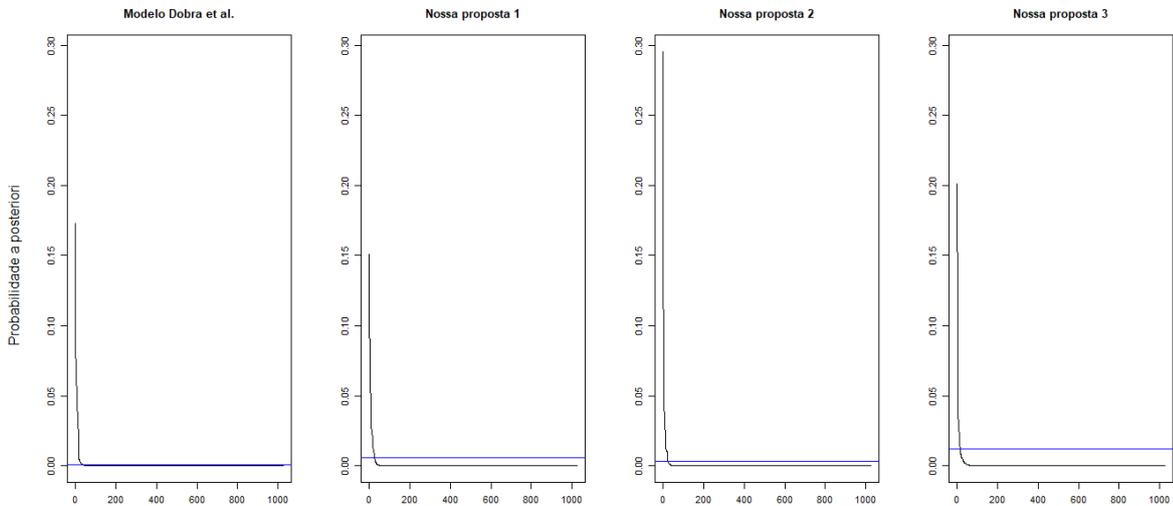


Figura 43 – Gráficos de avaliação das probabilidades *a posteriori* considerando $n = 100$, $m = 300.000$, warm-up = 100 e upgrade = 5. Fonte: Autor.

Na Tabela 64, observa-se que o algoritmo da proposta 3 obteve os melhores resultados em comparação com as outras variações, pois teve os menores valores em todas as medidas SAE, SAEIU, SQE e SQEIU. Isso indica que o algoritmo da proposta 3 cometeu menos erros ao capturar as presenças e ausências das arestas do verdadeiro modelo 544, em relação aos outros algoritmos. Considerando a Tabela 64 e as Figuras 42 e 43, pode-se concluir que o algoritmo da proposta 3 obteve o melhor desempenho para o cenário $m = 300.000$ e $n = 100$.

Tabela 64 – A soma absoluta dos erros (SAE), a soma absoluta dos erros no intervalo unitário (SAEIU), a soma quadrática dos erros (SQE) e a soma quadrática dos erros no intervalo unitário (SQEIU), para o cenário $m = 300.000$ e $n = 100$.

ALgoritmo	SAE	SAEIU	SQE	SQEIU
Dobra et al.	2,955	0,296	1,924	0,192
Proposta 1	2,752	0,275	1,645	0,165
Proposta 2	2,845	0,285	1,921	0,192
Proposta 3	2,586	0,259	1,607	0,161

Fonte: Autor.

5.3.8 Resultados no cenário com $m = 300.000$ e $n = 250$

Nesta subseção, os resultados das simulações para o cenário $m = 300.000$ e $n = 250$ são apresentados e interpretados para as quatro variações do algoritmo *Reversible Jump*.

5.3.8.1 Algoritmo de Dobra et al. ($m = 300.000$ e $n = 250$)

Neste cenário, o modelo 32 foi o mais visitado, obtendo um total de 262.797 visitas. Entretanto, o modelo 544 obteve, somente, um total de 4.384 visitas. O $I(32|544) = 9$ e $ISIU(32|544) = 0,90$. Com isso, tem-se que o modelo 32 possui uma semelhança de 90% com o modelo 544 que pode ser observada na Figura 44.

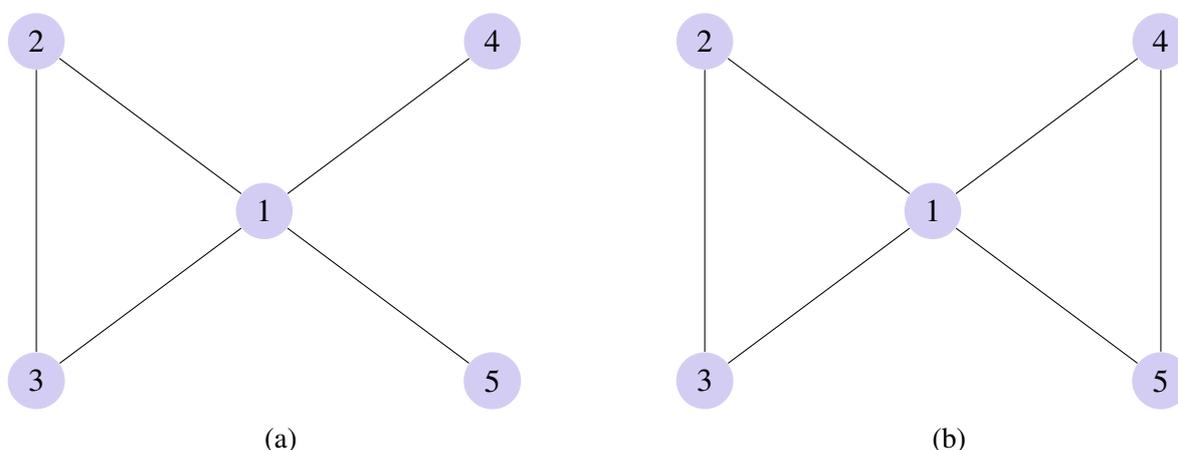


Figura 44 – (a) Grafo do modelo 32; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 65, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,3), (1,4), (1,5) e (2,3). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (2,4), (2,5), (3,4), (3,5) e (4,5).

Na Tabela 66, observa-se que o algoritmo acertou em captar a presença das arestas (1,2), (1,3), (1,4), (1,5) e (2,3), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a ausência das arestas (3,4) e (3,5). Todavia, o algoritmo errou em não captar a presença da aresta (4,5).

Tabela 65 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 300.000$ e $n = 250$.

Vértices	1	2	3	4	5
1	-	300.000	300.000	300.000	300.000
2	1,000	-	300.000	133	119
3	1,000	1,000	-	31.609	18.855
4	1,000	0,000	0,105	-	29.424
5	1,000	0,000	0,063	0,098	-

Fonte: Autor.

Tabela 66 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 300.000$ e $n = 250$.

Vértices	2	3	4	5
1	0,000	0,000	0,000	0,000
2	-	0,000	0,000	0,000
3	-	-	-0,105	-0,063
4	-	-	-	0,902

Fonte: Autor.

5.3.8.2 Algoritmo da Proposta 1 ($m = 300.000$ e $n = 250$)

Neste cenário, o modelo 672 foi o mais visitado, obtendo um total de 104.660 visitas. Entretanto, o modelo 544 obteve, somente, um total de 31.215 visitas. O $IS(672|544) = 9$ e $ISIU(672|544) = 0,90$. Com isso, tem-se que o modelo 672 possui uma semelhança de 90% com o modelo 544 que pode ser observada na Figura 45.

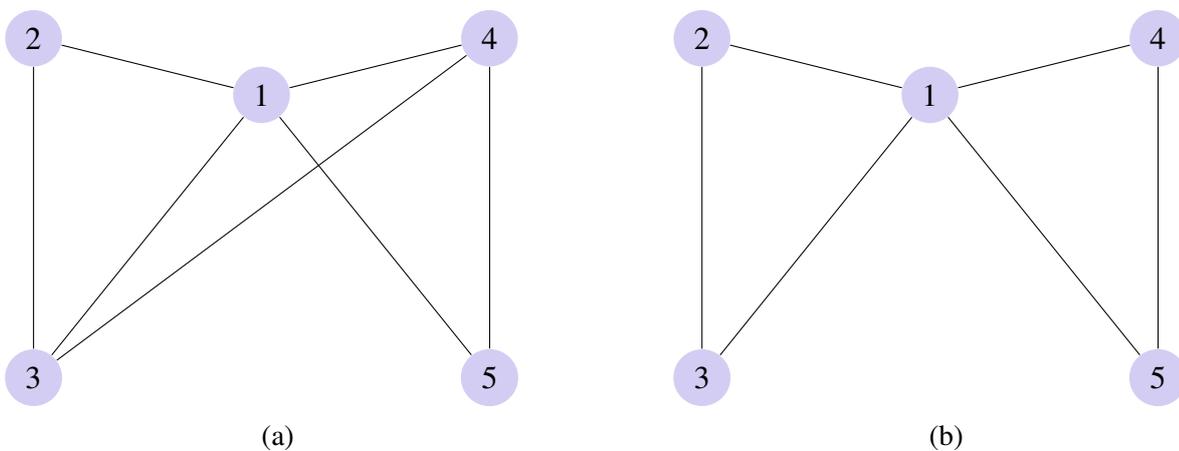


Figura 45 – (a) Grafo do modelo 672; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 67, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,3), (1,4), (1,5) e (2,3). Com uma moderada probabilidade, ele visitou grafos que tinham as arestas (3,4) e (4,5). Com uma probabilidade baixa, ele visitou grafos que tinham a aresta (3,5). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (2,4) e (2,5).

Tabela 67 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 300.000$ e $n = 250$.

Vértices	1	2	3	4	5
1	-	300.000	300.000	300.000	300.000
2	1,000	-	300.000	57	88
3	1,000	1,000	-	209.812	91.408
4	1,000	0,000	0,699	-	197.436
5	1,000	0,000	0,305	0,658	-

Fonte: Autor.

Na Tabela 68, observa-se que o algoritmo acertou em captar a presença das arestas (1,2), (1,3), (1,4), (1,5) e (2,3), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a presença da aresta (4,5), e a ausência das arestas (3,4) e (3,5).

Tabela 68 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 300.000$ e $n = 250$.

Vértices	2	3	4	5
1	0,000	0,000	0,000	0,000
2	-	0,000	0,000	0,000
3	-	-	-0,699	-0,305
4	-	-	-	0,342

Fonte: Autor.

5.3.8.3 Algoritmo da Proposta 2 ($m = 300.000$ e $n = 250$)

Neste cenário, o modelo 416 foi o mais visitado, obtendo um total de 78.990 visitas. Entretanto, o modelo 544 obteve, somente, um total de 41.022 visitas. O $IS(416|544) = 7$ e $ISIU(416|544) = 0,70$. Com isso, tem-se que o modelo 416 possui uma semelhança de 70% com o modelo 544 que pode ser observada na Figura 46.

Na Tabela 69, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,3), (1,4), (1,5) e (2,3). Com uma moderada probabilidade, ele visitou grafos que tinham a aresta (3,5). Com uma probabilidade baixa, ele visitou grafos que tinham as arestas (3,4) e (4,5). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (2,4) e (2,5).

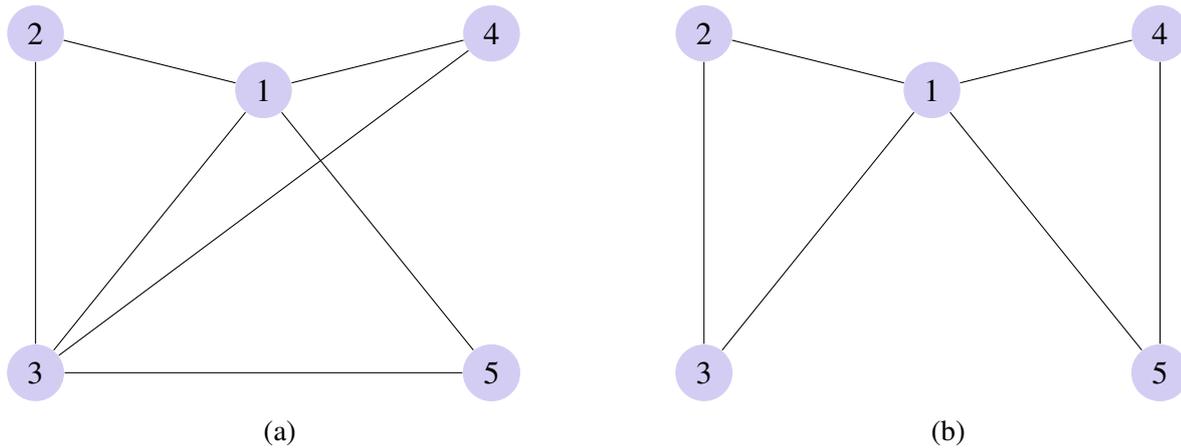


Figura 46 – (a) Grafo do modelo 416; (b) Grafo do modelo 544. Fonte: Autor.

Tabela 69 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 300.000$ e $n = 250$.

Vértices	1	2	3	4	5
1	-	300.000	300.000	300.000	300.000
2	1,000	-	300.000	777	235
3	1,000	1,000	-	110.026	150.566
4	1,000	0,003	0,367	-	42.546
5	1,000	0,001	0,502	0,142	-

Fonte: Autor.

Na Tabela 70, observa-se que o algoritmo acertou em captar a presença das arestas (1,2), (1,3), (1,4), (1,5) e (2,3), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a ausência das arestas (3,4) e (3,5). Todavia, o algoritmo errou em não captar a presença das aresta (4,5).

Tabela 70 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 300.000$ e $n = 250$.

Vértices	2	3	4	5
1	0,000	0,000	0,000	0,000
2	-	0,000	-0,003	-0,001
3	-	-	-0,367	-0,502
4	-	-	-	0,858

Fonte: Autor.

5.3.8.4 Algoritmo da Proposta 3 ($m = 300.000$ e $n = 250$)

Neste cenário, o modelo 32 foi o mais visitado, obtendo um total de 193.260 visitas. Entretanto, o modelo 544 obteve, somente, um total de 86 visitas. O $IS(32|544) = 9$ e $ISIU(32|544) = 0,90$. Com isso, tem-se que o modelo 32 possui uma semelhança de 90% com o modelo 544 que pode ser observada na Figura 47.

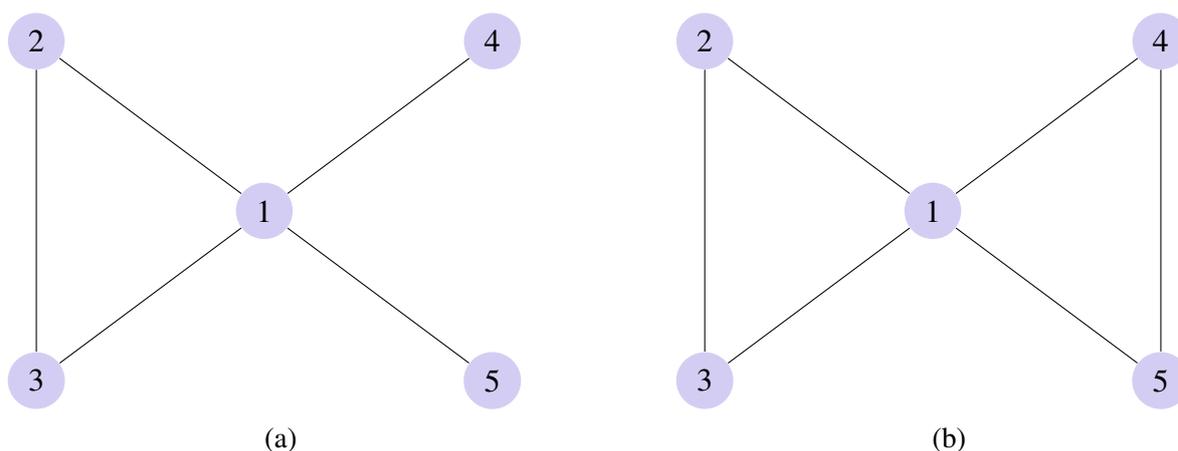


Figura 47 – (a) Grafo do modelo 32; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 71, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,3), (1,4), (1,5) e (2,3). Com uma probabilidade baixa, ele visitou grafos que tinham a aresta (3,5). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (2,4), (2,5), (3,4) e (4,5).

Tabela 71 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 300.000$ e $n = 250$.

Vértices	1	2	3	4	5
1	-	300.000	300.000	300.000	300.000
2	1,000	-	300.000	695	9
3	1,000	1,000	-	26.681	87.227
4	1,000	0,002	0,089	-	107
5	1,000	0,000	0,291	0,000	-

Fonte: Autor.

Na Tabela 72, observa-se que o algoritmo acertou em captar a presença das arestas (1,2), (1,3), (1,4), (1,5) e (2,3), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a ausência das arestas (3,4) e (3,5). Todavia, o algoritmo errou em não captar a presença da aresta (4,5).

Tabela 72 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 300.000$ e $n = 250$.

Vértices	2	3	4	5
1	0,000	0,000	0,000	0,000
2	-	0,000	-0,002	0,000
3	-	-	-0,089	-0,291
4	-	-	-	1,000

Fonte: Autor.

5.3.8.5 Comparação dos Algoritmos ($m = 300.000$ e $n = 250$)

Na Figura 48, observa-se que não existem tendências nos gráficos de todas as variações do algoritmo *Reversible Jump*. Isso é um indício de que houve convergência nos algoritmos. Entretanto, o algoritmo da proposta 3 apresentou um comportamento diferente que os outros.

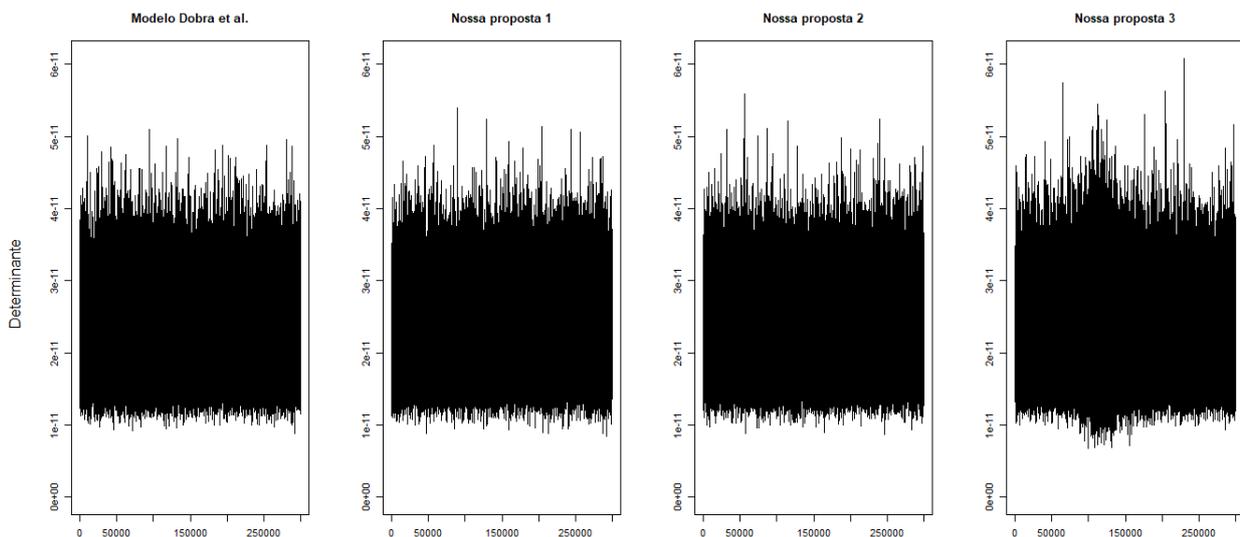


Figura 48 – Gráficos dos determinantes para avaliar a convergência dos algoritmos considerando $n = 250$, $m = 300.000$, warm-up = 100 e upgrade = 5. Fonte: Autor.

Na Figura 49, observa-se que um decaimento brusco em todas as linhas decrescentes dos gráficos. Isso indica que menos de 100 modelos foram visitados durante o processamento dos algoritmos. Além disso, o verdadeiro modelo 544 obteve baixos valores de probabilidade *a posteriori* em dois algoritmos (Dobra et al. e proposta 3). Entretanto, o algoritmo da proposta 1 e 2 tiveram desempenhos superiores aos outros dois algoritmos.

Na Tabela 73, observa-se que os algoritmos de Dobra et al. e da proposta 1 obtiveram um empate, pois Dobra et al. teve os menores valores para as medidas SAE e SAEIU, e a proposta 1 teve os menores valores para as medidas SQE e SQEIU. Isso indica que os algoritmos de Dobra et al. e da proposta 1 cometeram os menores erros ao captar as presenças e ausências das arestas do verdadeiro modelo 544, em relação aos outros algoritmos. Considerando a Tabela 73 e as

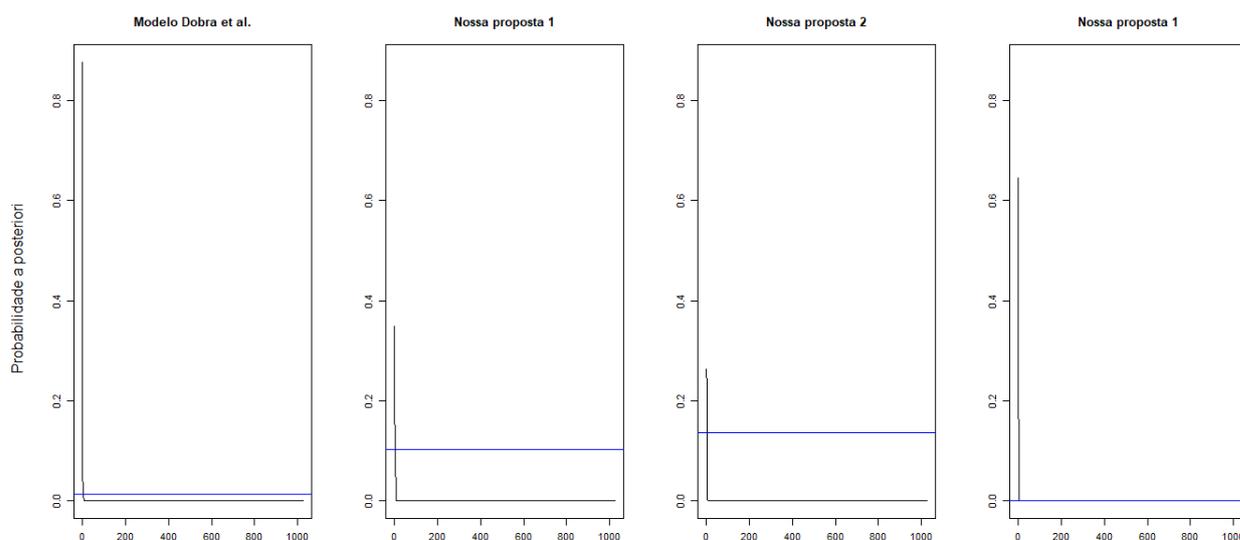


Figura 49 – Gráficos de avaliação das probabilidades *a posteriori* considerando $n = 250$, $m = 300.000$, $\text{warm-up} = 100$ e $\text{upgrade} = 5$. Fonte: Autor.

Figuras 48 e 49, pode-se concluir que o algoritmo da proposta 1 obteve o melhor desempenho para o cenário $m = 300.000$ e $n = 250$.

Tabela 73 – A soma absoluta dos erros (SAE), a soma absoluta dos erros no intervalo unitário (SAEIU), a soma quadrática dos erros (SQE) e a soma quadrática dos erros no intervalo unitário (SQEIU), para o cenário $m = 300.000$ e $n = 250$.

Algoritmo	SAE	SAEIU	SQE	SQEIU
Dobra et al.	1,071	0,107	0,829	0,083
Proposta 1	1,346	0,135	0,699	0,070
Proposta 2	1,730	0,173	1,123	0,112
Proposta 3	1,382	0,138	1,092	0,109

Fonte: Autor.

5.3.9 Resultados no cenário com $m = 300.000$ e $n = 500$

Nesta subseção, os resultados das simulações para o cenário $m = 300.000$ e $n = 500$ são apresentados e interpretados para as quatro variações do algoritmo *Reversible Jump*.

5.3.9.1 Algoritmo de Dobra et al. ($m = 300.000$ e $n = 500$)

Neste cenário, o modelo 928 foi o mais visitado, obtendo um total de 95.549 visitas. Entretanto, o modelo 544 obteve, somente, um total de 54.440 visitas. O $\text{IS}(928|544) = 8$ e $\text{ISIU}(928|544) = 0,80$. Com isso, tem-se que o modelo 928 possui uma semelhança de 80% com o modelo 544 que pode ser observada na Figura 50.

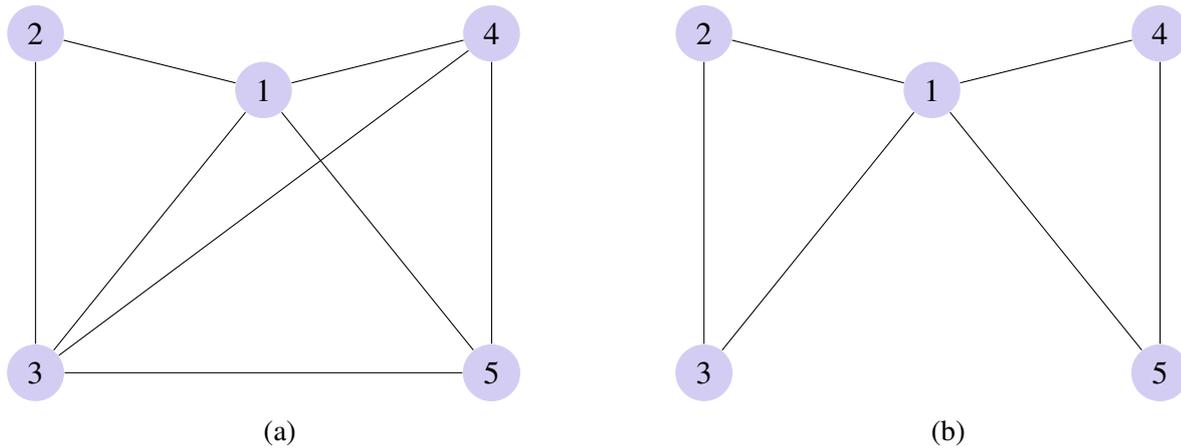


Figura 50 – (a) Grafo do modelo 928; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 74, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,3), (1,4), (1,5), (2,3) e (4,5). Com uma moderada probabilidade, visitou grafos que tinham as arestas (3,4) e (3,5). E, com probabilidade muito baixa, visitou grafos que tinham as arestas (2,4) e (2,5).

Tabela 74 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 300.000$ e $n = 500$.

Vértices	1	2	3	4	5
1	-	300.000	300.000	300.000	300.000
2	1,000	-	300.000	68	41
3	1,000	1,000	-	185.924	155.337
4	1,000	0,000	0,620	-	273.596
5	1,000	0,000	0,518	0,912	-

Fonte: Autor.

Na Tabela 75, observa-se que o algoritmo acertou em captar a presença das arestas (1,2), (1,3), (1,4), (1,5), (2,3) e (4,5), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a ausência das arestas (3,4) e (3,5).

Tabela 75 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 300.000$ e $n = 500$.

Vértices	2	3	4	5
1	0,000	0,000	0,000	0,000
2	-	0,000	0,000	0,000
3	-	-	-0,620	-0,518
4	-	-	-	0,088

Fonte: Autor.

5.3.9.2 Algoritmo da Proposta 1 ($m = 300.000$ e $n = 500$)

Neste cenário, o modelo 160 foi o mais visitado, obtendo um total de 81.871 visitas. Entretanto, o modelo 544 obteve, somente, um total de 41.145 visitas. O $IS(160|544) = 8$ e $ISIU(160|544) = 0,80$. Com isso, tem-se que o modelo 160 possui uma semelhança de 80% com o modelo 544 que pode ser observada na Figura 51.

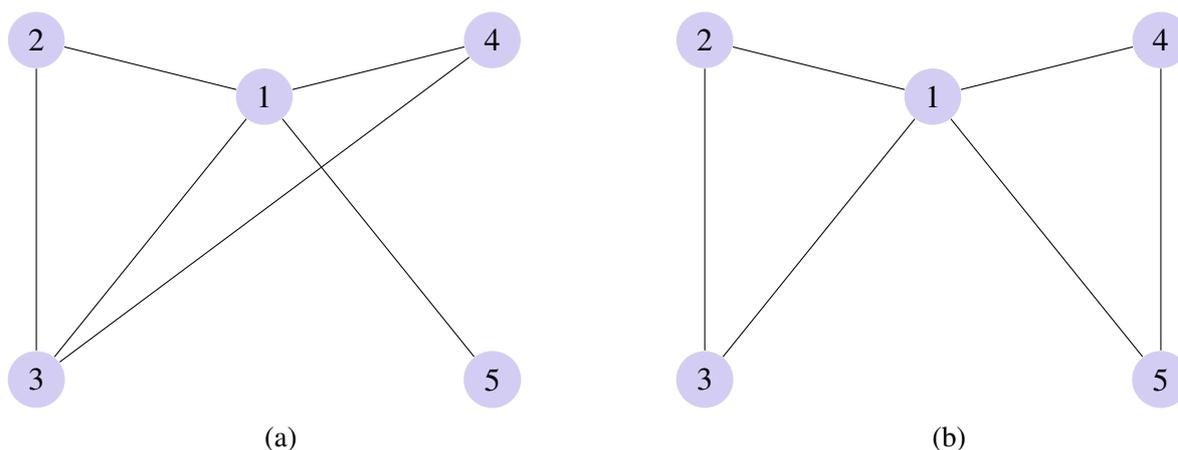


Figura 51 – (a) Grafo do modelo 160; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 76, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,3), (1,4), (1,5) e (2,3). Com uma moderada probabilidade, ele visitou grafos que tinha a aresta (3,4). Com uma probabilidade baixa, ele visitou grafos que tinham as arestas (3,5) e (4,5). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (2,4) e (2,5).

Tabela 76 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 300.000$ e $n = 500$.

Vértices	1	2	3	4	5
1	-	300.000	300.000	300.000	300.000
2	1,000	-	300.000	24	75
3	1,000	1,000	-	152.646	88.579
4	1,000	0,000	0,509	-	80.244
5	1,000	0,000	0,295	0,267	-

Fonte: Autor.

Na Tabela 77, observa-se que o algoritmo acertou em captar a presença das arestas (1,2), (1,3), (1,4), (1,5) e (2,3), e a ausência das arestas (2,4) e (2,5). O algoritmo também conseguiu parcialmente captar a ausência das arestas (3,4) e (3,5). Todavia, o algoritmo errou ao não captar a presença da aresta (4,5).

Tabela 77 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 300.000$ e $n = 500$.

Vértices	2	3	4	5
1	0,000	0,000	0,000	0,000
2	-	0,000	0,000	0,000
3	-	-	-0,509	-0,295
4	-	-	-	0,733

Fonte: Autor.

5.3.9.3 Algoritmo da Proposta 2 ($m = 300.000$ e $n = 500$)

Neste cenário, o modelo 32 foi o mais visitado, obtendo um total de 154.504 visitas. Entretanto, o modelo 544 obteve, somente, um total de 161 visitas. O $IS(32|544) = 9$ e $ISIU(32|544) = 0,90$. Com isso, tem-se que o modelo 32 possui uma semelhança de 90% com o modelo 544 que pode ser observada na Figura 52.

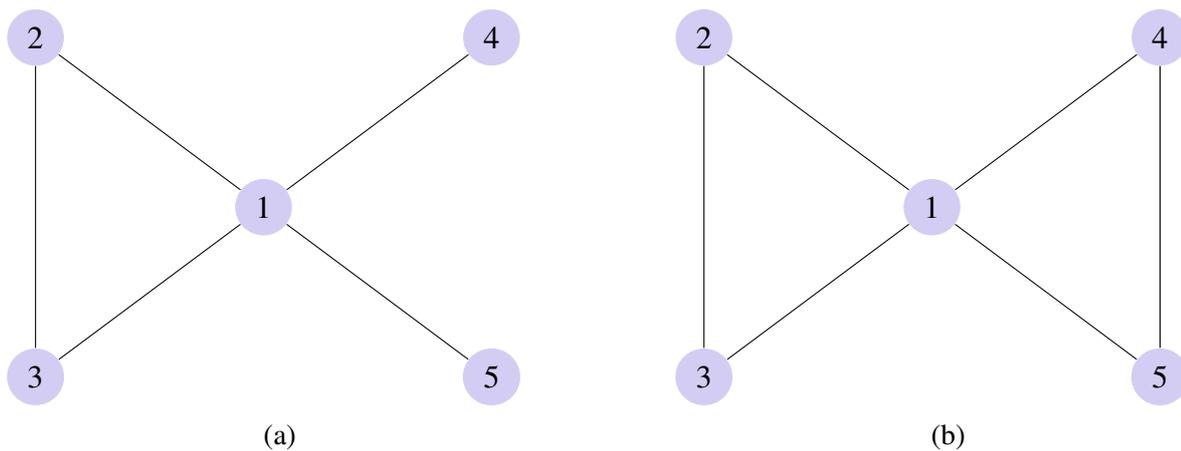


Figura 52 – (a) Grafo do modelo 32; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 78, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,3), (1,4), (1,5) e (2,3). Com uma moderada probabilidade, ele visitou grafos que tinham a aresta (3,5). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (2,4), (2,5), (3,4) e (4,5).

Na Tabela 79, observa-se que o algoritmo acertou em captar a presença das arestas (1,2), (1,3), (1,4), (1,5) e (2,3), e a ausência das arestas (2,4), (2,5) e (3,4). O algoritmo também conseguiu parcialmente captar a ausência das arestas (3,5). Todavia, o algoritmo errou em não captar a presença da aresta (4,5).

Tabela 78 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 300.000$ e $n = 500$.

Vértices	1	2	3	4	5
1	-	300.000	300.000	300.000	300.000
2	1,000	-	300.000	518	118
3	1,000	1,000	-	4.667	142.685
4	1,000	0,002	0,016	-	1.673
5	1,000	0,001	0,476	0,006	-

Fonte: Autor.

Tabela 79 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 300.000$ e $n = 500$.

Vértices	2	3	4	5
1	0,000	0,000	0,000	0,000
2	-	0,000	-0,002	0,000
3	-	-	-0,016	-0,476
4	-	-	-	0,994

Fonte: Autor.

5.3.9.4 Algoritmo da Proposta 3 ($m = 300.000$ e $n = 500$)

Neste cenário, o modelo 32 foi o mais visitado, obtendo um total de 134.556 visitas. Entretanto, o modelo 544 obteve, somente, um total de 115.697 visitas. O $IS(32|544) = 9$ e $ISIU(32|544) = 0,90$. Com isso, tem-se que o modelo 32 possui uma semelhança de 90% com o modelo 544 que pode ser observada na Figura 53.

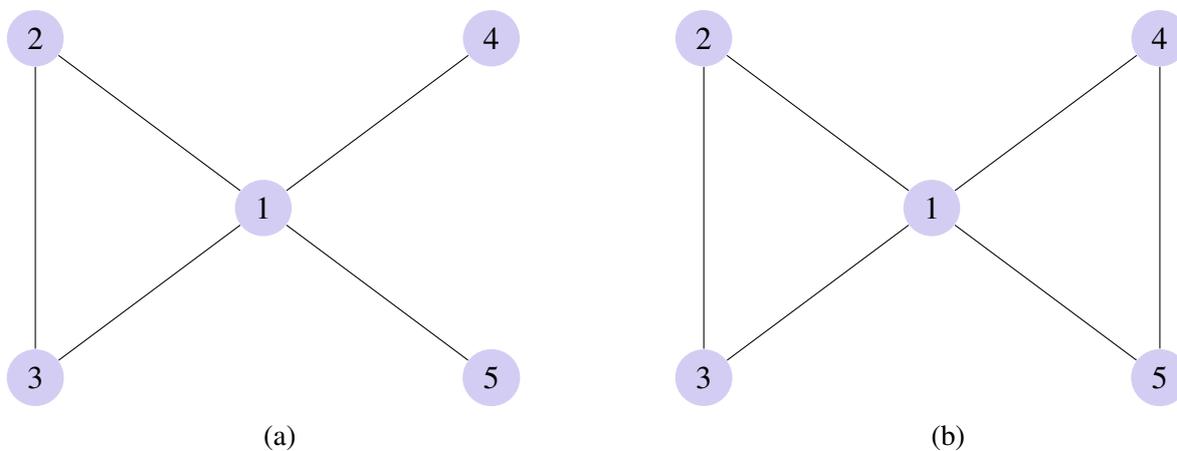


Figura 53 – (a) Grafo do modelo 32; (b) Grafo do modelo 544. Fonte: Autor.

Na Tabela 80, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo visitou grafos que tinham as arestas (1,2), (1,3) (1,4), (1,5) e (2,3). Com um moderada probabilidade, ele visitou grafos que tinham a aresta (4,5). Com uma probabilidade baixa, ele visitou grafos que tinham a aresta (3,5). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (2,4), (2,5) e (3,4).

Tabela 80 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o cenário $m = 300.000$ e $n = 500$.

Vértices	1	2	3	4	5
1	-	300.000	300.000	300.000	300.000
2	1,000	-	300.000	454	103
3	1,000	1,000	-	536	48.872
4	1,000	0,002	0,002	-	140.688
5	1,000	0,000	0,163	0,469	-

Fonte: Autor.

Na Tabela 81, observa-se que o algoritmo acertou em captar a presença das arestas (1,2), (1,3) (1,4), (1,5) e (2,3), e a ausência das arestas (2,4), (2,5) e (3,4). O algoritmo também conseguiu parcialmente captar a presença da aresta (4,5), e a ausência da aresta (3,5).

Tabela 81 – Os índices de erro das arestas (IEA) é apresentada na parte triangular superior, para o cenário $m = 300.000$ e $n = 500$.

Vértices	2	3	4	5
1	0,000	0,000	0,000	0,000
2	-	0,000	-0,002	0,000
3	-	-	-0,002	-0,163
4	-	-	-	0,531

Fonte: Autor.

5.3.9.5 Comparação dos Algoritmos ($m = 300.000$ e $n = 500$)

Na Figura 54, observa-se que não existem tendências nos gráficos de todas as variações do algoritmo *Reversible Jump*. Isso é um indício de que houve convergência nos algoritmos. Entretanto, o algoritmo de Dobra et al. apresentou um comportamento diferente que os outros.

Na Figura 55, observa-se que um decaimento brusco em todas as linhas decrescentes dos gráficos. Isso indica que menos de 100 modelos foram visitados durante o processamento dos algoritmos. Além disso, o verdadeiro modelo 544 obteve baixo valor de probabilidade *a posteriori* em somente um algoritmo (proposta 2). Os outros algoritmos obtiveram bons valores de probabilidade *a posteriori*, se destacando a proposta 3.

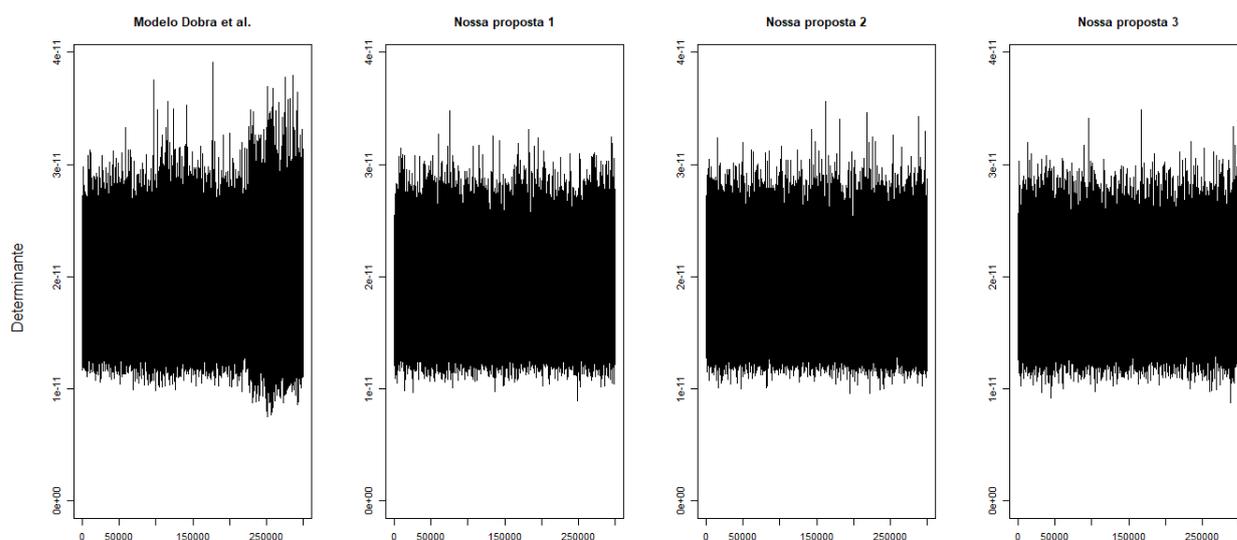


Figura 54 – Gráficos dos determinantes para avaliar a convergência dos algoritmos considerando $n = 500$, $m = 300.000$, warm-up = 100 e upgrade = 5. Fonte: Autor.

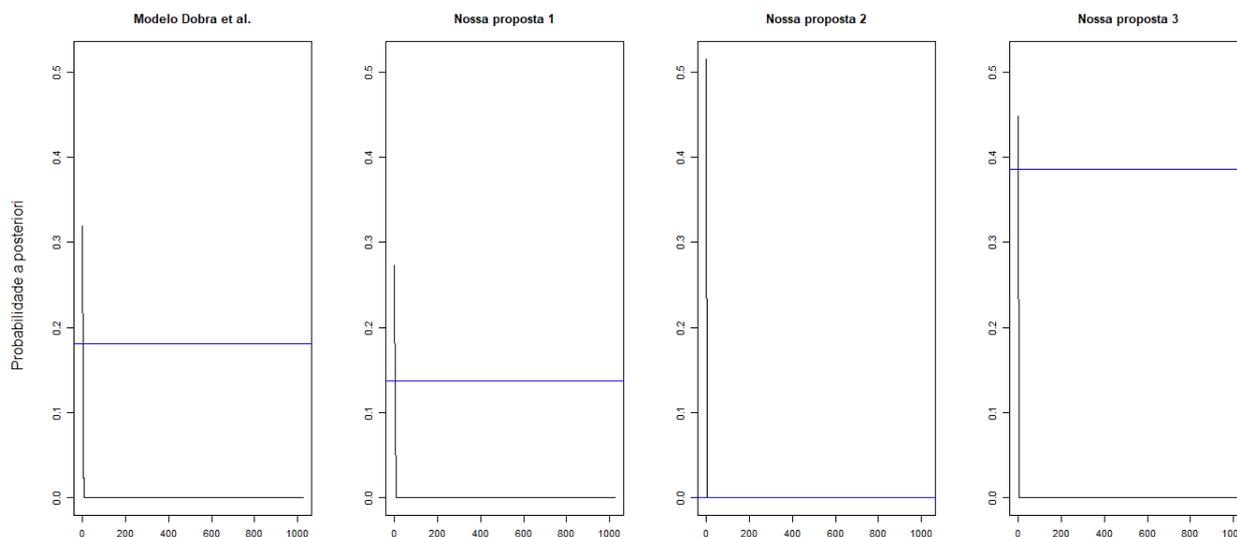


Figura 55 – Gráficos de avaliação das probabilidades *a posteriori* considerando $n = 500$, $m = 300.000$, warm-up = 100 e upgrade = 5. Fonte: Autor.

Na Tabela 82, observa-se que o algoritmo da proposta 3 obteve os melhores resultados em comparação ao outras variações, pois teve os menores valores em todas as medidas SAE, SAEIU, SQE e SQEIU. Isso indica que o algoritmo da proposta 3 cometeu menos erros ao captar as presenças e ausências das arestas do verdadeiro modelo 544, em relação ao outros algoritmos. Considerando a Tabela 82 e as Figuras 54 e 55, pode-se concluir que o algoritmo da proposta 3 obteve o melhor desempenho para o cenário $m = 300.000$ e $n = 500$.

Tabela 82 – A soma absoluta dos erros (SAE), a soma absoluta dos erros no intervalo unitário (SAEIU), a soma quadrática dos erros (SQE) e a soma quadrática dos erros no intervalo unitário (SQEIU), para o cenário $m = 300.000$ e $n = 500$.

Algoritmo	SAE	SAEIU	SQE	SQEIU
Dobra et al.	1,226	0,123	0,660	0,066
Proposta 1	1,537	0,154	0,883	0,088
Proposta 2	1,488	0,149	1,215	0,122
Proposta 3	0,698	0,070	0,309	0,031

Fonte: Autor.

5.4 Comparação dos Critérios

Com relação aos critérios quantitativos, SAE, SAEIU, SQE e SQEIU são os únicos critérios que dependem da convergência do modelo, para que seus resultados sejam confiáveis. CPA e PPPA são bons indicadores descritivos, mas o PPPA carrega mais informação que CPA, pois PPPA por construção agrega as informações de CPA para calcular probabilidades, que são indicadores fáceis de interpretar. Normalmente, se o valor de $PPPA_{ij}$ for aproximadamente 0,5 ou superior, observa-se a presença da aresta (i, j) no grafo mais visitado. No estudo de simulação, o IEA comportou-se como um excelente indicador para identificar a captação da verdadeira estrutura do grafo. No geral, SAE, SAEIU, SQE e SQEIU são indicadores que têm seus resultados concordando, mas no caso de resultados não conclusivos deve-se utilizar o gráfico de avaliação das probabilidades *a posteriori* como auxílio para interpretação. SAEIU e SQEIU são os indicadores com resultados mais fáceis de serem interpretados, devido as suas escalas no intervalo unitário. O gráfico de linhas dos determinantes é uma ferramenta eficiente para avaliar a convergência do algoritmo *Reversible Jump*. Já o gráfico das probabilidades *a posteriori* tem um comportamento adequado para comparar os resultados.

APLICAÇÃO

Os dados utilizados na aplicação são referentes as notas de cinco disciplinas (álgebra, vetores, mecânica, análise e estatística) de 88 indivíduos. Esses dados são considerados clássicos no estudo da teoria de grafos e eles podem ser encontrados no programa R, pacote "gRbase", comandos: `data(mathmark)`; `mathmark`. O problema consiste em descobrir como estas cinco disciplinas se relacionam entre si.

Na Tabela 83, as estatísticas descritivas das cinco disciplinas são apresentadas. O intervalo de variação das cinco disciplinas é $\{0, 1, 2, \dots, 100\}$. A disciplina mecânica foi a única que atingiu o limite inferior do intervalo. Entretanto, nenhuma disciplina conseguiu atingir o limite superior e o máximo valor atingido foi 82 pela disciplina vetores. As médias de algebra e vetores são 50,60 e 50,59, respectivamente, e são aproximadamente iguais. As outras médias são distintas com 46,68 (análise), 42,31 (estatística) e 38,95 (mecânica). O menor desvio padrão é 10,62 (álgebra) e o maior é 17,49 (mecânica).

Tabela 83 – Estatísticas descritivas das cinco disciplinas.

Disciplina	min	q_{25}^*	q_{50}^*	q_{75}^*	max	média	dp*
Álgebra (<i>al</i>)	15,00	45,00	50,00	57,25	80,00	50,60	10,62
Vetores (<i>ve</i>)	9,00	42,00	51,00	60,00	82,00	50,59	13,15
Mecânica (<i>me</i>)	0,00	30,00	41,50	49,25	77,00	38,95	17,49
Análise (<i>an</i>)	9,00	35,75	49,00	57,00	70,00	46,68	14,85
Estatística (<i>es</i>)	9,00	31,00	40,00	51,50	81,00	42,31	17,26

*: q_{25} é 1º quartil; q_{50} é 2º quartil; q_{75} é 3º quartil; dp é desvio padrão.

Fonte: Autor.

Na Tabela 84, todas as correlações parciais das disciplinas são apresentadas. Observa-se valores moderados de 0,4319 (álgebra, análise), 0,3568 (álgebra, estatística) e 0,3293 (vetores, mecânica). Valores baixos de 0,2808 (álgebra, vetores), 0,2304 (álgebra, mecânica) e 0,2528 (análise, estatística). E, valores muito baixos de 0,0781 (vetores, análise), 0,0202 (vetores, estatística), -0,0016 (mecânica, análise) e -0,0016 (mecânica, estatística). Com isso, nós podemos

ver indícios de que a estrutura do grafo das relações entre as disciplinas não contém as arestas (mecânica, análise), (mecânica, estatística) e (vetores, estatística), pois suas correlações parciais são aproximadamente zeros.

Tabela 84 – Correlações parciais das disciplinas.

Disciplina	Vetores	Mecânica	Análise	Estatística
Álgebra	0,2808	0,2304	0,4319	0,3568
Vetores	-	0,3293	0,0781	0,0202
Mecânica	-	-	-0,0016	-0,0016
Análise	-	-	-	0,2528

Fonte: Autor.

Sejam as variáveis aleatórias referentes aos vértices (disciplinas) tem distribuição $X_v \sim N(\mu_v, \sigma_v^2)$. Para atender a suposição do modelo estudado, será realizada a seguinte transformação nas variáveis aleatórias

$$Y_v = X_v - \bar{X}_v,$$

em que $v \in \{al, ve, me, an, es\}$. Logo, $Y_v \sim N(0, \sigma_v^2)$ para todo v .

Considera-se que *a priori* todos os grafos G têm probabilidades iguais de ocorrerem. Assim, a distribuição *a priori* de G é uma uniforme discreta. As outras especificações são as mesmas mencionadas nas Seções 3.2 e 4.2, em que $\sigma_m^2 = \sigma_g^2 = 10$.

O *Reversible Jump* (RJ) é processado para o algoritmo da Proposta 3, pois obteve os melhores resultados nas simulações do Capítulo 5. No total foram computados 300.000 iterações do RJ ($m = 300.000$), entre as iterações do RJ foram feitas 5 iterações do Metropolis-Hastings (upgrade = 5) e 500 iterações do MH antes da primeira iteração do RJ (warm-up = 100, pois warm-up \times upgrade = 500).

Do total de 1.500.495 simulações de MH, 750.247 foram descartas e 750.248 foram usadas para estimar a matriz K (inversa da matriz de covariâncias de Y_v). A semente 8481670 foi utilizada no processo do algoritmo.

No algoritmo Proposta 3, o modelo 261 foi o mais visitado, obtendo um total de 124.360 visitas e uma probabilidade *a posteriori* de 0,4145. Isso, indica que o modelo 261 foi visitado 41,45% do total de visitas (300.000). Na Figura 56, a estrutura do modelo grafo mais visitado é apresentada.

Na Tabela 85, observa-se que com uma alta probabilidade *a posteriori*, o algoritmo Proposta 3 visitou grafos que tinham a aresta (al, an). Com uma moderada probabilidade, ele visitou grafos que tinham a aresta (me, es). Com uma probabilidade baixa, ele visitou grafos que tinham as arestas (al, ve) e (me, an). E, com uma probabilidade muito baixa, visitou grafos que tinham as arestas (al, me), (al, es), (ve, me), (ve, an), (ve, es), e (an, es).

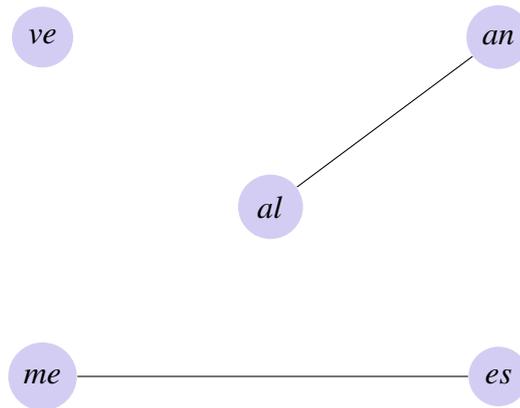


Figura 56 – Grafo do modelo 261. Fonte: Autor.

Tabela 85 – A contagem de permanência das arestas (CPA) é apresentada na parte triangular superior e a probabilidade *a posteriori* de permanência das arestas (PPPA) é apresentada na parte triangular inferior, para o algoritmo Proposta 3.

Disciplina	Álgebra	Vetores	Mecânica	Análise	Estatística
Álgebra	-	67.106	11.770	299.278	1.454
Vetores	0,224	-	9.472	738	27
Mecânica	0,039	0,032	-	32.372	196.069
Análise	0,998	0,002	0,108	-	203
Estatística	0,005	0,000	0,654	0,001	-

Fonte: Autor.

Na Figura 57, observa-se que não existe tendência no gráfico da Proposta 3 do algoritmo RJ. Isso é um indício de que houve convergência no algoritmo RJ e, portanto, as estimativas obtidas são confiáveis.

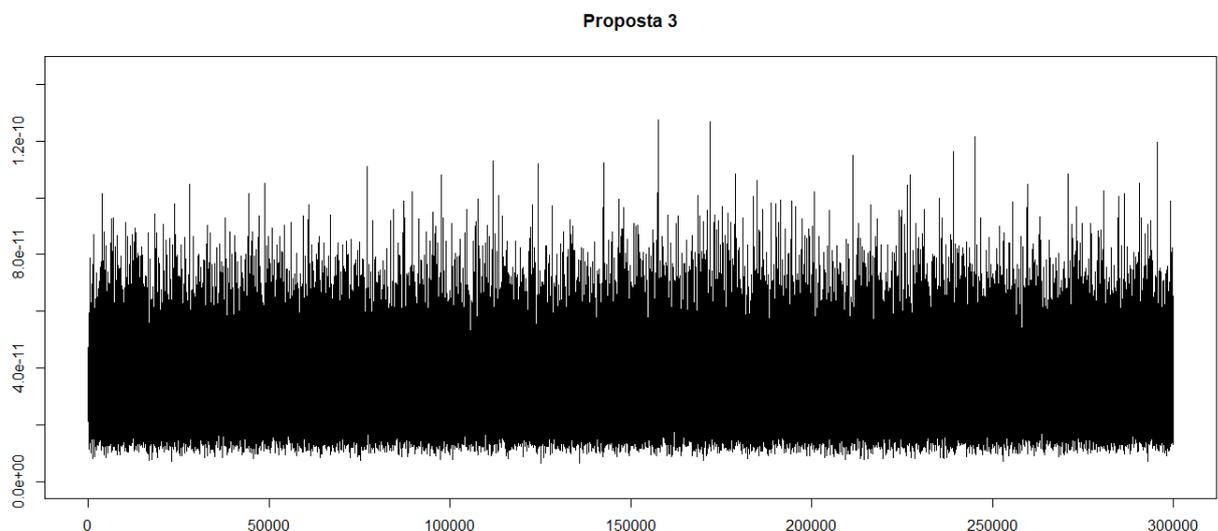


Figura 57 – Gráfico dos determinantes para avaliar a convergência do algoritmo Proposta 3, considerando $m = 300.000$, warm-up = 100 e upgrade = 5. Fonte: Autor.

Na Tabela 86, as estimativas dos elementos de K são apresentadas para o algoritmo Proposta 3. Observa-se que todos os elementos da diagonal principal de K são positivos, indicando que a suposição de K é positiva definida se manteve.

Tabela 86 – Estimativa dos elementos de K para o algoritmo Proposta 3.

Parâmetro	Estimativa	Desvio Padrão	Intervalo de 95%
$K_{al,al}$	2,859e-02	4,190e-03	[0,02093 ; 0,03736]
$K_{al,ve}$	-3,854e-03	1,719e-03	[-0,00762 ; -0,00097]
$K_{al,me}$	-3,045e-03	1,198e-03	[-0,00548 ; -0,00076]
$K_{al,an}$	-7,778e-03	1,969e-03	[-0,01184 ; -0,00411]
$K_{al,es}$	-4,142e-03	1,171e-03	[-0,00663 ; -0,00202]
$K_{ve,ve}$	1,065e-02	1,566e-03	[0,00780 ; 0,01395]
$K_{ve,me}$	-2,404e-03	7,840e-04	[-0,00399 ; -0,00097]
$K_{ve,an}$	-9,325e-04	9,179e-04	[-0,00268 ; 0,00107]
$K_{ve,es}$	-1,020e-03	6,293e-04	[-0,00212 ; 0,00041]
$K_{me,me}$	5,483e-03	7,705e-04	[0,00408 ; 0,00710]
$K_{me,an}$	-3,736e-04	7,733e-04	[-0,00184 ; 0,00107]
$K_{me,es}$	2,427327	2,074235	[-0,00135 ; 5,99447]
$K_{an,an}$	1,094e-02	1,559e-03	[0,00806 ; 0,01417]
$K_{an,es}$	-1,014595	0,911825	[-3,00134 ; -0,00116]
$K_{es,es}$	2363,461	2761,080	[6,409e-03 ; 7,724e+03]

Fonte: Autor.

Na Tabela 87, as estimativas das correlações parciais das disciplinas são apresentadas. Observa-se valores moderados de 0,4398 (álgebra, análise) e -0,6743 (mecânica, estatística). Valores baixos de 0,2209 (álgebra, vetores), 0,2432 (álgebra, mecânica), 0,3146 (vetores, mecânica) e 0,1995 (análise, estatística). E, valores muito baixos de 0,0005 (álgebra, estatística), 0,0876 (vetores, análise), 0,0002 (vetores, estatística) e 0,0482 (mecânica, análise). Observando ambas as Tabelas 84 e 87, nós podemos destacar que houve a troca de sinal na estimativa de (mecânica, análise) comparado com o seu valor descritivo.

Tabela 87 – Estimativa das correlações parciais das disciplinas.

Disciplina	Vetores	Mecânica	Análise	Estatística
Álgebra	0,2209	0,2432	0,4398	0,0005
Vetores	-	0,3146	0,0876	0,0002
Mecânica	-	-	0,0482	-0,6743
Análise	-	-	-	0,1995

Fonte: Autor.

DISCUSSÃO

Neste capítulo, discutiremos sobre as contribuições e os resultados obtidos nos Capítulos 4, 5 e 6. Ao final, propostas de pesquisas futuras foram pontuadas.

7.1 Discussão dos Resultados

No capítulo 4, a primeira contribuição do trabalho é apresentada. Nós desenvolvemos distribuições alternativas para o gerador de candidato do método *Reversible Jump* aplicado ao modelo grafo gaussiano de [Dobra, Lenkoski e Rodriguez \(2011\)](#). Nós propomos distribuições baseadas nas correlações parciais, pois queríamos que os dados fizessem um melhor direcionamento na escolha do candidato. Além disso, a ideia de utilizar a correlação parcial nas distribuições propostas foi elaborada observando que sua estrutura teórica se encaixa na estrutura de um grafo, mais especificamente na representação da aresta.

No capítulo 5, os resultados apresentados são bastante interessantes, pois, após desenvolver as contribuições no capítulo 4, nós atentamos na necessidade de indicadores avaliativos e comparativos que poderiam ser utilizados em um estudo de simulação, cujos resultados são provenientes de um modelo grafo gaussiano. Com isso, observamos que existe uma falta de critérios avaliadores na literatura de modelos grafos e foi preciso desenvolver alguns indicadores para avaliar e comparar resultados de um estudo de simulação. Os critérios criados atenderam as necessidades e, com isso, foi possível comparar as contribuições do capítulo 4 com o método já proposto por [Dobra, Lenkoski e Rodriguez \(2011\)](#). No fim do estudo de simulação, nós chegamos na conclusão que nossas alterações traziam vantagens na busca da estrutura do grafo, pois as propostas apresentadas cometeram menos erros ao buscar a estrutura de covariância que melhor representava os dados dos grafos gaussianos.

No capítulo 6, um banco de dados já conhecido e usado por [Whittaker \(1990\)](#) foi aplicado as nossas modificações no método de [Dobra, Lenkoski e Rodriguez \(2011\)](#). Os resultados obtidos

da aplicação mostraram um bom comportamento ao utilizarmos as nossas modificações sugeridas no trabalho.

7.2 Proposta de Pesquisa Futura

A seguir estão apresentados alguns pontos de possíveis pesquisas futuras:

- i. Mesclar as abordagens de modelo grafo gaussiano com modelo de mistura;
- ii. Expandir o modelo de seleção de covariância para grafos não gaussianos;
- iii. Desenvolver outros métodos para simular dados provenientes de um grafo gaussiano;
- iii. Expandir os resultados do modelo grafo gaussiano para a família exponencial.

REFERÊNCIAS

- ALBERT, R.; BARABÁSI, A.-L. Statistical mechanics of complex networks. **Reviews of modern physics**, APS, v. 74, n. 1, p. 47, 2002. Citado na página 16.
- ATAY-KAYIS, A.; MASSAM, H. A monte carlo method for computing the marginal likelihood in nondecomposable gaussian graphical models. **Biometrika**, Oxford University Press, v. 92, n. 2, p. 317–335, 2005. Citado nas páginas 21, 22 e 27.
- DAUDIN, J.-J.; PICARD, F.; ROBIN, S. A mixture model for random graphs. **Statistics and computing**, Springer, v. 18, n. 2, p. 173–183, 2008. Citado nas páginas 15 e 16.
- DIACONIS, P.; YLVISAKER, D. Conjugate priors for exponential families. **The Annals of statistics**, JSTOR, p. 269–281, 1979. Citado na página 21.
- DOBRA, A.; LENKOSKI, A. Copula gaussian graphical models and their application to modeling functional disability data. **The Annals of Applied Statistics**, Institute of Mathematical Statistics, v. 5, n. 2A, p. 969–993, 2011. Citado na página 16.
- DOBRA, A.; LENKOSKI, A.; RODRIGUEZ, A. Bayesian inference for general gaussian graphical models with application to multivariate lattice data. **Journal of the American Statistical Association**, Taylor & Francis, v. 106, n. 496, p. 1418–1433, 2011. Citado nas páginas 16, 17, 19, 20, 21, 22, 23, 24, 25, 26, 36 e 101.
- ERDOS, P.; RENYI, A. On random graphs i. **Publicationes Mathematicae**, v. 6, p. 419–427, 1959. Citado na página 15.
- ERDŐS, P.; RÉNYI, A. *et al.* On the evolution of random graphs. **Publ. Math. Inst. Hung. Acad. Sci.**, v. 5, n. 1, p. 17–60, 1960. Citado na página 15.
- GREEN, P. J. Reversible jump markov chain monte carlo computation and bayesian model determination. **Biometrika**, Oxford University Press, v. 82, n. 4, p. 711–732, 1995. Citado nas páginas 25, 28 e 29.
- HASLBECK, J.; WALDORP, L. J. mgm: Estimating time-varying mixed graphical models in high-dimensional data. **arXiv preprint arXiv:1510.06871**, 2015. Citado na página 16.
- LAURITZEN, S. L. **Graphical models**. [S.l.]: Clarendon Press, 1996. v. 17. Citado na página 20.
- LETAC, G.; MASSAM, H. Wishart distributions for decomposable graphs. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 35, n. 3, p. 1278–1323, 2007. Citado na página 21.
- MITSAKAKIS, N.; MASSAM, H.; ESCOBAR, M. D. A metropolis-hastings based method for sampling from the g-wishart distribution in gaussian graphical models. **Electronic Journal of Statistics**, Institute of Mathematical Statistics and Bernoulli Society, v. 5, p. 18–30, 2011. Citado na página 24.

- MOHAMMADI, A.; WIT, E. C. Bayesian structure learning in sparse gaussian graphical models. **Bayesian Analysis**, International Society for Bayesian Analysis, v. 10, n. 1, p. 109–138, 2015. Citado na página 16.
- MUIRHEAD, R. J. **Aspects of multivariate statistical theory**. [S.l.]: John Wiley & Sons, 2009. Citado na página 21.
- NOWICKI, K.; SNIJDERS, T. A. B. Estimation and prediction for stochastic blockstructures. **Journal of the American statistical association**, Taylor & Francis, v. 96, n. 455, p. 1077–1087, 2001. Citado nas páginas 15 e 16.
- ROVERATO, A. Hyper inverse wishart distribution for non-decomposable graphs and its application to bayesian inference for gaussian graphical models. **Scandinavian Journal of Statistics**, Wiley Online Library, v. 29, n. 3, p. 391–411, 2002. Citado nas páginas 21 e 22.
- WANG, H.; LI, S. Z. Efficient gaussian graphical model determination under g-wishart prior distributions. **Electronic Journal of Statistics**, Institute of Mathematical Statistics and Bernoulli Society, v. 6, p. 168–198, 2012. Citado na página 31.
- WHITTAKER, J. **Graphical Models in Applied Multivariate Statistics**. Chichester: John Wiley and Sons, 1990. Citado nas páginas 16, 20, 26 e 101.

