

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

JOSÉ VINÍCIUS DE MIRANDA CARDOSO

**Estimação de Grafos não Direcionais
Aplicada à Aprendizagem de Máquina**

Campina Grande
2019

JOSÉ VINÍCIUS DE MIRANDA CARDOSO

ESTIMAÇÃO DE GRAFOS NÃO DIRECIONAIS APLICADA À
APRENDIZAGEM DE MÁQUINA

Relatório de Projeto apresentado ao Departamento de Engenharia Elétrica da Universidade Federal de Campina Grande como parte dos requisitos exigidos para a obtenção do título de Bacharel em Engenharia Elétrica.

Orientador: Prof. Dr. Wamberto José Lira de Queiroz

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DO RELATÓRIO DE PROJETO DEFENDIDA PELO ALUNO JOSÉ VINÍCIUS DE MIRANDA CARDOSO, E ORIENTADA PELO PROF. DR. WAMBERTO JOSÉ LIRA DE QUEIROZ

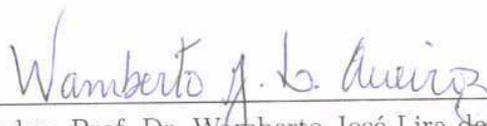
Campina Grande
2019

Comissão Julgadora – Trabalho de Conclusão de Curso

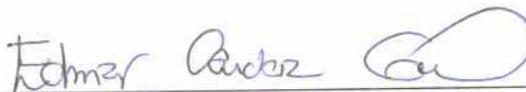
Candidato: José Vinícius de Miranda Cardoso

Data da defesa: 13 de Agosto de 2019

Título: “Estimação de Grafos não Direcionais Aplicada à Aprendizagem de Máquina.”



Orientador: Prof. Dr. Wamberto José Lira de Queiroz



Avaliador: Prof. Dr. Edmar Candeia Gurjão

Dedicatória

Este trabalho é dedicado a José Mineiro da Costa (Seu Tetêna), *in memoriam*, e a André Miranda Costa *in memoriam*.

Stay Hungry. Stay Foolish.
— Steve Paul Jobs

Don't let college get in the way of your education.
— Mark Twain

Agradecimentos

Eu gostaria de agradecer ao professor Wamberto por aceitar me orientar neste trabalho, sem o qual a qualidade desse trabalho seria bastante dirimida. Também expressei os meus agradecimentos ao professor Edmar por aceitar ser o membro da banca examinadora.

Eu agradeço aos professores e amigos Marcelo, Danilo, Paulo e Raissa pela paciência e atenção na transmissão dos conhecimentos tanto de cunho profissional quanto pessoal.

Eu agradeço demais aos amigos que essa jornada me deu e que tornaram ela cheia de risadas: José (Jorge) Igor, Jorge (Jorge) Henrique, Tullio (Tutu) Aureliano, Artur (Ceará) Freitas, Augusto (Gugu) César, Paulo (Paulim) Vinicius, Lucas Vinicius, Igor (Igão) Ulisses, Guilherme (Gui) Dantas, João André, Ícaro Torres, Pedro (Pedrão) Barros, André (Kiko) Augusto, Pedro (Ganso) Henrique, Felipe (Peux) Emanuel, Matheus (Souza) Souza, Hiago (Diniz) Maracajá, Adbys Vasconcelos, Gabriela (Gabi) Sobreira, Natália (Nat) Medeiros, João Roberto, Nathália (Nath) Moraes, Bruna Larissa, Ivana Crisóstomo, Felipe (Lipschutz) Trigueiro, Danilo Barreto, Fernando Ramalho, Josué Júnior, José Fidelis, Dionísio, Allender Alencar e tantos outros.

Agradeço aos professores do DEE, em especial: Saulo, Edgar, Damásio, Mário, Francisco Marcos e Alexandre. Muito obrigado pela paciência e atenção na transmissão dos conhecimentos.

Agradeço aos professores do antigo DME atuais DM e DE, em especial: Jefferson, Fernando, Aparecido, Denilson, Angelo, Zé Luis, Areli, Diogo e Iraponil.

Agradeço também aos amigos da atual iteração do LABCOM 3 do Iecom: Hugerles, Danilo e Dr. Hélio.

Meus agradecimentos aos funcionários da coordenação de graduação em Engenharia Elétrica, em especial Adail, Tchai, e professor Gutemberg por todo suporte prestado em todos os momentos.

It goes without saying that without the love, care, support, and companionship of my partner in crime, 杨雪, I would have never made it.

Meu muito obrigado a toda a minha família pelo amor diário. Aos meus pais Vanda e Cicéro, à minha irmã Vaneska e à alegria da casa, Pluto.

Resumo

Sistemas dinâmicos podem ser modelados como grafos não direcionais em que os vértices (ou nós) representam os dispositivos do sistema e as arestas representam as possíveis interconexões entre tais dispositivos. Após o advento das redes heterogêneas, que possibilitou a interconexão massiva de dispositivos distintos operando em diferentes protocolos, a análise de dados provenientes destes dispositivos se tornou uma tarefa indispensável para extração de informações sobre a estrutura de um determinado sistema dinâmico. Desta forma, este trabalho propõe um estimador para a matriz Laplaciana de grafos não direcionais com o intuito de inferir a estrutura e conseqüentemente as interconexões (correlações) entre os dispositivos envolvidos. Utilizou-se a abordagem Bayesiana para considerar as características espectrais da matriz Laplaciana no projeto do estimador. O estimador proposto foi aplicado em tarefas de clusterização de dados que está inserida na área de aprendizagem de máquina não supervisionada. Os resultados obtidos indicam que o estimador possui um desempenho satisfatório, correspondente aos resultados obtidos por algoritmos de clusterização espectral publicados na literatura.

Abstract

Dynamical systems may be modelled as undirected graphs in which the nodes represent the devices of the system and the edges represent possible connections amongst them. After the rapid development of heterogeneous networks, which provided a massive interconnection for different types of devices, the analysis of data outputted from these devices has become a paramount task to understand the structure underpinning such dynamical systems. In this way, this work proposes an estimator for the Laplacian matrix of undirected graphs in order to infer the structure and thus the interconnections (correlations) amongst the devices. We have used the Bayesian approach in order to take into account the spectral properties of the Laplacian matrix into the project of the estimator. The proposed estimator was applied in data clustering, which is a subfield of unsupervised machine learning. The experimental results indicate that the estimator's performance is satisfactory, and it is expected when compared with results available in the literature.

Lista de Figuras

2.1	Ilustração de grafo não direcional. Autoria própria.	16
2.2	Um grafo com três componentes e seus respectivos autovalores que exemplificam a propriedade de igualdade entre a multiplicidade do autovalor $\lambda_1 = 0$ e o número de componentes de um grafo.	17
3.1	Grafo circular com duas componentes.	26
3.2	Grafo circular com três componentes.	27
3.3	Grafo em formato de semi-círculos com duas componentes.	27

Lista de Símbolos

Θ matriz de precisão (inversa da matriz de covariância)

w vetor que contém os pesos das arestas de um grafo

\mathcal{L} operador Laplaciano

\mathcal{L}^* operador Laplaciano adjunto

L matriz Laplaciana de um grafo

W matriz de adjacência de um grafo

\mathcal{S}_{++}^p conjunto das matrizes simétricas positivas-definidas de ordem p

α parâmetro de regularização associado à esparsidade da matriz de precisão

β parâmetro de regularização associado à variações temporais da matriz de precisão

$\det \mathbf{A}$ determinante da matriz quadrada \mathbf{A}

$\text{gdet } \mathbf{A}$ determinante generalizado (pseudo-determinante) da matriz quadrada \mathbf{A}

$\log \det \mathbf{A}$ logaritmo determinante da matriz quadrada \mathbf{A}

\mathbf{A}^\dagger pseudo inversa da matriz \mathbf{A}

$a \propto b$ a é proporcional a b a menos de uma constante aditiva

$D_{\text{KL}}(P||Q)$ divergência de Kullback-Leibler (entropia relativa) entre as distribuições P e Q

Sumário

1	Introdução	11
1.1	Objetivos	12
1.2	Metodologia	12
2	Fundamentação Teórica	14
2.1	Notação	14
2.2	Álgebra Linear	14
2.2.1	Decomposição Espectral	15
2.3	Teoria dos grafos	15
2.4	Processo Estocástico Gaussiano	17
2.5	Teoria da Estimação	18
2.5.1	Estimação de Máxima Verossimilhança	19
2.5.2	Estimação de Máxima Probabilidade <i>a Posteriori</i>	19
2.6	Trabalhos Relacionados	20
3	Algoritmo Proposto	22
3.1	Máxima Probabilidade <i>a Posteriori</i>	22
3.1.1	Iteração para U	23
3.1.2	Iteração para λ	24
3.1.3	Iteração para w	24
3.2	Resultados	25
4	Conclusão	28

Capítulo 1

Introdução

O tópico de processamento de sinais para *big data* tem atraído bastante atenção de diversas disciplinas científicas e de engenharia [1]. Tal interesse reside principalmente nos desafios inerentes aos sinais de *big data*, tais como, representação, processamento, armazenamento, estimação e interpretação [2]. Nota-se que tais desafios vão muito além do mero aumento da capacidade de processamento computacional. De fato, algoritmos inovadores juntamente com implementações eficientes estão sendo desenvolvidas pela comunidade científica com o intuito de solucionar tais desafios.

Mais recentemente, a Teoria dos Grafos tem sido sugerida como um *framework* matemático adequado para lidar com a alta heterogeneidade que é observada nos sinais de *big data* [3]. Grafos são estruturas matemáticas que podem ser utilizadas para representar sinais de uma forma bastante generalista, isto é, em uma estrutura irregular que engloba como casos particulares o domínio do tempo e o espaço Euclidiano. Ao representar sinais sobre grafos, pode-se retirar vantagens da fundação solidificada deste campo da matemática, além dos benefícios práticos principalmente sob a perspectiva de implementação computacional. Entretanto, apesar da Teoria dos Grafos ser um tópico clássico da matemática, com aplicações desde de *design* de circuitos à otimização combinatória, o seu potencial na área de processamento de sinais somente começou a ser investigado. Como mencionado em [4], a estimação de grafos, isto é, da matriz Laplaciana associada ao grafo, a partir de sinais mensurados ainda constitui um desafio em aberto.

Nessa linha de pesquisa, a maioria dos métodos propostos assume que a topologia do grafo é uma propriedade conhecida *a priori*, o que é uma suposição um tanto proibitiva em várias aplicações práticas [5]. De forma adicional, menciona-se em [4] que a estimação de grafos é um tópico bastante produtivo, visto que vários desafios ainda estão em aberto, tais como a estimação de grafos com propriedades espectrais específicas, isto é a distribuição dos autovalores e autovetores da matriz laplaciana.

Sendo assim, este trabalho de conclusão de curso investiga respostas para as seguintes questões de pesquisa:

1. É possível estimar a topologia de grafos em dimensões elevadas de forma eficiente?
2. Há estimadores práticos para tal tarefa?

Apesar da existência de vários métodos heurísticos para realizar o agrupamento de dados em grafos, e dessa forma estimar aproximadamente suas topologias, em geral tais métodos não são estatisticamente justificáveis, isto é, não são construídos a partir de uma suposição da distribuição estatística das observações, o que acarreta em algoritmos que não possuem garantia de convergência ou ainda em formulações que são intratáveis computacionalmente [6].

1.1 Objetivos

Este trabalho possui o seguinte objetivo geral

- propor um estimador da matriz Laplaciana de grafos não direcionais utilizando a abordagem Bayesiana.

O objetivo específico deste trabalho é

- aplicar o estimador proposto em tarefas de processamento de sinais e aprendizagem de máquina, tais como clusterização de dados.

1.2 Metodologia

A metodologia adotada para o desenvolvimento deste trabalho consiste no emprego de ferramentas matemáticas principalmente aquelas advindas dos tópicos de álgebra linear, probabilidade, estatística, otimização, teoria da estimação e teoria dos grafos. O principal desafio deste trabalho está relacionado a grande quantidade de variáveis a serem estimadas. Dessa forma, a metodologia deste trabalho também emprega o uso de ferramentas computacionais eficientes como a linguagem de programação C++ e a biblioteca de álgebra linear Eigen3. Do ponto de vista prático, os estimadores propostos foram implementados em software de tal forma que estes possam ser portáteis para ambientes computacionais de relativo baixo custo, tais como computadores pessoais. Sob a perspectiva experimental, serão utilizados sinais sintéticos, isto é, simulados, para avaliar o desempenho dos estimadores propostos.

O restante deste relatório está organizado da seguinte maneira: o Capítulo 2 apresenta a notação adotada ao longo do texto, uma revisão dos principais tópicos matemáticos empregados para o projeto do estimador tais como: álgebra linear, teoria dos grafos, processos estocásticos e teoria da estimação, assim como uma revisão bibliográfica de alguns trabalhos relacionados. O Capítulo 3 apresenta o desenvolvimento do projeto do

estimador proposto e os resultados de simulação obtidos. Por fim, o Capítulo 4 discute as conclusões e apresenta algumas propostas para trabalhos futuros.

Capítulo 2

Fundamentação Teórica

Este capítulo discorre sobre os tópicos básicos da matemática que fornecem a base necessária para o entendimento dos trabalhos relacionados à estimação de grafos não direcionais, bem como dos algoritmos e estimadores propostos.

2.1 Notação

Escalares e variáveis aleatórias são representadas como letras minúsculas em itálico, por exemplo, x . A função densidade de probabilidade (muitas vezes referida como densidade de probabilidade, ou somente densidade) de uma variável aleatória x é representada como p_x ; em alguns casos em que o contexto é claro iremos omitir o subscrito x . Conjuntos, subespaços vetoriais e eventos são denotados por letras maiúsculas em itálico, por exemplo A . Matrizes são representadas por letras maiúsculas, em itálico e negrito, por exemplo \mathbf{A} . A inversa (generalizada) de uma matriz quadrada positiva definida (positiva semidefinida) é denotada como \mathbf{A}^{-1} (\mathbf{A}^\dagger).

2.2 Álgebra Linear

Algumas definições de álgebra linear são empregados nesse texto. Elas são listadas aqui para auxiliar a leitura dos resultados e desenvolvimentos apresentados no Capítulo 3. Essas definições também podem ser encontradas em livros introdutórios de matemáticas tais como [7].

Definição 1 – matriz simétrica: uma matriz quadrada $\mathbf{A} \in \mathbb{R}^{n \times n}$ é simétrica se $\mathbf{A} = \mathbf{A}^\top$.

Definição 2 – determinante: o determinante de uma matriz quadrada $\mathbf{A} \in \mathbb{R}^{n \times n}$, denotado por $\det \mathbf{A}$, é definido como o produto dos autovalores de \mathbf{A} .

Definição 3 – traço: o traço de uma matriz quadrada $\mathbf{A} \in \mathbb{R}^{n \times n}$, denotado por $\text{Tr} \mathbf{A}$, é definido como a soma dos valores da diagonal principal de \mathbf{A} .

Definição 4 – pseudo determinante: o pseudo determinante, também referido como determinante generalizado de uma matriz quadrada $\mathbf{A} \in \mathbb{R}^{n \times n}$, denotado por $\text{gdet}\mathbf{A}$, é definido como o produto dos autovalores não nulos de \mathbf{A} .

Definição 5 – matriz positiva definida: uma matriz quadrada $\mathbf{A} \in \mathbb{R}^{n \times n}$ é positiva definida quando para qualquer vetor $\mathbf{x} \in \mathbb{R}^n$, em que pelo menos uma das coordenadas de \mathbf{x} é não nula, vale $\mathbf{x}^\top \mathbf{A} \mathbf{x} > 0$.

Definição 6 – matriz positiva semidefinida: uma matriz quadrada $\mathbf{A} \in \mathbb{R}^{n \times n}$ é positiva semidefinida quando ela é positiva definida e $\mathbf{x}^\top \mathbf{A} \mathbf{x} \geq 0$.

Definição 7 – operador: um operador \mathbf{A} que mapeia um vetor \mathbf{x} de um subespaço vetorial V em outro vetor \mathbf{y} de um subespaço U (com possivelmente $U = V$) é definido como $\mathbf{y} = \mathbf{A} \mathbf{x}$.

Definição 8 – operador linear: um operador é dito linear se

$$\mathbf{A}(\alpha \mathbf{x}_1 + \beta \mathbf{x}_2) = \alpha \mathbf{A} \mathbf{x}_1 + \beta \mathbf{A} \mathbf{x}_2, \text{ em que } \alpha \in \mathbb{R}, \beta \in \mathbb{R}.$$

Definição 9 – rank: o rank de uma matriz quadrada \mathbf{A} , representado como $\text{rank}(\mathbf{A})$, é o número de vetores coluna linearmente independentes de \mathbf{A} , ou ainda, é o número de autovalores não nulos de \mathbf{A} .

2.2.1 Decomposição Espectral

A decomposição espectral de uma matriz diagonalizável \mathbf{A} corresponde à fatorização de \mathbf{A} da seguinte maneira

$$\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^{-1}, \quad (2.1)$$

em que \mathbf{U} é a matriz que contém os autovetores de \mathbf{A} em suas colunas e $\mathbf{\Lambda}$ é uma matriz diagonal em que Λ_{ii} representa o i -ésimo autovalor de \mathbf{A} .

No caso em que a matriz \mathbf{A} é real e simétrica, temos que $\mathbf{U}^{-1} = \mathbf{U}^\top$, logo

$$\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^\top. \quad (2.2)$$

Por meio da decomposição espectral de \mathbf{A} é possível calcular a sua inversa como

$$\mathbf{A}^{-1} = \mathbf{U} \mathbf{\Lambda}^{-1} \mathbf{U}^\top. \quad (2.3)$$

2.3 Teoria dos grafos

Grafos são estruturas matemáticas empregadas para a representação de conexões entre sistemas dinâmicos [8]. A teoria dos grafos tem sido aplicada nas mais diversas áreas da ciência, tecnologia e engenharia. Por exemplo, no campo da eletrônica, grafos são utilizados para modelar as interações entre os elementos de um circuito [9].

Matematicamente, um grafo não direcionado \mathcal{G} pode ser definido como uma tripla ordenada da forma $(V(\mathcal{G}), E(\mathcal{G}), \mathbf{W})$, em que $V(\mathcal{G})$ é um conjunto não vazio de vértices, $E(\mathcal{G})$ é um conjunto de pares não ordenados que denotam as arestas do grafo, e \mathbf{W} é a matriz de adjacência definida por

$$\mathbf{W} \triangleq \begin{cases} W_{ij} \in \mathbb{R}_+, & \text{se } (i, j) \in E(\mathcal{G}), \\ 0, & \text{caso contrário.} \end{cases} \quad (2.4)$$

Exemplo 1: o grafo da Figura 2.1 possui a seguinte representação matemática:

$V(\mathcal{G}) = \{x_i | i = 1, 2, \dots, 4\}$, $E(\mathcal{G}) = \{(x_i, x_j) | i = 1, 2, \dots, 4; j = 1, 2, \dots, 4\}$ e

$$\mathbf{W} = \begin{bmatrix} 0 & 0.9 & 0.27 & 0.37 \\ 0.9 & 0 & 0.57 & 0.91 \\ 0.27 & 0.57 & 0 & 0.2 \\ 0.37 & 0.91 & 0.2 & 0 \end{bmatrix} \quad (2.5)$$

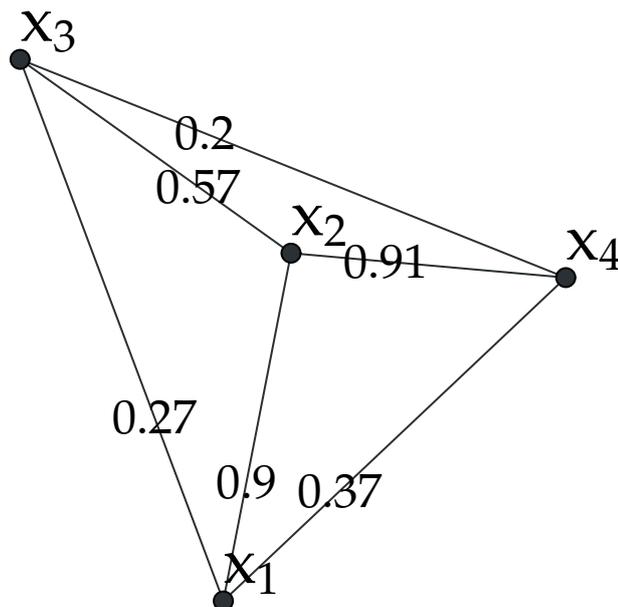


Figura 2.1: Ilustração de grafo não direcional. Autoria própria.

Por definição, observam-se as seguintes propriedades da matriz de adjacência \mathbf{W} : $\mathbf{W} = \mathbf{W}^T$ (simetria) e $\det \mathbf{W} = 0$ (singular).

Uma outra matriz importante em Teoria dos Grafos é a matriz Laplaciana \mathbf{L} definida como

$$\mathbf{L} \triangleq \mathbf{D} - \mathbf{W}, \quad (2.6)$$

em que \mathbf{D} é uma matriz diagonal tal que $D_{ii} = \mathbf{1}^T \mathbf{W}_i$, em que $\mathbf{1}$ é o vetor cujos elementos são 1 e \mathbf{W}_i é a i -ésima linha de \mathbf{W} .

Por exemplo, a matriz Laplacian para o grafo da Figura 2.1 é

$$\mathbf{L} = \begin{bmatrix} 1.54 & -0.9 & -0.27 & -0.37 \\ -0.9 & 2.38 & -0.57 & -0.91 \\ -0.27 & -0.57 & 1.04 & -0.2 \\ -0.37 & -0.91 & -0.2 & 1.48 \end{bmatrix} \quad (2.7)$$

A matriz Laplaciana é utilizada para modelar a matrix inversa de covariância (também denominada matrix de precisão) [10]. Assim como a matriz de adjacência, a matriz Laplaciana é simétrica e singular. Os autovalores de \mathbf{L} , $\lambda_1, \lambda_2, \dots, \lambda_n$, possuem a seguinte propriedade

$$\lambda_1 = 0, \leq \lambda_2, \leq \dots, \leq \lambda_n, \quad (2.8)$$

o que implica que \mathbf{L} é positiva semidefinida. Além disso, a multiplicidade do autovalor $\lambda_1 = 0$ indica quantidade de componentes do grafo.

As Figuras (2.2a) e (2.2b) ilustram um grafo que contém três componentes e os respectivos autovalores da matriz Laplaciana \mathbf{L} . Como pode ser observado, a multiplicidade do autovalor $\lambda_1 = 0$ é de fato igual a 3.

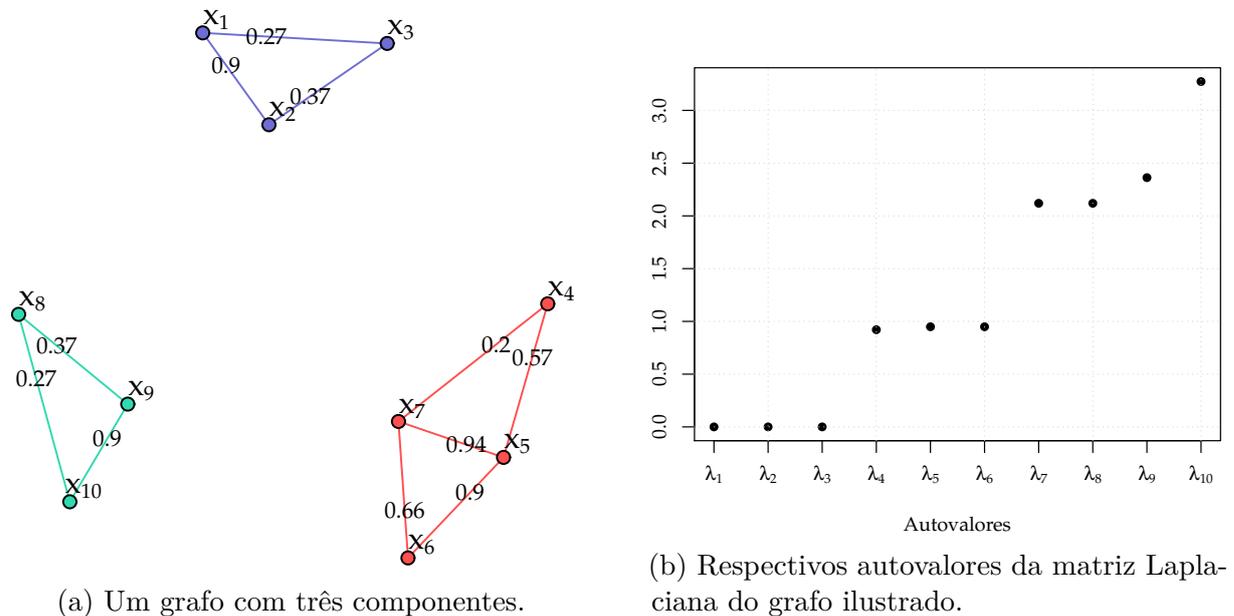


Figura 2.2: Um grafo com três componentes e seus respectivos autovalores que exemplificam a propriedade de igualdade entre a multiplicidade do autovalor $\lambda_1 = 0$ e o número de componentes de um grafo.

2.4 Processo Estocástico Gaussiano

Um processo estocástico contínuo em tempo discreto pode ser definido como um vetor $\mathbf{x} = [x_1, x_2, \dots, x_n]^\top$ em que a i -ésima componente x_i é uma variável aleatória com

densidade de probabilidade p_{x_i} .

A densidade de probabilidade de \mathbf{x} é definida como sendo a densidade conjunta entre suas componentes, isto é

$$p_{\mathbf{x}}(x_1, x_2, \dots, x_n) \triangleq p_{x_1, x_2, \dots, x_n}(x_1, x_2, \dots, x_n). \quad (2.9)$$

Um processo estocástico é dito Gaussiano quando sua densidade de probabilidade pode ser expressa como

$$p_{\mathbf{x}}(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} [\det \boldsymbol{\Sigma}]^{-\frac{1}{2}} \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right), \quad (2.10)$$

em que $\boldsymbol{\mu}$ denota a média do processo, isto é, $\boldsymbol{\mu} \triangleq \mathbb{E}(\mathbf{x})$, e $\boldsymbol{\Sigma}$ denota a matriz de covariância do processo, isto é, $\Sigma_{ij} \triangleq \text{cov}(x_i, x_j)$.

2.5 Teoria da Estimação

A teoria da estimação concerne o projeto de estimadores, isto é, funções de variáveis aleatórias observáveis, para a extração de informações sobre parâmetros de um determinado sistema dinâmico. Em geral, os parâmetros de interesse não são mensuráveis ou sequer observáveis diretamente. Assim, tais parâmetros devem ser estimados a partir de medições de um fenômeno aleatório. Na teoria de estimação paramétrica, a distribuição de probabilidade do fenômeno observável é parametrizada como $P_{\boldsymbol{\theta}}$, em que $\boldsymbol{\theta}$ é o vetor de parâmetros de interesse.

Matematicamente, um estimador g para um vetor de parâmetros $\boldsymbol{\theta} \in \mathbb{R}^p$ é uma função, $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$, somente do vetor de variáveis observáveis $\mathbf{y} \in \mathbb{R}^n$, de tal forma que g forneça uma “boa aproximação” para o valor verdadeiro, não observável, de $\boldsymbol{\theta}$. Apesar de considerarmos os parâmetros $\boldsymbol{\theta}$ e as observações \mathbf{y} como sendo vetores, a teoria também é aplicável para o caso em que $\boldsymbol{\theta}$ e \mathbf{y} são matrizes ou ainda tensores.

Ilustrativamente, suponha que em um sistema de comunicações um receptor measure o seguinte sinal

$$\mathbf{y} = \theta + \boldsymbol{\eta}, \quad (2.11)$$

em que $\theta \in \mathbb{R}$ é a amplitude do sinal transmitido, $\boldsymbol{\eta} \in \mathbb{R}^n$ é o ruído térmico no receptor e $\mathbf{y} \in \mathbb{R}^n$ é o sinal recebido.

Sendo assim, suponha que o receptor tenha o interesse em estimar o valor de θ com base nas observações mensuradas \mathbf{y} . Frequentemente, assume-se que as componentes de $\boldsymbol{\eta}$ são independentes e identicamente distribuídas com distribuição Gaussiana de média

nula e variância σ^2 , isto é $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. Logo, a densidade de probabilidade de $\boldsymbol{\eta}$ é

$$p_{\boldsymbol{\eta}}(\boldsymbol{\eta}) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n \eta_i^2\right) \quad (2.12)$$

Dessa maneira, assumindo que θ é um parâmetro determinístico, temos que o sinal recebido \mathbf{y} possui distribuição Gaussiana com média $\mathbb{E}(\mathbf{y}) = \mathbb{E}(\theta + \boldsymbol{\eta}) = \theta$ e variância $\text{var}(\mathbf{y}) = \text{var}(\theta + \boldsymbol{\eta}) = \sigma^2$, isto é

$$p(\mathbf{y}) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \theta)^2\right). \quad (2.13)$$

De posse da densidade de probabilidade do sinal recebido \mathbf{y} , existem basicamente duas abordagens práticas para realizar a estimação do parâmetro θ : (1) estimador de máxima verossimilhança e (2) estimador de máxima probabilidade *a posteriori*.

2.5.1 Estimação de Máxima Verossimilhança

O estimador de máxima verossimilhança, do inglês, *Maximum Likelihood Estimator* (MLE), é definido como

$$g_{MLE}(\mathbf{y}) \triangleq \arg \max_{\theta} p_{\theta}(\mathbf{y}), \quad (2.14)$$

isto é, o MLE retorna o valor de θ para o qual a densidade de probabilidade da observação \mathbf{y} é máxima.

Aplicando o MLE ao exemplo da seção anterior temos que

$$g_{MLE}(\mathbf{y}) = \arg \max_{\theta} \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \theta)^2\right), \quad (2.15)$$

$$g_{MLE}(\mathbf{y}) = \arg \min_{\theta} \sum_{i=1}^n (y_i - \theta)^2, \quad (2.16)$$

$$g_{MLE}(\mathbf{y}) = \frac{1}{n} \sum_{i=1}^n y_i, \quad (2.17)$$

em que a igualdade (2.16) segue após a aplicação da função $\log(\cdot)$, visto que \log é uma função monotonicamente crescente; a igualdade (2.17) segue como solução única da equação $\frac{\partial}{\partial \theta} \log p_{\theta}(\mathbf{y}) = 0$.

2.5.2 Estimação de Máxima Probabilidade *a Posteriori*

Em várias aplicações, existe alguma informação *a priori* sobre o parâmetro de interesse. Dessa forma, a abordagem Bayesiana é a maneira natural de incorporar tal informação no projeto do estimador. Suponha então que o parâmetro de interesse é uma variável aleatória com densidade de probabilidade $p(\boldsymbol{\theta})$. Sendo assim, o estimador de

Máxima Probabilidade *a Posteriori*, g_{MAP} , é definido como

$$g_{MAP}(\mathbf{y}) \triangleq \arg \max_{\boldsymbol{\theta} \in \mathcal{S}_{\Theta}} p(\mathbf{y}, \boldsymbol{\theta}) = \arg \max_{\boldsymbol{\theta} \in \mathcal{S}_{\Theta}} p(\boldsymbol{\theta}|\mathbf{y}), \quad (2.18)$$

em que \mathcal{S}_{Θ} é o suporte da densidade de probabilidade *a priori*, $p(\boldsymbol{\theta})$, isto é, $\mathcal{S}_{\Theta} = \{\boldsymbol{\theta} : p(\boldsymbol{\theta}) > 0\}$.

2.6 Trabalhos Relacionados

No ano de 2008, em [11], Friedman apresentou o algoritmo denominado *graphical lasso* que resolve o seguinte problema de máxima verossimilhança

$$\arg \min_{\Theta} \text{Tr}(\mathbf{S}\Theta) + \rho \|\Theta\|_1 - \log \det \Theta. \quad (2.19)$$

Na formulação de Friedman, a esparsidade da matriz de precisão estimada Θ é incluída por meio do termo $\rho \|\Theta\|_1$, em que $\|\cdot\|_1$ representa a norma ℓ_1 , ρ é um parâmetro de regularização que controla o nível de esparsidade desejado em Θ . O parâmetro ρ é usualmente ajustado de forma empírica. A matriz \mathbf{S} representa a matriz de covariância amostral, dada por $\mathbf{S} = \frac{1}{n} \mathbf{X}\mathbf{X}^T$.

O algoritmo de Friedman, comumente conhecido como *graphical lasso* (ou *glasso*), foi amplamente aceito na comunidade científica e encontrou aplicações nos mais variados campos da ciência, engenharia e computação. Por exemplo, em [12], o algoritmo *glasso* foi aplicado para estimar as conexões entre neurônios com base em dados de séries temporais provenientes de medições realizadas por meio da técnica de imageamento de ressonância magnética funcional. Apesar da grande importância prática e teórica do algoritmo *glasso*, ele possui algumas limitações. Mais precisamente, o algoritmo *glasso* não permite a inclusão de informações *a priori* sobre, por exemplo, a estrutura da matriz de precisão Θ . Além disso, ele não admite uma representação de baixo *rank* de Θ , o que impede, por exemplo, que Θ modele a matriz Laplaciana de um grafo não direcional.

Mais recentemente, em 2017, Hallac *et al* [13] propôs um estimador que estende o algoritmo *glasso* ao incorporar um modelo em que a matriz de precisão Θ varia com o tempo. Mais precisamente, o estimador proposto por Hallac segue como a solução do seguinte problema de otimização

$$\begin{aligned} \underset{\Theta \in \mathcal{S}_{++}^p}{\text{argmin}} \quad & \sum_{i=1}^m n_i (-\log \det \Theta_i + \text{Tr}(\mathbf{S}_i \Theta_i)) + \alpha \|\Theta_i\|_1 + \\ & \beta \sum_{i=2}^m \psi(\Theta_i - \Theta_{i-1}), \end{aligned} \quad (2.20)$$

em que \mathcal{S}_{++}^p é o conjunto de matrizes simétricas positivas-definidas de ordem p , e α e β são parâmetros de regularização que controlam o nível de esparsidade e o nível de alterações

temporais na estrutura de Θ , respectivamente.

Similarmente ao algoritmo *lasso*, o algoritmo proposto por Hallac não incorpora informações estruturais disponíveis a priori sobre a matriz de precisão Θ .

Dessa forma, faz-se importante desenvolver estimadores que possam incorporar informações *a priori* sobre a estrutura da matriz Laplaciana, de modo que o estimador projetado seja o mais eficiente possível.

Capítulo 3

Algoritmo Proposto

Este capítulo discorre sobre o desenvolvimento do algoritmo proposto para estimação da matriz Laplaciana de grafos não direcionais.

3.1 Máxima Probabilidade *a Posteriori*

Dado um processo estocástico Gaussiano, como um vetor de variáveis aleatórias da forma $\mathbf{x} = [x_1, x_2, \dots, x_p]^\top$, $p > 1$, tal que $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \Theta^{-1})$, em que Θ é a matriz de precisão (inversa da matriz de covariância), a função densidade de probabilidade de \mathbf{x} é dada por

$$p(\mathbf{x}) = (2\pi)^{-\frac{p}{2}} \left[\det(\Theta^{-1}) \right]^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{x}^\top \Theta \mathbf{x}\right). \quad (3.1)$$

No caso em que a matriz de precisão venha a ser singular, a definição do operador determinante \det , isto é, o produto dos autovalores da matriz de precisão, é alterada para o produto dos autovalores não nulos da matriz de precisão, que é comumente referenciada na literatura como o pseudo determinante, ou ainda, determinante generalizado.

Dadas n realizações do processo \mathbf{x} , isto é, dada a matriz $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p]^\top$, com $\mathbf{x}_i \in \mathbb{R}^n$, o estimador de Máxima Probabilidade *a Posteriori* da matriz de precisão Θ é então dado por

$$g_{\text{MAP}}(\mathbf{X}) = \arg \max_{\Theta \in \mathcal{S}_\Theta} \log p(\mathbf{X}|\Theta) + \log p(\Theta), \quad (3.2)$$

em que $p(\Theta)$ é a distribuição *a priori* da matriz de precisão e \mathcal{S}_Θ é o conjunto das matrizes simétricas, positivas semi-definidas.

O termo $\log p(\mathbf{X})$ pode ser desenvolvido como

$$\log p(\mathbf{X}|\Theta) = -\frac{1}{2} \log \det(\Theta^{-1}) - \frac{1}{2} \mathbf{X}^\top \Theta \mathbf{X} + c_1, \quad (3.3)$$

$$\log p(\mathbf{X}|\Theta) \propto \log \det(\Theta) - \mathbf{X}^\top \Theta \mathbf{X} + c_2, \quad (3.4)$$

$$\log p(\mathbf{X}|\Theta) \propto \log \det(\Theta) - \text{Tr}(\Theta \mathbf{S}) + c_3 \quad (3.5)$$

em que $\mathbf{S} \triangleq \frac{1}{n} \mathbf{X} \mathbf{X}^\top$ é a matriz de covariância amostral, c_1, c_2 , e c_3 são constantes não relevantes para o problema de otimização, visto que elas não são funções de Θ . Além disso, usando o operador Laplaciano \mathcal{L} , temos que $\Theta = \mathcal{L} \mathbf{w}$, portanto

$$\log p(\mathbf{X} | \mathbf{w}) \propto \log \det(\mathcal{L} \mathbf{w}) - \text{Tr}(\mathcal{L} \mathbf{w} \mathbf{S}). \quad (3.6)$$

A densidade de probabilidade *a priori*, por sua vez, deve contemplar as propriedades da matriz de precisão. Neste trabalho, considera-se as propriedades espectrais, isto é, $\lambda_1 = \lambda_2 = \dots = \lambda_k = 0$ e a propriedade de esparsidade de Θ . Logo, escolheu-se a seguinte densidade de probabilidade *a priori*

$$p(\Theta) \propto \exp(-\alpha \|\mathcal{L} \mathbf{w}\|_1) \exp\left(-\beta D_{KL}\left(\mathcal{N}(\mathbf{0}, \mathcal{L} \mathbf{w}^\dagger), \mathcal{N}(\mathbf{0}, \mathbf{U}^\top [\text{Diag}(\boldsymbol{\lambda})]^{-1} \mathbf{U})\right)\right), \quad (3.7)$$

em que α e β são constantes de regularização, e D_{KL} denota a divergência de Kullback-Leibler dada por

$$D_{KL}\left(\mathcal{N}(\mathbf{0}, \mathcal{L} \mathbf{w}^\dagger), \mathcal{N}(\mathbf{0}, \mathbf{U}^\top [\text{Diag}(\boldsymbol{\lambda})]^{-1} \mathbf{U})\right) \propto -\log \text{gdet}(\mathcal{L} \mathbf{w}) + \log \det(\text{Diag}(\boldsymbol{\lambda})) + \text{Tr}\left(\mathbf{U} [\text{Diag}(\boldsymbol{\lambda})]^{-1} \mathbf{U}^\top \mathcal{L} \mathbf{w}\right). \quad (3.8)$$

Substituindo (3.6), (3.7) e (3.8) em (3.2), segue-se que

$$g_{\text{MAP}}(\mathbf{X}) = \arg \min_{\mathbf{w}, \mathbf{U}, \boldsymbol{\lambda}} - (1 + \beta) \log \det \mathcal{L} \mathbf{w} + \beta \log \det \text{Diag}(\boldsymbol{\lambda}) + \text{Tr}\left(\left(\mathbf{K} + \beta \mathbf{U} [\text{Diag}(\boldsymbol{\lambda})]^{-1} \mathbf{U}^\top\right) \mathcal{L} \mathbf{w}\right), \quad (3.9)$$

em que $\mathbf{K} \triangleq \mathbf{S} + \alpha(2\mathbf{I} - \mathbf{1}\mathbf{1}^\top)$, com as restrições que $\mathbf{w} > \mathbf{0}$, $\mathbf{U}^\top \mathbf{U} = \mathbf{I}$ e $0 \leq \lambda_2, \dots, \leq \lambda_p$.

Note que o número de componentes do grafo pode ser imposto de maneira arbitrária, considerando-se $\lambda_1 = \lambda_2, \dots, = \lambda_k = 0$.

Nesse caso, a solução de (3.9) é impraticável de forma analítica. A utilização de sistemas de otimização em *black box* (tais como MATLAB ou CVX) também não é viável para sistemas práticos.

Sendo assim, este trabalho propõe um algoritmo iterativo baseado no método *Block Successive Upper Bound Minimization* (BSUM) [14], em que as variáveis de otimização são atualizadas sequencialmente.

3.1.1 Iteração para \mathbf{U}

Mantendo as variáveis \mathbf{w} e $\boldsymbol{\lambda}$ fixas, temos o seguinte subproblema para \mathbf{U}

$$\mathbf{U}_{l+1} = \arg \min_{\mathbf{U}_l, \mathbf{U}_l^\top \mathbf{U}_l = \mathbf{I}} \text{Tr}\left(\mathbf{U}_l [\text{Diag}(\boldsymbol{\lambda})]^{-1} \mathbf{U}_l^\top \mathcal{L} \mathbf{w}\right). \quad (3.10)$$

Em [15], mostra-se que a solução para o problema (3.10) é a matriz de autovetores de $\mathcal{L}\mathbf{w}$ associada aos autovalores não nulos de $\mathcal{L}\mathbf{w}$.

3.1.2 Iteração para λ

Fixando as variáveis \mathbf{U} e \mathbf{w} , tem-se o seguinte subproblema para λ

$$\lambda_{l+1} = \arg \min_{\lambda_l} \log \det \text{Diag}(\lambda_l) + \text{Tr} \left(\mathbf{U} [\text{Diag}(\lambda_l)]^{-1} \mathbf{U}^\top \mathcal{L}\mathbf{w} \right), \quad (3.11)$$

com a restrição de que $0 = \lambda_1 = \dots = \lambda_k \leq \lambda_{k+1} \leq \dots \leq \lambda_p$.

Note que o termo $\log \det \text{Diag}(\lambda)$ pode ser escrito como

$$\log \det \text{Diag}(\lambda^t) = \sum_{i=k+1}^p \log \lambda_i. \quad (3.12)$$

Além disso, o termo $\text{Tr} \left(\mathbf{U} [\text{Diag}(\lambda)]^{-1} \mathbf{U}^\top \mathcal{L}\mathbf{w} \right)$ pode ser escrito como

$$\text{Tr} \left(\mathbf{U} [\text{Diag}(\lambda)]^{-1} \mathbf{U}^\top \mathcal{L}\mathbf{w} \right) = \text{Tr} \left([\text{Diag}(\lambda)]^{-1} \mathbf{U}^\top \mathcal{L}\mathbf{w}\mathbf{U} \right) = \sum_{i=k+1}^p \frac{z_i}{\lambda_i}, \quad (3.13)$$

em que $z_i \triangleq \left[\mathbf{U}^\top \mathcal{L}\mathbf{w}\mathbf{U} \right]_{ii}$.

Portanto, combinando (3.12) e (3.13) em 3.11, temos que

$$\lambda_{l+1} = \arg \min_{\lambda_l} \sum_{i=k+1}^p \log \lambda_{li} + \frac{z_i}{\lambda_{li}}. \quad (3.14)$$

Em [16], mostra-se que a solução para o problema (3.14), considerando que $z_k \leq z_{k+1} \leq \dots \leq z_p$, é dada por

$$\lambda_{l+1} = \mathbf{z}. \quad (3.15)$$

3.1.3 Iteração para \mathbf{w}

Fixando as variáveis \mathbf{U} e λ , temos o seguinte subproblema para \mathbf{w}

$$\arg \min_{\mathbf{w} \geq 0} -(1 + \beta) \log \text{gdet} \mathcal{L}\mathbf{w} + \text{Tr}(\mathbf{Q}\mathcal{L}\mathbf{w}), \quad (3.16)$$

em que $\mathbf{Q} \triangleq \mathbf{K} + \beta \mathbf{U} [\text{Diag}(\lambda)]^{-1} \mathbf{U}^\top$.

Note-se que o subproblema (3.16) é convexo, visto que $\log \text{gdet}$ é uma função côncava, Tr é uma função convexa, e o conjunto $\mathbf{w} \geq 0$ também é convexo. No entanto, uma expressão analítica não é admissível devido à restrição $\mathbf{w} \geq 0$.

Além disso, temos a seguinte identidade $\text{Tr}(\mathbf{Q}\mathcal{L}\mathbf{w}) = \langle \mathbf{w}, \mathcal{L}^* \mathbf{Q} \rangle$. Portanto,

segue-se que

$$\frac{\partial}{\partial \mathbf{w}} \text{Tr}(\mathbf{Q}\mathcal{L}\mathbf{w}) = \frac{\partial}{\partial \mathbf{w}} \langle \mathbf{w}, \mathcal{L}^* \mathbf{Q} \rangle = \mathcal{L}^* \mathbf{Q} \quad (3.17)$$

A derivada de $\log \text{gdet} \mathcal{L}\mathbf{w}$ é dada por [17]

$$\frac{\partial}{\partial \mathbf{w}} \log \text{gdet} \mathcal{L}\mathbf{w} = \mathcal{L}^* \mathcal{L}\mathbf{w}^\dagger. \quad (3.18)$$

Portanto, a atualização para \mathbf{w} é dada por

$$\mathbf{w}_{l+1} = \left(\mathbf{w}_l - \frac{1}{L} \left(\mathcal{L}^* \left(-(1 + \beta) \mathcal{L}\mathbf{w}_l^\dagger + \mathbf{Q} \right) \right) \right)^+, \quad (3.19)$$

em que $(\cdot)^+$ denota a operação $\max(0, \cdot)$ elemento à elemento, isto é, esta operação retorna 0 caso o elemento negativo, caso contrário mantém o elemento. Além disso, L é a constante de Lipschitz [1] de $\log \text{gdet} \mathcal{L}\mathbf{w}_l$. Na prática, utiliza-se um valor elevado para essa constante, por exemplo, $L = 10^6$ [1].

O algoritmo a seguir sumariza o procedimento iterativo para o estimador:

Data: $\mathbf{S}, \beta, \alpha$

Result: Estimativa da matriz Laplaciana $\mathcal{L}\mathbf{w}$

$$\mathbf{w}_0 = \mathcal{L}^{-1} \mathbf{S}, \mathbf{S} = \mathbf{U}_0 \text{Diag}(\boldsymbol{\lambda}_0) \mathbf{U}_0^\top$$

while não convergiu **do**

atualize $\mathbf{U}_{t+1} = f_1(\boldsymbol{\lambda}_t, \mathbf{w}_t);$
atualize $\boldsymbol{\lambda}_{t+1} = f_2(\mathbf{U}_{t+1}, \mathbf{w}_t);$
atualize $\mathbf{w}_{t+1} = f_3(\mathbf{w}_t, \mathbf{U}_{t+1}, \boldsymbol{\lambda}_{t+1});$

end

3.2 Resultados

Os dados de entrada do estimador são somente a matriz de covariância amostral, os parâmetros de regularização α e β , e o número de componentes da matriz Laplaciana do grafo. Dessa forma, com o intuito de validar o estimador proposto, gerou-se vértices no plano Euclidiano ao longo de círculos de raios 1 *u.a.*, 2 *u.a.*, e 3 *u.a.*, respectivamente. Em seguida, adicionou-se ruído gaussiano às coordenadas x e y dos vértices. A matriz de covariância amostral foi calculada como

$$\mathbf{S} = \frac{1}{n} \mathbf{X} \mathbf{X}^\top, \quad (3.20)$$

em que $\mathbf{X} \triangleq [\mathbf{x}^\top \mathbf{y}^\top]^\top$, $\mathbf{x} \in \mathbb{R}^p$ e $\mathbf{y} \in \mathbb{R}^p$ são vetores contendo as coordenadas espaciais (abscissa e ordenada) dos vértices, n é o número de colunas da matriz \mathbf{X} , isto é, $n = 2$, e p é o número de vértices do grafo.

As Figuras 3.1, 3.2 e 3.3 ilustram a aplicação do estimador proposto para a tarefa

de clusterização de dados. Para o caso das Figuras 3.1, 3.2 e 3.3, o número total de vértices é $p = 200$, $p = 300$ e $p = 200$, respectivamente, em que cada subgrafo possui cem vértices. Os parâmetros de regularização utilizados para o estimador foram $\beta = 10$ e $\alpha = 0.01$.

A partir da observação das Figuras 3.1, 3.2 e 3.3, nota-se que o desempenho do estimador proposto é satisfatório, no entanto, mais experimentos devem ser realizados com o intuito de aferir a performance do algoritmo em diferentes cenários de ruído, número total de vértices, número total de subgrafos, etc.

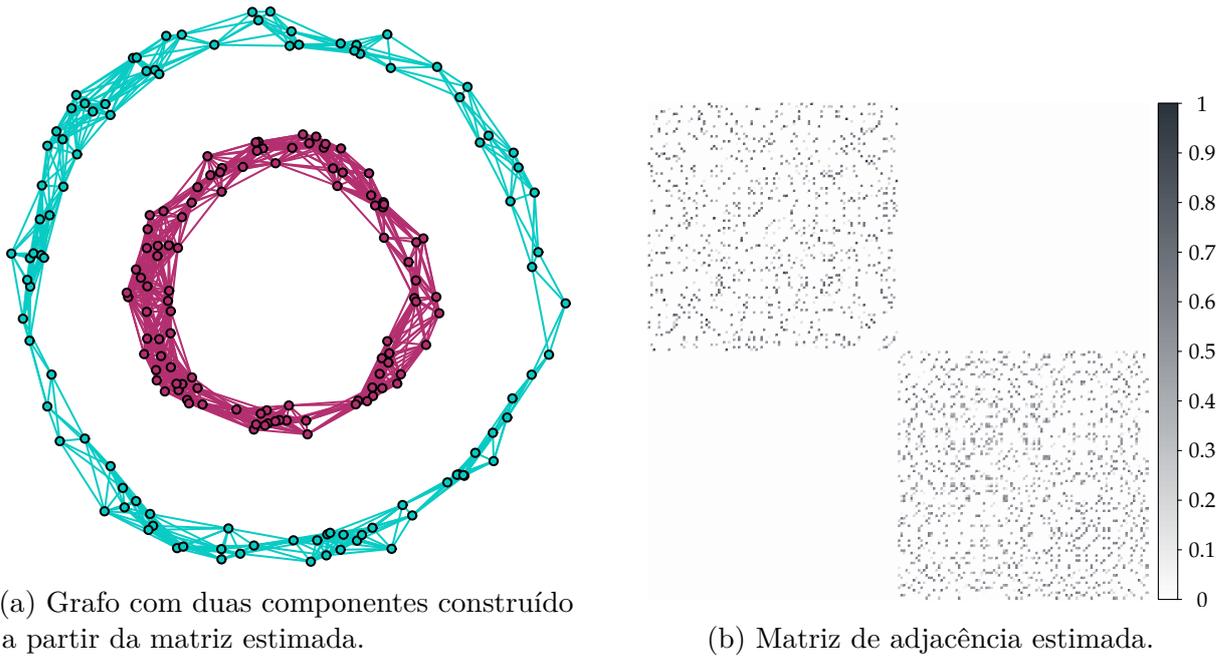
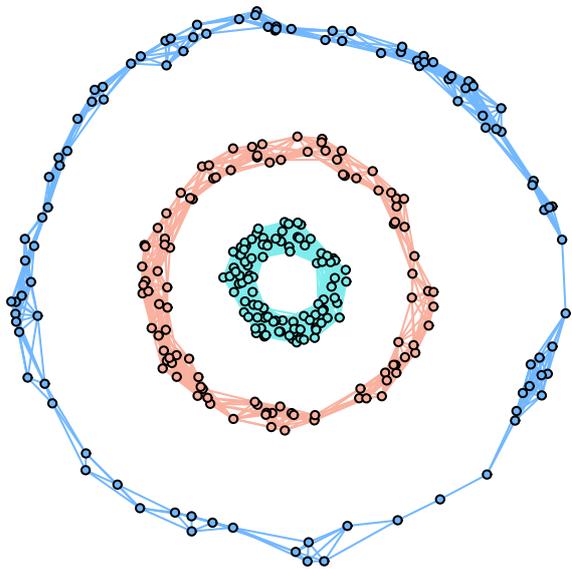
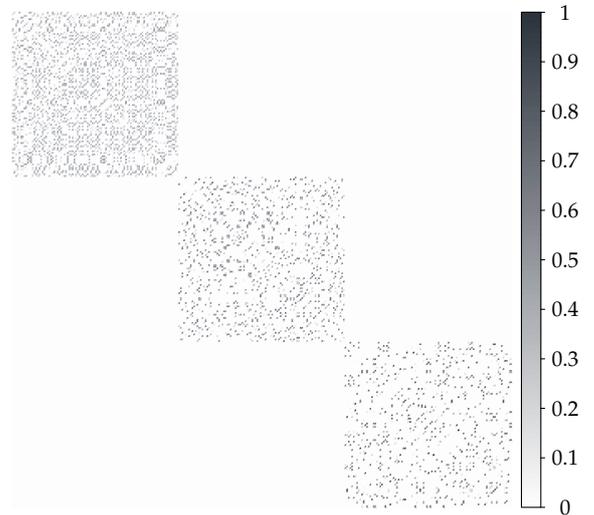


Figura 3.1: Grafo circular com duas componentes.

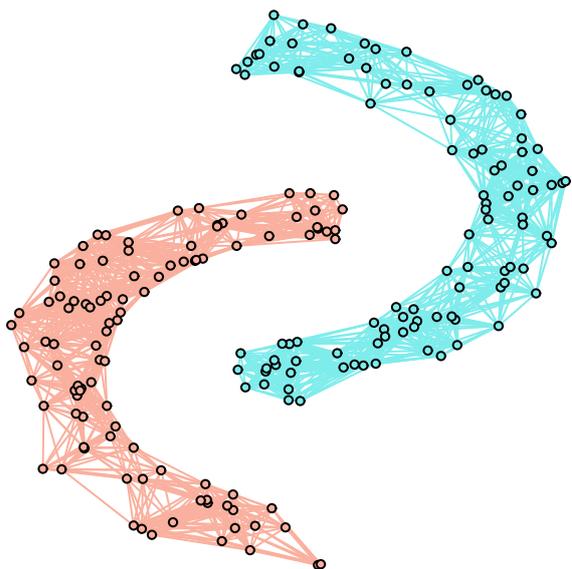


(a) Grafo com três componentes construído a partir da matriz estimada.

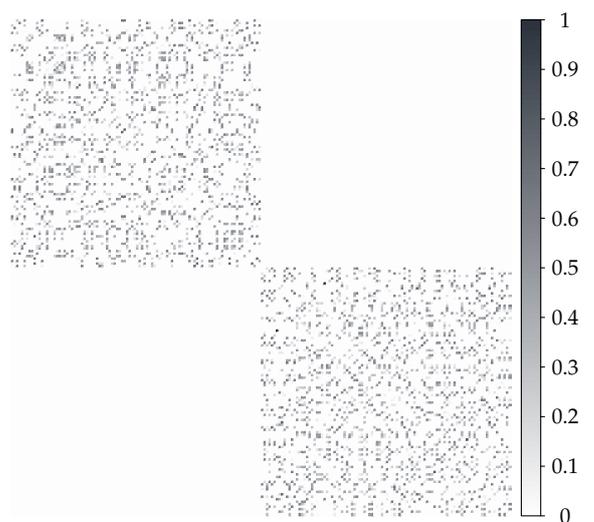


(b) Matriz de adjacência estimada.

Figura 3.2: Grafo circular com três componentes.



(a) Grafo com duas componentes construído a partir da matriz estimada.



(b) Matriz de adjacência estimada.

Figura 3.3: Grafo em formato de semi-círculos com duas componentes.

Capítulo 4

Conclusão

Neste trabalho foi proposto um estimador Bayesiano para a matriz Laplaciana de um grafo não direcional. O estimador proposto é baseado no método de Máxima Probabilidade *a Posteriori*, e foi projeto com o intuito de incluir informações *a priori* sobre a estrutura da matriz Laplaciana do grafo a ser estimado. Por fim, utilizou-se o algoritmo de solução por aproximações sucessivas para resolver o problema de otimização inerente ao projeto do estimador. Como ilustração prática, o estimador proposto foi aplicado em tarefas de aprendizagem de máquina não supervisionado, mais precisamente em clusterização de dados.

O autor deste trabalho prospecta as seguintes possibilidades para trabalhos futuros:

- Estudar a inclusão de diferentes formatos da distribuição *a priori* sobre a matriz Laplaciana e investigar o impacto de tal escolha no desempenho do estimador.
- Estudar técnicas para identificação automática do número de *clusters* (isto é, o número de componentes do grafo).
- Investigar possibilidades de otimização da implementação dos algoritmos propostos por meio do uso de bibliotecas de álgebra linear e otimização.

Referências

- [1] Y. Sun, P. Babu, and D. P. Palomar, “Majorization-minimization algorithms in signal processing, communications, and machine learning,” *IEEE Transactions on Signal Processing*, vol. 65, pp. 794–816, Feb 2017.
- [2] Al-Fuqaha A., Guizani M., Mohammadi M., Aledhari M., and Ayyash M., “Internet of things: A survey on enabling technologies, protocols, and applications,” *IEEE Communications Surveys and Tutorials*, 2015.
- [3] Shuman D. I., Narang S. K., Frossard P., and Ortega A., “The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains,” *IEEE Signal Processing Magazine*, 2013.
- [4] A. Ortega, F. P., J. Kovacevic, and M. J. M. F., “Graph signal processing: Overview, challenges, and applications,” *Proceedings of the IEEE*, 2018.
- [5] T. D. A. and M. G., “Estimation of graphical models through structured norm minimization,” *Journal of Machine Learning Research*, vol. 18, pp. 7692–7739, Jan. 2017.
- [6] M. Mahajan, P. Nimbhorkar, and K. Varadarajan, “The planar k-means problem is np-hard,” *Theoretical Computer Science*, vol. 442, pp. 13 – 21, 2012. Special Issue on the Workshop on Algorithms and Computation (WALCOM 2009).
- [7] Kolmogorov A. N. and Fomin S. V., *Introductory Real Analysis*. Dover, 1975.
- [8] Douglas B. West, *Introduction to Graph Theory*. Pearson, 2nd ed., September 2000.
- [9] J. A. Bondy and U. S. R. Murty, *Graph Theory with Applications*. Elsevier Science Publishing Co., Inc., 1st ed., 1976.
- [10] Hilmi E. Elgimez, Eduardo Pavez, and Antonio Ortega, “Graph Learning from Data under Structural and Laplacian Constraints,” *IEEE Journal of Selected Topics in Signal Processing*, 2017.
- [11] J. Friedman, T. Hastie, and R. Tibshirani, “Sparse inverse covariance estimation with the lasso,” *Biostatistics*, 2008.

-
- [12] S. M. Smith, K. L. Miller, G. Salimi-Khorshidi, M. Webster, C. F. Beckmann, T. E. Nichols, J. D. Ramsey, and M. W. Woolrich, “Network modelling methods for fmri,” *NeuroImage*, vol. 54, no. 2, pp. 875 – 891, 2011.
- [13] Hallac D., Park Y, Boyd S., and Leskovec J., “Network inference via the time-varying graphical lasso,” in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, pp. 205–213, ACM, 2017.
- [14] M. Hong, T-H. Chang, X. Wang, M. Razaviyayn, S. Ma, and Z-Q. Luo, “A block successive upper bound minimization method of multipliers for linearly constrained convex optimization,” *ArXiv*, 2014.
- [15] P. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 1st ed., 2008.
- [16] K. Benidis, Y. Sun, P. Babu, and D. P. Palomar, “Orthogonal sparse pca and covariance estimation via procrustes reformulation,” *IEEE Transactions on Signal Processing*, vol. 64, pp. 6211–6226, Dec 2016.
- [17] Holbrook A., “Differentiating the pseudo determinant,” *arXiv e-prints*, February 2018.