



RAFAEL POLLI CARNEIRO

Teoremas limites para a homologia persistente e uma aplicação

PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da CAPES e da UFABC

Santo André, 12 de março de 2024



Universidade Federal do ABC

Universidade Federal do ABC

Centro de Matemática, Computação e Cognição

Rafael Polli Carneiro

Teoremas limites para a homologia persistente e uma aplicação

Orientador: Prof. Dr. Cristian Favio Coletti

Coorientador: Prof. Dr. Daniel Miranda Machado

Tese de doutorado apresentada ao Centro de
Matemática, Computação e Cognição para
obtenção do título de Doutor em Matemática

ESTA É A VERSÃO FINAL DA TESE COM AS
REVISÕES LEVANTADAS PELA BANCA EXAMINADORA.

Santo André, 12 de março de 2024

Sistema de Bibliotecas da Universidade Federal do ABC
Elaborada pelo Sistema de Geração de Ficha Catalográfica da UFABC
com os dados fornecidos pelo(a) autor(a).

Polli Carneiro, Rafael

Teoremas limites para a homologia persistente e uma
aplicação / Rafael Polli Carneiro. — 2024.

192 fls. : il.

Orientador: Cristian Favio Coletti

Coorientador: Daniel Miranda Machado

Tese (Doutorado) — Universidade Federal do ABC,
Programa de Pós-Graduação em Matemática, Santo André,
2024.

1. Homologia Persistente. 2. Processos Pontuais. 3.

Análise Topológica de Dados. I. Favio Coletti, Cristian. II.

Miranda Machado, Daniel. III. Programa de Pós-Graduação em
Matemática, 2024. IV. Título.



Documento assinado digitalmente
CRISTIAN FAVIO COLETTI
Data: 13/03/2024 11:21:18-0300
Verifique em <https://validar.iti.gov.br>



Documento assinado digitalmente
RAFAEL POLLI CARNEIRO
Data: 12/03/2024 22:51:21-0300
Verifique em <https://validar.iti.gov.br>

Este exemplar foi revisado e alterado em relação à versão original, de acordo com as observações levantadas pela banca examinadora no dia da defesa, sob responsabilidade única do autor e com a anuência do orientador.



MINISTÉRIO DA EDUCAÇÃO

Fundação Universidade Federal do ABC

Avenida dos Estados, 5001 – Bairro Santa Terezinha – Santo André – SP
CEP 09210-580 · Fone: (11) 4996-0017

FOLHA DE ASSINATURAS

Assinaturas dos membros da Banca Examinadora que avaliou e aprovou a Defesa de Tese de Doutorado do candidato, RAFAEL POLLI CARNEIRO realizada em 15 de Dezembro de 2023:

Documento assinado digitalmente
 **ALEXSANDRO GIACOMO GRIMBERT GALLO**
Data: 31/01/2024 16:53:22-0300
Verifique em <https://validar.iti.gov.br>

Prof.(a) ALEXSANDRO GIACOMO GRIMBERT GALLO

Documento assinado digitalmente
 **ANATOLI IAMBARTSEV**
Data: 09/02/2024 10:48:02-0300
Verifique em <https://validar.iti.gov.br>

Prof.(a) ANATOLI YAMBARTSEV

Documento assinado digitalmente
 **JEFERSON CASSIANO**
Data: 11/02/2024 09:49:10-0300
Verifique em <https://validar.iti.gov.br>

Prof.(a) JEFERSON CASSIANO
UNIVERSIDADE FEDERAL DO ABC

Documento assinado digitalmente
 **NATALIA ANDREA VIANA BEDOYA**
Data: 04/02/2024 14:19:52-0300
Verifique em <https://validar.iti.gov.br>

Prof.(a) NATALIA ANDREA VIANA BEDOYA
UNIVERSIDADE FEDERAL DE SÃO CARLOS

Prof.(a) DANIEL MIRANDA MACHADO
UNIVERSIDADE FEDERAL DO ABC

* Por ausência do membro titular, foi substituído pelo membro suplente descrito acima: nome completo, instituição e assinatura



MINISTÉRIO DA EDUCAÇÃO

Fundação Universidade Federal do ABC

Avenida dos Estados, 5001 – Bairro Santa Terezinha – Santo André – SP
CEP 09210-580 · Fone: (11) 4996-0017

Prof.(a) ERIKA ALEJANDRA RADA MORA
UNIVERSIDADE FEDERAL DO ABC

Prof.(a) GLEICIANE DA SILVA ARAGÃO
UNIVERSIDADE FEDERAL DE SÃO PAULO

Prof.(a) LEONEL ZUAZNABAR MOLINER
UNIVERSIDADE FEDERAL DO ABC

Prof.(a) PABLO MARTIN RODRIGUEZ

Documento assinado digitalmente
 **CRISTIAN FAVIO COLETTI**
Data: 16/02/2024 10:19:50-0300
Verifique em <https://validar.iti.gov.br>

Prof.(a) CRISTIAN FAVIO COLETTI
UNIVERSIDADE FEDERAL DO ABC - Presidente

* Por ausência do membro titular, foi substituído pelo membro suplente descrito acima: nome completo, instituição e assinatura



Universidade Federal do ABC

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

AGRADECIMENTOS

Agradeço profundamente à minha esposa, Deyse Cadete de Campos, por ser uma presença constante em minha vida. Seu apoio e amor são pilares essenciais que me impulsionam a crescer e aprimorar-me diariamente. Através do seu apoio, você me inspira a buscar sempre o melhor em mim e a enxergar a beleza que permeia cada aspecto da vida. Sou verdadeiramente grato por tê-la ao meu lado, compartilhando cada momento desta jornada.

Agradeço de coração aos meus pais, Jacira Polli Carneiro e Celso Carneiro, e a toda a minha família, por serem meu suporte constante e por me transmitirem a valorização e a importância dos estudos. Sem o encorajamento e a orientação deles, não teria alcançado este momento significativo em minha jornada acadêmica. Suas palavras de incentivo e exemplos de dedicação foram fundamentais para moldar meu caráter e minha determinação. Sou imensamente grato por ter uma família tão incrível ao meu lado.

Agradeço aos meus orientadores, Cristian Favio Coletti e Daniel Miranda Machado, por compartilharem generosamente seus conhecimentos e por me incentivarem incansavelmente ao longo do meu doutorado. Mesmo diante das inúmeras dificuldades enfrentadas ao longo desses anos, sua presença, amizade e apoio foram fundamentais para minha jornada acadêmica. Agradeço, também, a todos os meus colegas da universidade, com destaque para o Jeferson Cassiano, a Sandra Maria Zapata Yepes e a Mirelle Alves de Freitas, cuja colaboração e amizade enriqueceram significativamente minha experiência acadêmica. Suas contribuições foram inestimáveis e jamais serão esquecidas.

Finalmente, agradeço à UFABC e seu corpo docente por possibilitarem meu aperfeiçoamento profissional e pessoal, e agradeço, tanto a CAPES como a UFABC, pelo auxílio financeiro a minha pesquisa.

RESUMO

Nos últimos anos o rápido avanço tecnológico levou à produção e ao acúmulo de uma enorme quantidade de dados. A Análise Topológica de Dados é uma área emergente que se utiliza da estrutura geométrica e topológica dos dados para extrair informações destes. Sua importância emerge de sua capacidade de revelar padrões intrincados e informações ocultas em dados multidimensionais. Ela ajuda a identificar clusters, hierarquias e conexões entre os conjuntos de dados, independentemente da métrica usada e é especialmente útil em situações em que os métodos tradicionais podem falhar devido à complexidade dos dados ou à presença de outliers. Sua aplicabilidade têm crescido e abarca diversas áreas, incluindo ciência, engenharia, medicina, finanças e neurociência.

Dentre as técnicas da Análise Topológica de Dados, podemos destacar a Homologia Persistente, cuja ideia central reside em utilizar-se da Homologia Simplicial para extrair informações dos dados em estudo. Esta técnica será o objeto central de nosso estudo.

Nesta tese provamos uma Lei Forte dos Grandes Números e um Teorema do Limite Central para a Homologia Persistente associada a uma classe de Processos Pontuais. A continuação utilizamos dessa ferramenta para estudarmos a propagação de Fake News no Twitter.

Palavras-chave: Homologia Persistente, Fake News, Processos Pontuais

ABSTRACT

In recent years, rapid technological advancement has led to the production and accumulation of large amounts of data. Topological Data Analysis is an emerging area that uses the geometric and topological structure of data to extract information from them. Its importance emerges from its ability to reveal intricate patterns and hidden information in multidimensional data. It helps identify clusters, hierarchies, and connections between data sets, regardless of the metric used, and is especially useful in situations where traditional methods may fail due to data complexity or the presence of outliers. Its applicability has grown and covers several areas, including science, engineering, medicine, finance and neuroscience.

Among the techniques of Topological Data Analysis, we can highlight Persistent Homology, whose central idea lies in using Simplicial Homology to extract information from data under study. This technique will be the central object of our study. In this thesis we prove a Strong Law of Large Numbers and a Central Limit Theorem for Persistent Homology associated with a class of Point Processes. Next, we use this tool to study the spread of Fake News on Twitter.

Keywords: Homologia Persistente, Fake News, Processos Pontuais

CONTEÚDO

1	Introdução a homologia singular e simplicial	3
1.1	Complexos Simpliciais	3
1.2	Complexos simpliciais abstratos	10
1.3	Equivalência Homotópica	16
1.4	Sequências Exatas	23
1.5	Complexo de Cadeias simpliciais	27
1.6	Homologia Simplicial	39
1.7	Induzindo Morfismos na Categoria dos Complexos de Cadeias . .	48
2	Homologia Persistente	51
2.1	Motivação	52
2.2	Formalização	57
2.3	Diagramas de Persistência	63
2.4	Filtrações	68
2.5	Algoritmo para calcular o Diagrama de Persistência em $\mathbb{Z}_2$	69
2.6	Utilizando o ambiente Python	81
3	Homologia Persistente de Processos Pontuais	85
3.1	Notação Básica	86
3.2	Filtrações Simpliciais	87
3.3	Número de Persistência de Betti	89
3.4	Contagem de simplexes e desigualdades	91
3.5	Classes de Processos Pontuais	93
3.5.1	Processos Pontuais	93
3.5.2	Processos Pontuais majorados por um PPP homogêneo . .	96
3.5.3	Processos pontuais de Gibbs	97
3.6	Lei Forte do número de persistência de Betti	101
3.7	Teorema do Limite Central	111
4	Detectando Notícias Falsas no Twitter	123
4.1	Coleta de Dados	127
4.2	Descrição dos grafos amostrados	129

4.3	Transformação dos dados brutos para os Grafos de interação entre usuários	132
4.4	Análise exploratória dos grafos de interação de usuários	133
4.5	A homologia persistente de Caminhos	141
4.5.1	Caminhos em um grafo	142
4.6	Classificação dos trend topics usando a homologia persistente . .	144
4.6.1	Resultados	144
4.6.2	Diagramas de Persistência	145
4.6.3	Silhuetas de Persistência e Entropia de Persistência	149
4.6.4	Extraíndo as informações das silhuetas e da entropia	151
4.6.5	Conclusão	159
4.7	Classificação dos trend topics usando o Botômetro	160
4.7.1	Resultados	161
4.7.2	Dados do Botômetro	161
4.7.3	Histogramas	163
4.7.4	Conclusão	164
4.8	Comparação	164
Apêndice		171
A	Resultado do Algoritmo 2, Capítulo 2, Aplicado ao Exemplo 2.5.1	173
Referência Bibliográfica		174

INTRODUÇÃO

Por volta dos anos 90, pesquisas envolvendo o desenvolvimento da Homologia Persistente tomaram início, sendo o pesquisador italiano Patrizio Frosini um dos seus precursores. Todavia, necessitou-se mais de uma década para que chegassemos a uma formalização apropriada desta estrutura, sendo o pesquisador Gunnar Carlsson um dos grandes responsáveis pelo amadurecimento desta área.

Devido ao recente surgimento da Homologia Persistente, poucos resultados existem quando tentamos induzir estruturas probabilísticas nesta categoria. Com isto em mente, iniciamos nosso trabalho propondo um texto base e introdutório sobre Homologia Persistente, com o objetivo de servir de auxílio a pessoas sem contato algum com a matéria mas que desejam estudar este subramo da Análise Topológica de Dados.

O primeiro capítulo desta tese trará uma breve (pra não dizer ingênua) abordagem sobre o que é homologia simplicial. Tendo sempre como objetivo destacar, ou pelo menos tentar destacar, a intuição geométrica por detrás da homologia simplicial, o qual se estenderá para a Homologia Persistente. Em seguida, o próximo capítulo apresenta, finalmente, a definição da Homologia Persistente, junto com sua motivação e alguns exemplos.

Após esta breve introdução à Homologia Persistente, apresentaremos, no terceiro capítulo, nossa primeira contribuição, que consiste na demonstração de uma Lei Forte dos Grandes Números e um Teorema do Limite Central para a Homologia Persistente associada a uma classe de Processos Pontuais. Desta forma, extendemos os resultados de convergências já conhecidos para processos pontuais que, não necessariamente, sejam estacionários.

Em seguida, no quarto capítulo, estudaremos a dissiminação de Fake News (ou, se preferir, mensagens maliciosas) na rede social X, antigo Twitter. Nossa metodologia de análise se dará pela coleta de mais de 90000 tweets contendo certas palavras chaves e, em seguida, avaliaremos, por meio da homologia persistente, o grafo resultante dos tweets e suas interações.

1 | INTRODUÇÃO A HOMOLOGIA SINGULAR E SIMPLICIAL

Este Capítulo tem como intuito dar uma breve intuição sobre a Teoria de Homologia Simplicial, tendo em mente pessoas sem contato com a matéria, mas que tem interesse em estudar a área da Análise Topológica de Dados e Homologia Persistente.

O fundamental aqui é desenvolver uma intuição das estruturas e operadores desenvolvidas na Teoria da Homologia Simplicial. Para uma abordagem completa sobre o assunto, recomenda-se o livro de Hatcher [21] (em inglês) ou o livro do Elon [18].

1.1 COMPLEXOS SIMPLICIAIS

Iniciaremos esta seção desenvolvendo intuição sobre o seguinte objeto matemático: os *simplexos*. Em seguida, apresentaremos sua forma mais abstrata, os *simplexos abstratos*. Todavia, deixamos o adendo que ambas estruturas, sob certas circunstâncias, desempenharão o mesmo objetivo para o nosso estudo.

Seja $\mathbb{R}^n$ o espaço euclidiano e $a_0, a_1, \dots, a_n$ uma sequência finita de pontos em $\mathbb{R}^n$, dizemos que estes pontos são *pontos independentes* se eles satisfazem a seguinte condição:

$$a_1 - a_0, a_2 - a_0, \dots, a_n - a_0 \text{ são L.I.,}$$

com L. I. significando vetores linearmente independentes.

Exemplo 1.1.1. Seja a_0, a_1, a_2 uma sequência de pontos do $\mathbb{R}^2$ dada por

$$a_0 = (1, 1), \quad a_1 = (1, 0), \quad a_2 = (0, 1).$$

Logo, como é fácil observar, teremos $a_1 - a_0 = (0, -1)$ e $a_2 - a_1 = (-1, 0)$, ou seja, os pontos a_0, a_1, a_2 são pontos independentes.

Observação 1.1.1. Seja $a_0, a_1, \dots, a_n$ uma sequência de pontos independentes em $\mathbb{R}^n$. Note que, para qualquer reordenação desses pontos, sempre obteremos, ainda, uma sequência de pontos independentes. Em outras palavras, para toda permutação

$$s : \{0, 1, 2, \dots, n\} \rightarrow \{0, 1, 2, \dots, n\},$$

temos que a sequência $b_0, b_1, \dots, b_n$, dada por

$$b_0 = a_{s_0}, \quad b_1 = a_{s_1}, \quad \dots, \quad b_n = a_{s_n},$$

ainda é uma sequência de pontos independentes.

Nosso próximo passo será estudar o conjunto resultante da combinação afim desses elementos:

Definição 1.1.1 (simplexos). Seja $a_0, a_1, \dots, a_n$ uma sequência de pontos independentes do $\mathbb{R}^n$. Então, chamaremos de *simplexo n -dimensional*, ou simplesmente *n -simplexo*, como o conjunto denotado por $\langle a_0, a_1, \dots, a_n \rangle$ e definido pela igualdade abaixo

$$\langle a_0, a_1, \dots, a_n \rangle := \left\{ \sum_{i=0}^n \lambda_i a_i \in \mathbb{R}^n; \lambda_i \geq 0 \text{ para todo } i \in \{0, 1, \dots, n\} \text{ e } \sum_{i=0}^n \lambda_i = 1 \right\}.$$

Agora, seja $t = \langle a_0, a_1, \dots, a_n \rangle$ um n -simplexo e $p \in \langle a_0, a_1, \dots, a_n \rangle$ o ponto

$$p = \sum_{i=0}^n \lambda_i a_i.$$

Usualmente, chamamos de *coordenadas baricêntricas* do ponto p os valores que compõem suas coordenadas: $\lambda_0, \lambda_1, \dots, \lambda_n$. Chamamos, também, de *face* de um n -simplexo todo subconjunto $s \subseteq \langle a_0, a_1, \dots, a_n \rangle$ satisfazendo

$$s = \langle a_{i_0}, a_{i_1}, \dots, a_{i_m} \rangle,$$

com $i_0, i_1, \dots, i_m$ uma subsequência de índices de $\{0, 1, 2, \dots, n\}$.

Observação 1.1.2. Note que as coordenadas baricêntricas de um n -simplexo são únicas. Ou seja, dados

$$p = \sum_{i=1}^n \lambda_i a_i \quad \text{e} \quad q = \sum_{i=1}^n \delta_i a_i$$

elementos de $\langle a_0, a_1, \dots, a_n \rangle$, então

$$p = q \implies \forall i \in \{0, 1, 2, \dots, n\} \text{ vale } \lambda_i = \delta_i.$$

A demonstração desta propriedade pode ser vista em Elon [18, pg 82 - Teorema 1].

Abaixo, estabelecemos a definição de simplexo canônico e fixamos uma notação para ele, o qual usaremos repetidamente no decorrer deste Capítulo.

Definição 1.1.2 (simplexo canônico). Dada a base canônica $e_0, e_1, e_2, \dots, e_n$ do espaço $\mathbb{R}^{n+1}$, onde

$$e_i = (\delta_{i,j})_{j=0}^n$$

para todo $i = 0, 1, \dots, n$ (e, $\delta_{i,j}$ o delta de Kronecker). Chamaremos de *simplexo canônico de dimensão n* , denotado por Δ^n , o simplexo abaixo

$$\Delta^n := \langle e_0, e_1, e_2, \dots, e_n \rangle = \left\{ (t_0, t_1, \dots, t_n) \in \mathbb{R}^{n+1}; t_i \geq 0 \forall i \in \{0, 1, \dots, n\} \text{ e } \sum_{i=0}^n t_i = 1 \right\}.$$

Após estabelecido o objeto central desta seção, os simplexos, é interessante adquirir um pouco de intuição sobre estes conjuntos. A primeira propriedade importante a salientar, reside na convexidade desses conjuntos. Em outras palavras, simplexos n -dimensionais $\langle a_0, a_1, \dots, a_n \rangle$ são nada mais que a envoltória convexa gerada pelos pontos independentes $a_0, a_1, \dots, a_n$. Ou seja, dado o conjunto $X = \{a_0, a_1, \dots, a_n\}$, podemos pensar no simplexo $\langle a_0, a_1, \dots, a_n \rangle$ como a área hachurada na Figura 1, onde um elemento p desta região é dado por

$$p = \sum_{i=0}^n \lambda_i a_i, \quad \text{com } \sum_{i=0}^n \lambda_i = 1 \text{ e } \lambda_i \geq 0, \text{ para todo } i = 0, 1, \dots, n.$$

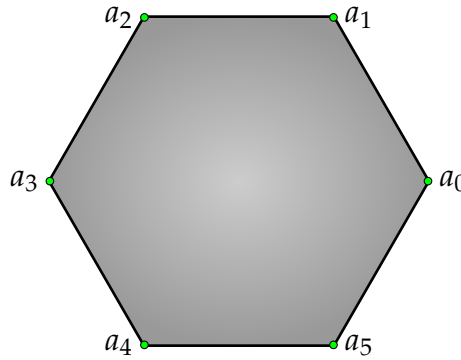
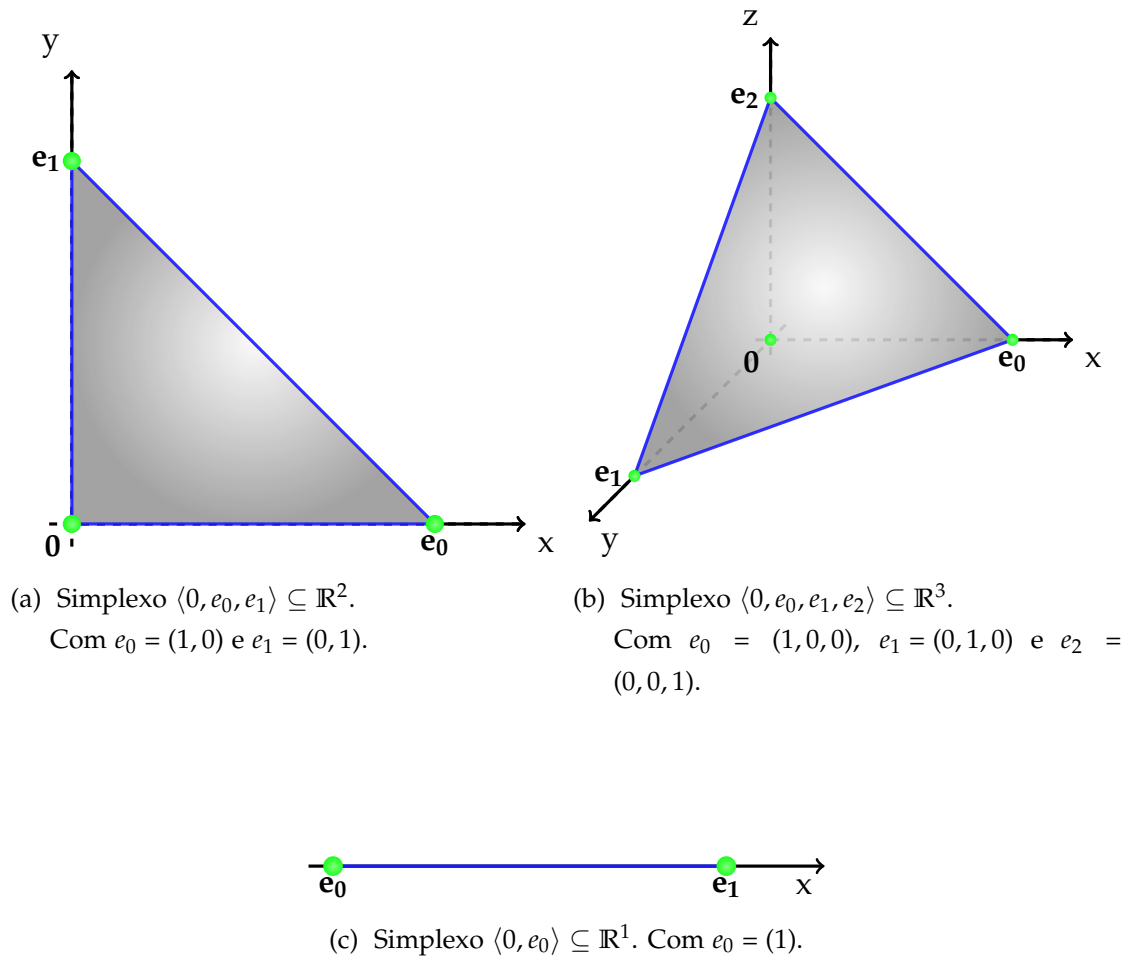


Figura 1: Envoltória convexa dos pontos $a_0, a_1, \dots, a_5$.

Outra propriedade importante dos simplexos reside no apelo geométrico, e em sua simplicidade algébrica (combinação afim de elementos), o que nos possibilita uma maior flexibilidade (principalmente intuitiva) sobre os estruturas abstratas que veremos

Figura 2: Simplexos de dimensão $n = 1, 2, 3$.

a seguir. Note que para dimensões baixas $n = 1, 2, 3$, os simplexos apresentam uma geometria familiar para nós, eles podem ser vistos como objetos da geometria euclidiana usual, como o segmento de reta, o triângulo e a pirâmide triangular (vide a Figura 2).

O próximo passo para esta Seção é definir a estrutura dos *complexos simpliciais*, ou, como também são chamados, os *poliedros*. Junto com esta estrutura estaremos interessados no processo de *triangularização*, que consiste basicamente em definirmos homeomorfismos entre espaços topológicos e complexos simpliciais. Começamos com a seguinte definição:

Definição 1.1.3 (*Complexos Simpliciais* ou *poliedros*). Seja (K, S) uma tupla, o qual K é um subespaço não nulo do espaço euclidiano $\mathbb{R}^n$ e S uma família de simplexos dada por:

$$S = \{s \subseteq K; \text{ existe uma sequência de pontos independentes } (a_i)_{i=0}^n \in K \\ \text{tal que } s = \langle a_0, a_1, \dots, a_n \rangle, \text{ com } n \text{ um natural} \}$$

Dizemos que K é um *complexo simplicial*, ou um *poliedro*, se a família de simplexos S satisfazer as seguintes propriedades:

(i) Todo ponto $p \in K$ pertence a um número finito de simplexos em S , i.e.,

$$\text{card}\{s \in S; x \in s\} < \infty;$$

(ii) Para todo $s \in S$ temos que todas as faces de s devem pertencer a S também;

(iii) Para todos simplexos $s, t \in S$ devemos ter $s \cap t = \emptyset$ ou a intersecção deles deve ser igual a alguma face de s e t .

Nosso primeiro exemplo de complexo simplicial consiste no mais simples possível, resultante de pontos e segmentos de retas os conectando.

Exemplo 1.1.2. Seja $G = (V, A)$ um grafo não orientado com vértices em $\mathbb{R}^n$. O conjunto V pode ser visto como um complexo simplicial, formado como a união de simplexos de dimensão 1 (segmentos de reta) que se interseccionam nos vértices ou possuem intersecção vazia, como representado na Figura 3.

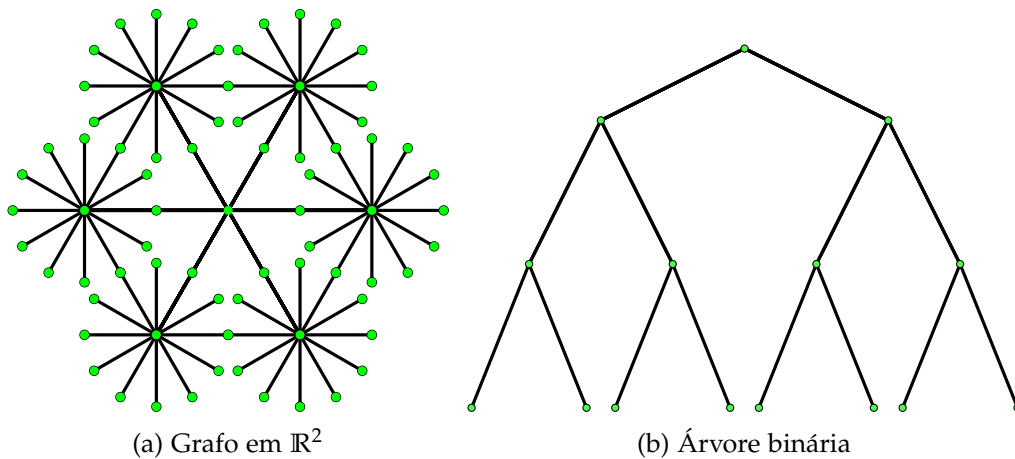


Figura 3: Grafos vistos como complexos simpliciais.

A seguir apresentamos um exemplo clássico de um espaço topológico homeomorfo a um complexo simplicial:

Exemplo 1.1.3. Considere o Toro da Figura 4, e denotado por $\mathcal{T} \subseteq \mathbb{R}^3$. Sabemos que o Toro admite a parametrização

$$\begin{aligned} \phi : [0, 2\pi] \times [0, 2\pi] / \sim &\rightarrow \mathcal{T} \subseteq \mathbb{R}^3 \\ (\theta, \alpha) &\mapsto ((R + r \cos(\alpha)) \cos(\theta), (R + r \cos(\alpha)) \sin(\theta), r \sin(\alpha)). \end{aligned}$$

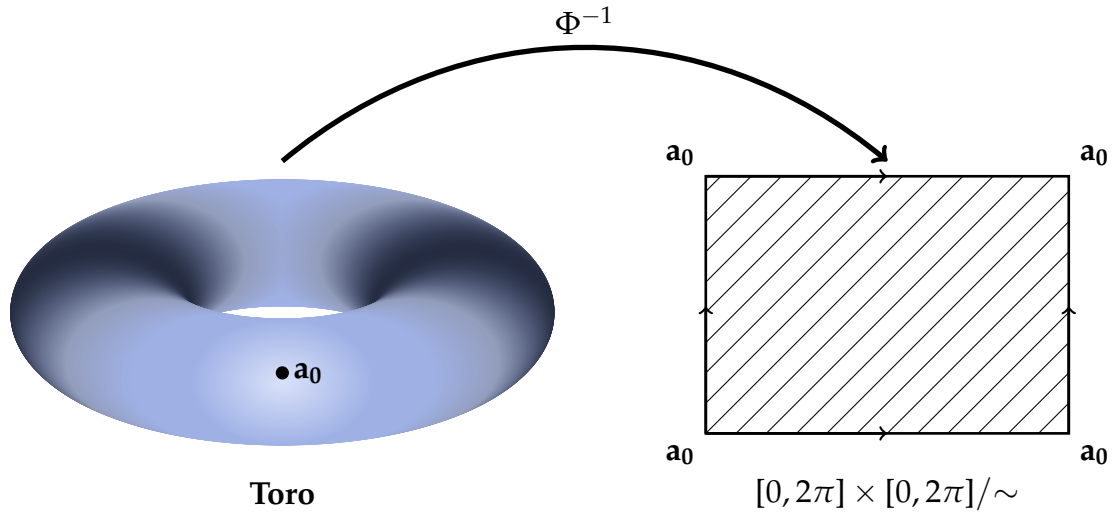


Figura 4: Homeomorfismo entre o Toro e o conjunto $[0, 2\pi] \times [0, 2\pi] / \sim$.

onde r, R são constantes (R define o raio de rotação em relação ao eixo z , por qual uma circunferência de raio r se moverá) e $\sim$ a relação de equivalência definida por

$$(x_1, y_1) \sim (x_2, y_2) \iff \begin{cases} y_1 = 0 \text{ e } y_2 = 2\pi, \text{ ou} \\ x_1 = 0 \text{ e } x_2 = 2\pi, \text{ ou} \\ x_1 = x_2 \text{ e } y_1 = y_2. \end{cases}$$

Mostraremos que $[0, 2\pi] \times [0, 2\pi] / \sim$ é homeomorfo a imagem de ϕ , o Toro. Note que:

- (i) ϕ é contínua. Consequência do fato de suas funções coordenadas serem contínuas;
- (ii) ϕ é uma bijeção com sua imagem (o Toro);
- (iii) $[0, 2\pi] \times [0, 2\pi] / \sim$ é um conjunto compacto (pois $[0, 2\pi] \times [0, 2\pi]$ é o conjunto resultante do produto de compactos, o qual, munido da topologia produto será um conjunto compacto - vide, Teorema de Tychonoff, capítulo 5, pg 143 do livro de topologia do Kelley [22])

(iv) ϕ é uma aplicação fechada. De fato, para todo subconjunto fechado

$$B \subseteq [0, 2\pi] \times [0, 2\pi] / \sim$$

temos que B será um conjunto compacto cuja imagem, também, é compacta. Portanto, $\phi(B)$ é um subconjunto fechado de $\mathbb{R}^3$ (este fato decorre da caracterização de espaços vetoriais de dimensão finita V , onde todo subconjunto compacto de V é fechado em V). Com isto, mostramos que

para todo conjunto fechado B vale que $\phi(B)$ é um conjunto fechado.

i.e., ϕ é uma aplicação fechada.

Com isto somos capazes de provar que ϕ é um homeomorfismo. Em decorrência de ϕ ser uma bijeção com sua imagem e, denotando por $g := \phi^{-1}$, temos que para todo fechado $B \subseteq [0, 2\pi] \times [0, 2\pi] / \sim$

$$g^{-1}(B) = \{y \in \mathbb{R}^3; g(y) \in B\} = \{y \in \mathbb{R}^3; \phi^{-1}(y) \in B\} = \phi(B) = \overline{\phi(B)},$$

ou seja, a pré-imagem de todo fechado pela função $g = \phi^{-1}$ é um conjunto fechado, i.e., ϕ^{-1} é contínua. Sendo assim, ϕ é um homeomorfismo, i.e.,

$$\text{Toro} \cong [0, 2\pi] \times [0, 2\pi] / \sim,$$

como ilustrado pela Figura 4.

Agora, observe que $[0, 2\pi] \times [0, 2\pi] / \sim$ é um complexo simplicial, descrito na Figura 5, e formado por simplexes de dimensão 0, 1, 2

0 – simplexes : $a_0, a_1, a_2, \dots, a_8$ (total: 9 vértices)

1 – simplexes : $\overrightarrow{a_0 a_1}, \overrightarrow{a_1 a_2}, \overrightarrow{a_2 a_0}, \overrightarrow{a_0 a_4}, \dots$ (total: 27 simplexes de dimensão 1)

2 – simplexes : $s_0, s_1, s_2, \dots, s_{17}$ (total: 18 simplexes de dimensão 2)

Além do mais, fixado um ponto a_0 qualquer do Toro (vide Figura 4) e fixada sua imagem em $[0, 2\pi] \times [0, 2\pi] / \sim$, o qual denotaremos por a_0 também, podemos definir um esquema de colagem de tal forma a obtermos nosso conjunto inicial novamente.

Motivados pelo Exemplo 1.1.3, definimos o conceito de triangularização de um espaço topológico.

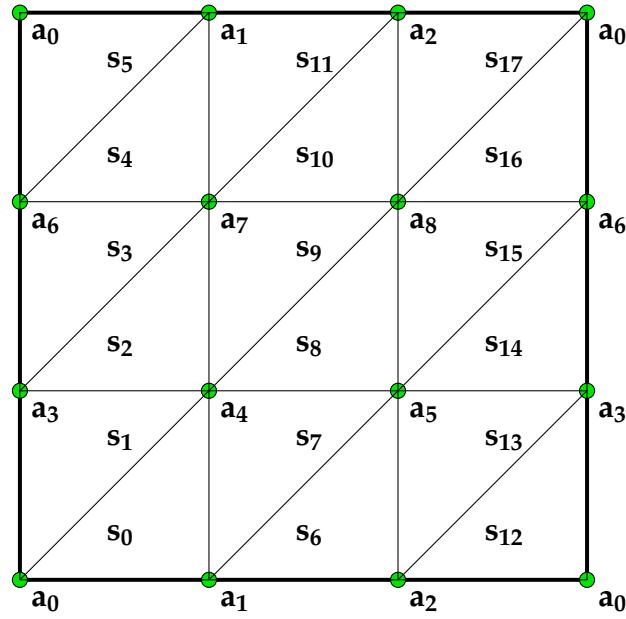


Figura 5: $[0, 2\pi] \times [0, 2\pi] / \sim$ como complexo simplicial.

Definição 1.1.4 (triangularização). Seja X um espaço topológico. Dizemos que X é *triangularizável* se existir um complexo simplicial (K, S) tal que K e X sejam homeomorfos.

Observação 1.1.3. Como já observado, temos que o Toro pode ser triangularizado pois é homeomorfo ao espaço $[0, 2\pi] \times [0, 2\pi] / \sim$, que admite ser escrito como união de simplexos.

1.2 COMPLEXOS SIMPLICIAIS ABSTRATOS

Uma importante propriedade dos complexos simpliciais (K, S) é seu apelo combinatório, devido ao fato de K ser a união de conjuntos, os simplexos, definidos de forma única como função de seus vértices. Sob esta perspectiva, estendemos nossa estrutura de complexos simpliciais para *complexos simpliciais abstratos*.

Definição 1.2.1 (Complexos Simpliciais abstratos). Seja $\Phi = (\mathcal{K}, \mathcal{S})$ uma tupla, o qual $\mathcal{K}$ é um conjunto não nulo e cujos elementos são chamados de vértices, e $\mathcal{S}$ uma família de subconjuntos finitos de $\mathcal{K}$:

$$\mathcal{S} \subseteq \mathcal{P}(\mathcal{K})$$

onde $s \in \mathcal{S}$ será chamado de *simplexo abstrato*, com $\mathcal{P}(\mathcal{K})$ denotando o conjunto das partes de $\mathcal{K}$. Dizemos que Φ é um *complexo simplicial abstrato* se a família de simplexos $\mathcal{S}$ satisfazer as seguintes propriedades:

(P1) Todo vértice $p \in \mathcal{K}$ pertence a um número finito de simplexos em $\mathcal{S}$, i.e.,

$$\text{card}\{s \in \mathcal{S}; p \in s\} < \infty;$$

(P2) Para todo $s \in \mathcal{S}$ e, para todo subconjunto $t \subseteq s$, temos

$$t \in \mathcal{S}.$$

Para nos familiarizarmos com a noção de complexos simpliciais, abaixo apresentaremos alguns exemplos, começando com um caso trivial:

Exemplo 1.2.1. Seja a tupla $\Phi = (\mathcal{K}, \mathcal{S})$ formada pelos conjuntos

$$\mathcal{K} = \{x, y, z\}, \quad \text{e} \quad \mathcal{S} = \{\emptyset, \{x\}, \{y\}, \{z\}, \{x, z\}, \{y, z\}\}.$$

É fácil observar que as propriedades (P1) e (P2) são satisfeitas e, portanto, Φ é um complexo simplicial abstrato.

Exemplo 1.2.2. Seja $l \in \mathbb{N}$ um número natural fixado e os conjuntos $\mathcal{K} = \mathbb{N}$ e

$$\mathcal{S} = \{s \in \mathcal{P}(\mathbb{N}); \text{card } s < \infty \text{ e } |\max s - \min s| \leq l\}.$$

Então $\Phi = (\mathcal{K}, \mathcal{S})$ é um complexo simplicial abstrato. Para provarmos esta afirmação, mostremos que as propriedades (P2) e (P1) são satisfeitas. Dado um elemento w qualquer de $\mathcal{S}$, então, pela definição de $\mathcal{S}$, temos que w tem cardinalidade finita. Além do mais, para todo $t \subseteq w$, vale

$$\max t \leq \max w, \min w \leq \min t \implies \max t - \min t \leq l.$$

Finalmente, para mostrarmos que a propriedade (P1) é satisfeita, basta notarmos que, para todo ponto $p \in \mathcal{K}$, a desigualdade

$$\text{card}\{w \in \mathcal{S}; p \in w\} \leq \sum_{i=0}^l \binom{2l+1}{i} = 2^{2l+1}$$

é verdadeira, o que implica que Φ cumpre as condições de (P1). Ou seja, Φ é um complexo simplicial abstrato.

Exemplo 1.2.3. $\Phi = (\mathcal{K}, \mathcal{S})$ a tupla definida por

$$\mathcal{K} = \{(x_i)_{i=1}^{\infty} \in \mathbb{N}^{\infty}; (x_i)_{i=1}^{\infty} \text{ é uma sequência quase nula}\},$$

onde, por uma sequência ser quase nula, queremos dizer que a sequência $(x_i)_{i=1}^{\infty}$ admite uma quantidade finita de termos não nulos, e $\mathcal{S} = \mathcal{P}(\mathcal{S})$. Neste exemplo, Φ não é um complexo simplicial abstrato, pois falha a propriedade: *vértices devem pertencer a um número finito de simplexes*. De fato, denotemos por

$$a_i := (\delta_{i,j})_{j=1}^{\infty} \quad (\delta_{i,j} \text{ o símbolo de kroenecker})$$

os pontos com unidade no índice i e zero nas restantes coordenadas, e fixemos o ponto $a := a_1$. Consequentemente, o vértice a pertence a todos os conjuntos

$$s_i = \{a, a_i\}, \quad i = 2, 3, 4, \dots$$

Ou seja,

$$\text{card}\{s \in \mathcal{S}; a \in s\} = \infty,$$

contrariando a propriedade (P1).

Estabeleceremos, abaixo, o importante conceito de morfismos entre complexos simpliciais abstratos. Dizemos que uma função

$$f : (\mathcal{K}_1, \mathcal{S}_1) \rightarrow (\mathcal{K}_2, \mathcal{S}_2)$$

é um *morfismo entre complexos simpliciais* ou, uma *transformação simplicial*, se $f(\mathcal{K}_1) \subseteq \mathcal{K}_2$ e, para todo simplexo $s \in \mathcal{S}_1$, tivermos $f(s) \in \mathcal{S}_2$. Da mesma forma, dizemos que $(\mathcal{K}_1, \mathcal{S}_1)$ e $(\mathcal{K}_2, \mathcal{S}_2)$ são isomorfos, se existirem transformações simpliciais

$$f : (\mathcal{K}_1, \mathcal{S}_1) \rightarrow (\mathcal{K}_2, \mathcal{S}_2) \quad \text{e} \quad g : (\mathcal{K}_2, \mathcal{S}_2) \rightarrow (\mathcal{K}_1, \mathcal{S}_1)$$

tal que

$$f \circ g = \mathbb{1}_{\mathcal{K}_2} \quad \text{e} \quad g \circ f = \mathbb{1}_{\mathcal{K}_1}.$$

Note que a função identidade $\mathbb{1} : (\mathcal{K}, \mathcal{S}) \rightarrow (\mathcal{K}, \mathcal{S})$ é uma transformação simplicial e, além disso, dadas transformações simpliciais f, g , temos que sua composta $f \circ g$ ainda é um morfismo entre complexos simpliciais abstratos. Finalmente, pelo fato de a composição entre funções ser associativa, temos garantido que os complexos simpliciais abstratos podem ser tratados com o viés da teoria das categorias ¹,

¹ Para uma breve revisão dos conceitos de teoria das categorias recomenda-se a leitura dos capítulos 4 e 15 do livro “Algebra”, dos autores MacLane e Birkhoff [20]. Estes dois capítulos cobrem tudo, e mais um pouco, sobre a teoria de categorias que necessitaremos.

A categoria dos complexos simpliciais abstratos, denotada por **SCpx**, é definida como a categoria

SCpx :

Obj(**SCpx**) $:= \{(\mathcal{K}, \mathcal{S}); (\mathcal{K}, \mathcal{S}) \text{ é um complexo simpliciai abstrato}\}$

Hom $((\mathcal{K}_1, \mathcal{S}_1), (\mathcal{K}_2, \mathcal{S}_2)) := \{f : (\mathcal{K}_1, \mathcal{S}_1) \rightarrow (\mathcal{K}_2, \mathcal{S}_2); f \text{ é uma transformação simplicial}\}$

Composição entre setas : $\circ(f, g) := f \circ g$ (composição usual entre funções),

onde **Obj** denota a classe dos objetos da categoria e **Hom** $((\mathcal{K}_1, \mathcal{S}_1), (\mathcal{K}_2, \mathcal{S}_2))$ representa o conjunto formado pelos morfismos entre os elementos $(\mathcal{K}_1, \mathcal{S}_1), (\mathcal{K}_2, \mathcal{S}_2) \in \mathbf{Obj}(\mathbf{SCpx})$.

Para encerrar esta Seção, mostraremos como os complexos simpliciais da primeira Seção estão conectados com os complexos simpliciais abstratos, e, como a ideia de triangulação de espaços topológicos por simplexes, pode ser estendida para simplexes abstratos. Iniciaremos esta discussão com o seguinte exemplo:

Exemplo 1.2.4. Considere as letras a, b, c representando símbolos distintos e o complexo simplicial $\Phi = (\mathcal{K}, \mathcal{S})$, munido pelo conjunto de vértices $\mathcal{K} = \{a, b, c\}$ e pela família de simplexes abstratos

$$\mathcal{S} = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}\}.$$

Afirmamos que este complexo simplicial abstrato induz um conjunto, denotado por $\mathcal{K}_g$, homeomorfo a S^1 .

De fato, consideremos o espaço vetorial livre V gerado por $\mathcal{K}$

$$V = \{\lambda_1 a + \lambda_2 b + \lambda_3 c; \lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}\} \cong \mathbb{R}^3.$$

(Note que a soma usada para descrever os elementos de V é a soma formal). Para facilitar a notação, denotaremos por

$$\tilde{a} := 1 \cdot a, \quad \tilde{b} := 1 \cdot b, \quad \tilde{c} := 1 \cdot c$$

os vetores que geram V . Assim, usando a notação de simplexes da primeira seção, e denotando para toda sequência $a_0, a_1, \dots, a_n \in V$ o conjunto

$$\langle a_0, a_1, \dots, a_n \rangle := \left\{ \sum_{i=0}^n \lambda_i a_i \in V; \forall i = 1, 2, \dots, n, \text{ vale } \lambda_i \geq 0 \text{ e } \sum_{i=0}^n \lambda_i = 1 \right\},$$

definimos, com o propósito de carregar os simplexes abstratos para V , o conjunto

$$S_g = \{\tilde{a}, \tilde{b}, \tilde{c}, \langle \tilde{a}, \tilde{b} \rangle, \langle \tilde{a}, \tilde{c} \rangle, \langle \tilde{b}, \tilde{c} \rangle\}.$$

Com isto, formamos a tupla (K_g, S_g) , onde K_g é dado por

$$K_g = \bigcup_{s \in S_g} s.$$

Decorre, do isomorfismo trivial de V com $\mathbb{R}^3$, dado pela transformação linear que associa as bases

$$\tilde{a} \mapsto (1, 0, 0), \quad \tilde{b} \mapsto (0, 1, 0) \quad \tilde{c} \mapsto (0, 0, 1),$$

que (K_g, S_g) é isomorfo ao esqueleto de dimensão 1 do simplexo canônico $\Delta^2 \subseteq \mathbb{R}^3$, como pode ser visto na Figura 6. Além do mais, este esqueleto é homeomorfo ao esqueleto de dimensão 1 do simplexo $\langle 0, e_0, e_1 \rangle \subseteq \mathbb{R}^2$ ($e_0 = (1, 0), e_1 = (0, 1)$). Finalmente, como o esqueleto de dimensão 1 do simplexo $\langle 0, e_0, e_1 \rangle$ é homeomorfo a S^1 , obtemos o desejado. Em outras palavras, temos que

$$K_g \cong 1\text{-esqueleto}(\Delta^2) \cong 1\text{-esqueleto}(\langle 0, (0, 1), (1, 0) \rangle) \cong S^1$$

onde o símbolo de congruência usado acima indica homeomorfismo.

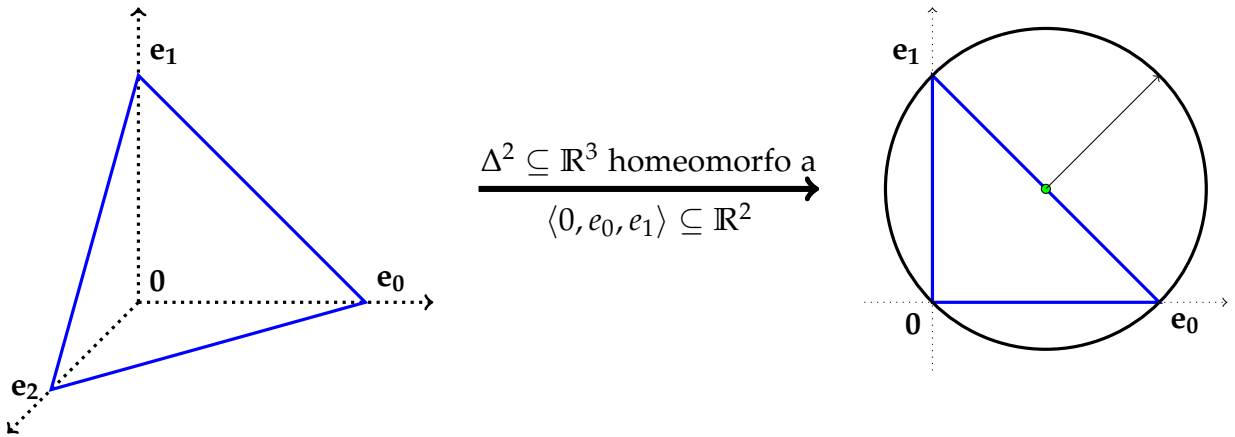


Figura 6: Simplexo canônico $\Delta^2 \subseteq \mathbb{R}^3$ e o homeomorfismo de seu esqueleto de dimensão 1 (triângulo azul na figura a esquerda) com o esqueleto de dimensão 1 (triângulo da figura a direita) do simplexo $\langle 0, e_0, e_1 \rangle$ de $\mathbb{R}^2$, que, por sua vez, é homeomorfo a S^1 . Os vetores e_0, e_1, e_2 representam os vetores da base canônica de $\mathbb{R}^2, \mathbb{R}^3$, dependendo do contexto.

O que acabamos de mostrar com o Exemplo 1.2.4 foi:

1. Complexos simpliciais podem ser vistos como complexos simpliciais abstratos;

2. Dizer que um espaço topológico é triangularizável por um poliedro equivale a dizer que é triangularizável por algum complexo simplicial abstrato.

Iremos formalizar estas ideias agora. Seja $\Phi = (\mathcal{K}, \mathcal{S})$ um complexo simplicial abstrato e V_Φ o $\mathbb{R}$ -espaço vetorial livre gerado por $\mathcal{K}$, i.e., o espaço dado pelas somas formais finitas de $\mathcal{K}$:

$$V_\Phi = \left\{ \sum_{p \in \mathcal{K}} \lambda_p \cdot p; (\lambda_p)_{p \in \mathcal{K}} \in \mathbb{R} \text{ é uma sequência quase nula} \right\}.$$

No espaço V_Φ , para facilitar a notação, definimos sua base como os vetores

$$\tilde{p} = 1 \cdot p \in V_\Phi, \forall p \in \mathcal{K}.$$

Agora, induzimos simplexos em V_Φ , por meio da família de simplexos abstratos $\mathcal{S}$, da forma mais direta possível. Para todo $s \in \mathcal{S}$ definimos (lembre-se que s é um conjunto finito)

$$\tilde{s} = \left\{ \sum_{p \in s} \lambda_p \tilde{p}; \sum_{p \in s} \lambda_p = 1 \text{ e } \lambda_p \geq 0, \forall p \in s \right\}$$

e

$$S_g = \{\tilde{s} \in V_\Phi; s \in \mathcal{S}\}.$$

Com isto, somos capazes de estabelecer a *realização geométrica* do complexo simplicial abstrato Φ , que será o conjunto

$$K_g := \bigcup_{s \in \mathcal{S}} \tilde{s}.$$

Dizemos, ainda, que um espaço topológico é *triangularizável* se existir um complexo simplicial abstrato tal que sua realização geométrica é homeomorfa ao espaço topológico.

Observação 1.2.1. Todo complexo simplicial (ou, poliedro) (K, S) é realização geométrica de algum complexo simplicial abstrato. Para isto, basta tomar a tupla $\Phi = (\mathcal{K}, \mathcal{S})$ com $\mathcal{K}$ igual a todos os vértices do complexo simplicial e definir os simplexos abstratos de $\mathcal{S}$ como sendo o conjunto formado pelos vértices dos simplexos de S . Portanto, no texto não faremos mais distinção em um complexo simplicial ser abstrato ou não, e chamaremos complexos simpliciais e abstratos de apenas complexos simpliciais.

Em todos exemplos até agora, estivemos discutindo sobre a congruência de espaços topológicos como a existência de um homeomorfismo entre eles. A seguir, estenderemos esta noção para espaços topologicamente equivalentes, flexibilizando nossa análise.

1.3 EQUIVALÊNCIA HOMOTÓPICA

Esta seção é baseada no Capítulo Introdutório do livro de Topologia Algébrica do Hatcher [21, Capítulo 0] (excelente capítulo para estabelecer a ideia geométrica das estruturas que estudaremos a seguir) e, um pouco, no capítulo 1 do livro sobre Grupos Fundamentais do Elon [19].

Começaremos definindo a homotopia entre funções, o qual nos fornecerá a extensão de congruência, discutida no final da seção anterior.

Definição 1.3.1 (homotopia). Sejam X, Y espaços topológicos e $f, g : X \rightarrow Y$ funções contínuas. Dizemos que f e g são *homotópicas* se existir uma função contínua

$$h : X \times [0, 1] \rightarrow Y, \quad \text{satisfazendo} \quad h(x, 0) = f(x), \quad h(x, 1) = g(x), \quad \forall x \in X,$$

com $[0, 1]$ o intervalo unitário da reta real. A função h será chamada de homotopia entre f e g e, por muitas vezes, para simplificar a escrita, escreveremos $h_t(x)$ ao invés de $h(x, t)$

Observação 1.3.1. Sejam $f, g : X \rightarrow Y$ e $h : X \times [0, 1] \rightarrow Y$ funções contínuas. Para simplificar a notação, utilizaremos a convenção

$$h : f \cong g \iff h \text{ for uma homotopia entre as funções } f \text{ e } g.$$

Dadas funções contínuas definidas nos reais, $f, g : \mathbb{R} \rightarrow \mathbb{R}$, podemos entender por homotopia entre as funções f e g , como “variar continuamente” a imagem de f até a imagem da função g , como representado na Figura 7.

Exemplo 1.3.1. Sejam X um espaço topológico compacto e V um espaço vetorial normado. Então, fixado o ponto $p \in V$, temos que toda função contínua $f : X \rightarrow V$ é homotópica a função constante

$$g : X \rightarrow V, \quad x \mapsto p.$$

Este fato decorre da função

$$h : X \times [0, 1] \rightarrow V, \quad h(x, t) = (1 - t)f(x) + tp$$

ser contínua ². Fixado o ponto $(x_0, t_0) \in X \times [0, 1]$ temos, para todo $\varepsilon > 0$, que

² Vale a pena ressaltar que a topologia que estamos munindo o espaço $X \times [0, 1]$ é a topologia produto.

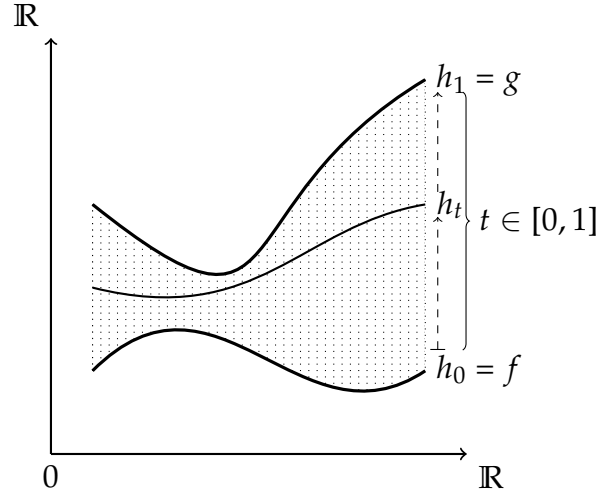


Figura 7: Homotopia entre duas funções $f, g : \mathbb{R} \rightarrow \mathbb{R}$.

1. Existe um conjunto aberto $U \subseteq X$ tal que

$$x \in U \implies \|f(x) - f(x_0)\| < \varepsilon/4;$$

2. Por X ser compacto e f e $\|\cdot\|$ serem funções contínuas, existe $M_0 \geq 0$ tal que

$$\|f(x)\| \leq M_0;$$

Desta forma, se considerarmos os valores $t \in [0, 1]$ como sendo os valores que satisfazem a desigualdade

$$|t - t_0| < K_0,$$

com $K_0 := \min \left\{ \frac{\varepsilon}{4\|p\|}, \frac{\varepsilon}{4M_0}, 1 \right\}$, teremos, para todo $(x, t) \in U \times (t_0 - K_0, t_0 + K_0)$, que

$$\begin{aligned} \|h(x, t) - h(x_0, t_0)\| &= \|(1 - t)f(x) + tp - (1 - t_0)f(x_0) - t_0p\| \\ &\leq \|f(x) - f(x_0)\| + \|p\||t - t_0| + t\|f(x) - f(x_0)\| + |t - t_0|\|f(x)\| \\ &< \varepsilon. \end{aligned}$$

Ou seja, a função h é uma homotopia.

Agora estamos de posse do instrumental necessário para definirmos a *equivalência homotópica* entre espaços topológicos.

Definição 1.3.2 (equivalência homotópica). Sejam X, Y dois espaços topológicos não vazios. Então, dizemos que X é *homotopicamente equivalente* ao espaço Y , ou, X tem o mesmo *tipo de homotopia* de Y , se existirem funções contínuas

$$f : X \rightarrow Y \quad \text{e} \quad g : Y \rightarrow X$$

tal que as compostas $f \circ g$ e $g \circ f$ são homotópicas as identidades $\mathbb{1}_Y$ e $\mathbb{1}_X$, respectivamente. Em outras palavras, para que X e Y tenham o mesmo tipo de homotopia, é necessário que existam as funções contínuas

$$f : X \rightarrow Y, \quad g : Y \rightarrow X, \quad h_1 : X \times [0, 1] \rightarrow Y \quad \text{e} \quad h_2 : Y \times [0, 1] \rightarrow X$$

tal que

$$h_1 : g \circ f \cong \mathbb{1}_X \quad \text{e} \quad h_2 : f \circ g \cong \mathbb{1}_Y.$$

Denotaremos por $X \cong Y$ quando os espaços, X e Y , forem do mesmo tipo de homotopia.

Observação 1.3.2. Espaços topológicos homeomorfos são homotopicamente equivalentes. De fato, considere X, Y como espaços homeomorfos, i.e., existem funções homeomorfas $f : X \rightarrow Y$ e $g : Y \rightarrow X$ tal que $f \circ g = \mathbb{1}_Y$ e $g \circ f = \mathbb{1}_X$. Decorre destas igualdades que as funções

$$h_1 : X \times [0, 1] \rightarrow X \quad \text{e} \quad h_2 : Y \times [0, 1] \rightarrow Y$$

dadas por $h_1(x, t) = x$ e $h_2(y, t) = y$, para todo $x \in X, y \in Y$ e $t \in [0, 1]$, são contínuas e

$$h_1 : g \circ f \cong \mathbb{1}_X \quad \text{e} \quad h_2 : f \circ g \cong \mathbb{1}_Y.$$

Observação 1.3.3. Espaços homotopicamente equivalentes não, necessariamente, são homeomorfos. Para isto, basta lembrar do Exemplo 1.3.1. Neste exemplo, provamos que para um espaço topológico compacto X e, um espaço vetorial V , toda função contínua $f : X \rightarrow V$ é homotópica a uma função constante $g : X \rightarrow V$. Sendo assim, consideremos $V = \{0\} \subseteq \mathbb{R}^2$ e $X = D^2 = \{x \in \mathbb{R}^2; \|x\| \leq 1\}$. Logo, para as funções

$$i : D^2 \rightarrow \{0\} \subseteq \mathbb{R}^2, \quad x \mapsto 0 \quad \text{e} \quad j : \{0\} \subseteq \mathbb{R}^2 \rightarrow D^2, \quad 0 \mapsto 0,$$

suas compostas $i \circ j : \{0\} \rightarrow \{0\} \subseteq \mathbb{R}^2$ e $j \circ i : D^2 \rightarrow \mathbb{R}^2$ serão funções constantes. Portanto, decorre de $\{0\}$ e D^2 serem espaços compactos, que as funções identidades (funções estas sempre contínuas)

$$\mathbb{1}_{\{0\}} : \{0\} \rightarrow \{0\} \quad \text{e} \quad \mathbb{1}_{D^2} : D^2 \rightarrow D^2$$

serão homotópicas as funções $i \circ j$ e $j \circ i$. De forma mais simbólica, existe homotopias h_1, h_2 tal que

$$h_1 : i \circ j \cong \mathbb{1}_{\{0\}} \quad \text{e} \quad h_2 : j \circ i \cong \mathbb{1}_{D^2}.$$

Ou seja, o disco D^2 é homotopicamente equivalente ao conjunto $\{0\}$. Porém o disco e ponto 0 não são homeomorfos.

Estas ideias condensam o que dissemos por estender o conceito de congruências para além de homeomorfismos entre espaços. Outra vantagem em se utilizar deste conceito, reside no fato de que existem casos particulares de homotopia entre espaços que carregam consigo um apelo geométrico muito grande, nos possibilitando abstrair e realizar os cálculos de forma mais simples do que tentar definir homeomorfismos, que quando possível, pode ser uma tarefa extremamente árdua.

Inicialmente, citaremos apenas algumas propriedades relevantes sobre espaços topologicamente equivalentes. As demonstrações destas propriedades são semelhantes ao feito (bem sucintamente) no início do livro “A concise Course in Algebraic Topology” do autor J.P. May [23, Seção 2 - Capítulo 1].

Sejam X, Y e Z espaços topológicos quaisquer. Então

- (i) Equivalência homotópica sempre será *reflexiva*, i.e., sempre teremos $X \cong X$;
- (ii) $X \cong Y \iff Y \cong X$, ou seja, a relação $\cong$ é *simétrica*;
- (iii) Sejam $X \cong Y$ e $Y \cong Z$. Então, $X \cong Z$ e, portanto, equivalência homotópica é, também, *transitiva*.

Desta forma, fixado um espaço topológico X , temos que no espaço

$$\mathcal{H} := \{A \subseteq X; A \text{ é um subespaço topológico de } X\}$$

a relação de equivalência homotópica $\cong$ é uma relação de equivalência.

Ao mostrarmos, na Observação 1.3.3, que o disco tem o mesmo tipo de homotopia que a origem do $\mathbb{R}^2$, desenvolvemos a intuição para o fato de ser possível “achatar” o disco em função do parâmetro $t \in [0, 1]$ (onde t pode ser interpretado como o parâmetro representando o tempo) até obtermos o ponto 0, preservando, ainda, propriedades geométricas (no caso, a homotopia). Esta ideia pode ser vista na Figura 8 e nos permite apresentar dois conceitos: *retrato por deformação* e *contração*.

Antes de definirmos o que é um retrato por deformação, estabeleceremos a definição de uma retração de um espaço X em um subespaço.

Definição 1.3.3. Seja X um espaço topológico e Y um subespaço de X . Y é um *retrato* de X se existir uma função $r : X \rightarrow X$ tal que

$$r(X) = Y \quad \text{e} \quad r|_Y = \mathbb{1}_Y.$$

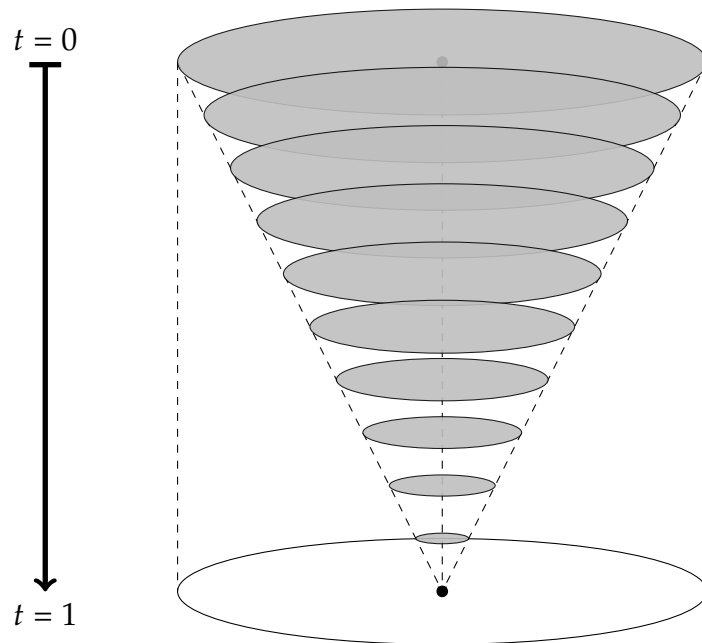


Figura 8: Comprimindo o disco D^2 , com o passar do tempo $t \in [0, 1]$, até obtermos o ponto $0 \in \mathbb{R}^2$.

Definição 1.3.4 (retrato por deformação). Dizemos que um subespaço topológico Y é um *retrato por deformação* do espaço topológico X , com $Y \subseteq X$, se existir uma função contínua, $h : X \times [0, 1] \rightarrow X$, tal que h é uma homotopia entre a função identidade $\mathbb{1}_X : X \rightarrow X$ com o retrato $r : X \rightarrow X$ de X em Y , i.e.,

$$h : r \cong \mathbb{1}_X.$$

O ponto $0 \in \mathbb{R}^2$ pode ser visto como um retrato do disco D^2 . De fato, como mostrado na Observação 1.3.3, temos garantido que, para as funções

$$i : D^2 \rightarrow \{0\}, x \mapsto 0 \quad \text{e} \quad j : \{0\} \rightarrow D^2, 0 \mapsto 0,$$

a composta $j \circ i : D^2 \rightarrow D^2$ será homotópica a identidade $\mathbb{1}_{D^2}$. Além do mais, a composta $j \circ i$ será uma retração do disco na origem:

$$j \circ i(D^2) = \{0\} \quad \text{e} \quad (j \circ i)|_{\{0\}} = \mathbb{1}_{\{0\}}.$$

A seguir, enunciamos uma Proposição que nos garante a seguinte propriedade: retratos por deformação de um espaço topológico X qualquer serão homotopicamente equivalentes ao espaço X . Isto nos fornece uma ferramenta muito útil, onde através de um espaço complicado X , obtemos um espaço “mais simples” que, ainda, preserva certas propriedades do conjunto inicial.

Proposição 1.3.1. *Seja X um espaço topológico e Y um subespaço de X . Então, se Y for um retrato por deformação de X , teremos que $X \cong Y$.*

Demonstração. Se Y é um retrato por deformação do espaço X , então a retração

$$r : X \rightarrow X, \quad \text{com} \quad r(X) = Y \text{ e } r|_Y = \mathbb{1}_Y$$

e a função identidade $\mathbb{1}_X$ são homotópicas, i.e., existe uma função contínua $h : X \times [0, 1] \rightarrow X$ tal que $h : \mathbb{1}_X \cong r$. Agora, considere a função inclusão $i : Y \hookrightarrow X$. Então, as seguintes igualdades são verdadeiras

$$r \circ i = \mathbb{1}_Y \quad \text{e} \quad i \circ r = r.$$

Consequentemente, por valer $i \circ r = r$, temos que h será uma homotopia entre $\mathbb{1}_X$ e $i \circ r$:

$$h : i \circ r \cong \mathbb{1}_X.$$

Note, também, que $\tilde{h} := h|_{Y \times [0, 1]}$ é uma homotopia entre

$$(h_0)|_Y = \mathbb{1}_Y \quad \text{e} \quad (h_1)|_Y = r|_Y = \mathbb{1}_Y = r \circ i,$$

onde

$$h_0 = h|_{X \times \{0\}} \quad \text{e} \quad h_1 = h|_{X \times \{1\}}.$$

Ou seja, temos que

$$\tilde{h} : r \circ i \cong \mathbb{1}_Y,$$

concluindo o desejado: $Y \cong X$. □

Definição 1.3.5 (Espaço contrátil). Seja X um espaço topológico e p um ponto qualquer de X . Então, dizemos que o espaço X é *contrátil*, se o mesmo tiver a mesma homotopia de $\{p\}$:

$$X \cong \{p\}.$$

Observação 1.3.4. O disco do $\mathbb{R}^2$ é contrátil. Como já mostrado anteriormente, a origem do $\mathbb{R}^2$ é um retrato por deformação disco D^2 . Logo, pela proposição acima

$$D^2 \cong \{0\}.$$

Abaixo forneceremos alguns exemplos de retratos por deformação.

Exemplo 1.3.2. Complementando o exemplo anterior, podemos olhar o disco como um retrato por deformação de um cone, como ilustrado pela Figura 9.

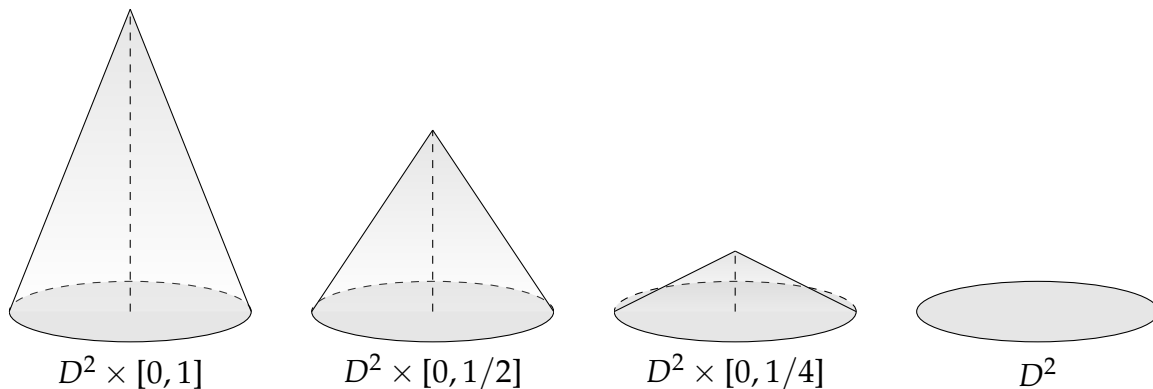


Figura 9: Retrato por deformação do cone para o disco.

Exemplo 1.3.3. Considere os grafos da Figura 3. Estes grafos podem ser representados por grafos de estrutura mais simples, o qual são nada mais, nada menos, que retratos por deformação dos grafos originais, e que estão representados na Figura 10.

Exemplo 1.3.4. Seja o conjunto $\mathcal{A}$ formado pelas letras do alfabeto:

$$\mathcal{A} = \{A, B, C, D, E, F, G, H, I, J, L, M, N, O, P, Q, R, S, T, U, V, X, Z\}$$

Considerando os elementos de $\mathcal{A}$ como subconjuntos de $\mathbb{R}^2$, munidos com a topologia induzida. Então, dadas as letras $\alpha, \beta \in \mathcal{A}$ podemos induzir a relação de equivalência $\sim \subseteq \mathcal{A} \times \mathcal{A}$, dada por

$$\alpha \sim \beta \iff \alpha, \beta \text{ forem homotopicamente equivalentes.}$$

Para esta relação de equivalência, temos as seguintes classes de equivalência:

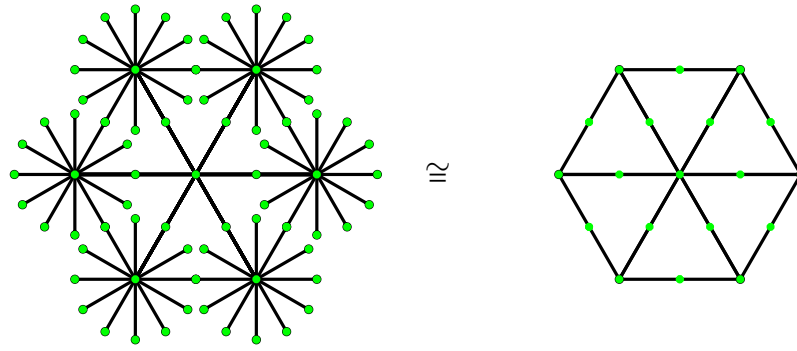
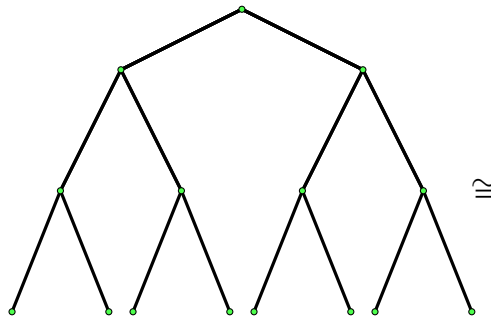
$$[A] = \{A, D, O, P, Q, R\}$$

$$[B] = \{B\}$$

$$[C] = \{C, E, F, G, H, I, J, L, M, N, S, T, U, V, X, Z\}$$

Avaliando o Exemplo 1.3.4 destacamos um padrão, letras com quantidades distintas de “buracos” não são homotopicamente equivalentes. Ou seja, avaliando a letra A, por exemplo, observamos um “buraco” em seu centro enquanto que a letra I, que não é equivalente a A, não possui nenhum tipo de buraco. Desta forma, percebemos, neste exemplo, que propriedades geométricas como buracos são preservadas por homotopias.

Esta ideia de estudarmos estruturas com propriedades geométricas semelhantes, como buracos, e quantificá-los, será muito utilizada no próximo Capítulo. A abordagem

(a) Grafo do $\mathbb{R}^2$ homotopicamente equivalente ao hexágono.

(b) Árvore binária homotopicamente equivalente ao ponto de origem da árvore.

Figura 10: Grafos da Figura 3 e suas versões com o mesmo tipo de homotopia.

utilizada para quantificarmos estas propriedades se pautará, fortemente, no estudo das propriedades algébricas dos **Complexos de Cadeias**. Antes de definirmos os complexos de cadeias, introduziremos o conceito de *sequências exatas* e forneceremos alguns resultados importantes.

1.4 SEQUÊNCIAS EXATAS

Seja $(A, +, \cdot, 1)$ um anel com unidade e $(M, +)$ um grupo abeliano, dizemos que M é um A -módulo à esquerda se existir uma função multiplicação por escalar

$$\begin{aligned} \odot : A \times M &\rightarrow M \\ (\alpha, v) &\mapsto \alpha \odot v \end{aligned}$$

que, para todo $\alpha, \beta \in A$ e para todo $u, v \in M$, satisfaz as condições

$$\begin{aligned}\alpha \odot (u + v) &= \alpha \odot u + \alpha \odot v; \\ (\alpha + \beta) \odot u &= \alpha \odot u + \beta \odot u; \\ (\alpha \cdot \beta) \odot u &= \alpha \odot (\beta \odot u); \\ 1 \odot u &= u.\end{aligned}$$

Note que, da mesma maneira, podemos construir A -módulos à direita.

Notação. A seguir, adotaremos a seguinte convenção. O operador multiplicação por escalar nos módulos e o operador multiplicação nos anéis serão representados da mesma maneira, i.e.,

$$(\alpha \cdot \beta) \odot (u) \stackrel{\text{por convenção}}{=} (\alpha\beta)u.$$

Dizemos que uma função entre A -módulos

$$f : M \rightarrow N$$

é um homomorfismo se a mesma preservar as propriedades algébricas entre os módulos. Ou seja, dados $u, v \in M$ e α um escalar, f deve satisfazer a igualdade

$$f(u + \alpha v) = f(u) + \alpha f(v).$$

Finalmente, o último conceito importante de se ter em mente, para o que faremos a seguir, é o de módulos livres. Dizemos que, M é um *módulo livre* à esquerda, se existir um subconjunto $X \subseteq M$ tal que, para toda função $f : X \rightarrow N$, com N um módulo à esquerda qualquer, existir um único homomorfismo de módulos $\tilde{f} : M \rightarrow N$ tal que o diagrama abaixo comute

$$\begin{array}{ccc} X & \xrightarrow{i} & M \\ & \searrow f & \downarrow \tilde{f} \\ & & N, \end{array}$$

com $i : X \hookrightarrow M$ a função inclusão. O conjunto X , sob estas condições, será chamado de *base* que gera o módulo M .

Módulos livres tem uma interessante interpretação por meio dos *elementos universais*, na teoria das categorias. Consideremos a categoria dos A -módulos à esquerda, **A-Mod**,

e a categoria dos conjuntos, **Set**, e fixemos um conjunto B . Portanto, o módulo livre M gerado pelo conjunto B , junto com a inclusão $i : B \hookrightarrow M$, será um *elemento universal* do *funtor covariante*

$$\mathcal{F} : \mathbf{A}\text{-Mod} \rightarrow \mathbf{Set},$$

onde, para todo módulo $N \in \mathbf{Obj}(\mathbf{A}\text{-Mod})$, definimos

$$\mathcal{F}(N) = \text{hom}(B, N) := \{f : B \rightarrow N; f \text{ é uma função}\}$$

e, para todo homomorfismo entre módulos N e P , dado por $\phi : N \rightarrow P$, o funtor vale

$$\mathcal{F}(\phi) : \text{hom}(B, N) \rightarrow \text{hom}(B, P)$$

$$f \mapsto \phi \circ f.$$

Temos que (M, i) é um elemento universal pois, decorre de que para toda tupla (N, j) , com N um módulo qualquer e j a função $j : B \rightarrow N$, existe, por M ser um módulo livre, um homomorfismo único $\tilde{j} : M \rightarrow N$ tal que $\tilde{j} \circ i = j$. Portanto, aplicando o funtor $\mathcal{F}$ ao morfismo $\tilde{j}$, obtemos

$$\mathcal{F}(\tilde{j})i = \tilde{j} \circ i = j.$$

Decorre de M ser um módulo livre gerado por B e, portanto, um elemento universal, que para qualquer módulo livre N gerado, também, por B , teremos, pela Propriedade de isomorfismo entre elementos universais (vide [20, Capítulo 4]), que M e N serão isomorfos. Ou seja, não importa como construímos nossos módulos livres gerados por B , eles sempre serão isomorfos. Para o leitor interessado em mais tópicos sobre teoria de Módulos, sugerimos o livro do professor Polcino, Anéis e Módulos, [24].

Agora, fixado um anel com unidade, denotado por A . Dizemos que uma sequência de A -módulos à esquerda, $(M_i)_{i \in \mathbb{Z}}$, é uma *sequência exata*, se existir uma sequência de homomorfismos de A -módulos $(f_i : M_i \rightarrow M_{i+1})_{i \in \mathbb{Z}}$ tal que, para todo inteiro i , valha

$$\ker(f_{i+1}) = \text{Im}(f_i).$$

Logo, para uma sequência exata $(M_i)_{i \in \mathbb{Z}}$, podemos pensar no segundo diagrama

$$\cdots \xrightarrow{f_{i-1}} M_{i-1} \xrightarrow{f_i} M_i \xrightarrow{f_{i+1}} M_{i+1} \xrightarrow{f_{i+2}} \cdots$$

Exemplo 1.4.1. Considere a sequência de $\mathbb{R}$ -módulos $(M_i)_{i \in \mathbb{Z}}$ dada por

- (i) $M_i = \mathbb{R}$, para $i = 0, 2;$

(ii) $M_i = \mathbb{R}^2$, para $i = 1$;

(iii) e, para $i \neq 0, 1, 2$, temos $M_i = (0)$;

Se associarmos a $(M_i)_{i \in \mathbb{Z}}$ a sequência de homomorfismos $(f_i)_{i \in \mathbb{Z}}$ definida por

$$f_i : M_i \rightarrow M_{i+1}, x \mapsto 0, \forall i \neq 0, 1;$$

$$f_0 : \mathbb{R} \rightarrow \mathbb{R}^2, x \mapsto (x, x);$$

$$f_1 : \mathbb{R}^2 \rightarrow \mathbb{R}, (x, y) \mapsto \langle (-1, 1), (x, y) \rangle$$

teremos que a sequência de módulos será exata. De fato, as únicas transformações que precisamos checar são f_0 e f_1 . Porém, f_0 mapeia a reta real na diagonal de $\mathbb{R}^2$, e f_1 calcula o produto interno dos pontos de $\mathbb{R}^2$ com o vetor $(-1, 1)$. Como este vetor é ortogonal a diagonal do espaço, temos que as condições para esta sequência ser exata são satisfeitas, como ilustrado no diagrama abaixo.

$$\begin{array}{ccccccc}
 (0) & \xrightarrow{f_{-1}} & \mathbb{R} & \xrightarrow{f_0} & \mathbb{R}^2 & \xrightarrow{f_1} & \mathbb{R} & \xrightarrow{f_2} & (0) \\
 \bullet & & \bullet & & \bullet & & \bullet & & \bullet \\
 \ker(f_{-1}) = (0) & & \ker(f_0) = (0) & & \ker(f_1) = \text{Diag}(\mathbb{R}^2) & & \ker(f_2) = \mathbb{R} & & \\
 \text{Im}(f_{-1}) = (0) & & \text{Im}(f_0) = \text{Diag}(\mathbb{R}^2) & & \text{Im}(f_1) = \mathbb{R} & & \text{Im}(f_2) = (0) & &
 \end{array}$$

De fato, sequências exatas como a do Exemplo acima desempenham um papel importante em topologia algébrica.

Definição 1.4.1 (sequência exata curta). Seja uma sequência exata $(M_i)_{i \in \mathbb{Z}}$. Então, chamaremos $(M_i)_{i \in \mathbb{Z}}$ de *sequência exata curta*, se a sequência satisfizer a condição: $M_i = (0)$, para todo $i \neq 0, 1, 2$. Denotaremos, por simplicidade, uma sequência exata curta da mesma maneira que a sequência do diagrama abaixo

$$(0) \xrightarrow{0} M \xrightarrow{f} N \xrightarrow{g} P \xrightarrow{0} (0)$$

onde $f : M \rightarrow N, g : N \rightarrow P$ são homomorfismos entre módulos e, o símbolo 0 acima das setas, indica o homomorfismo nulo.

Observação 1.4.1. Em uma sequência exata curta, como a descrita no diagrama acima, teremos, por estarmos trabalhando com um sequência exata, que

$$\ker(f) = \text{Im}(0) = (0), \quad \text{e} \quad \text{Im}(g) = \ker(0) = (0).$$

Ou seja, em uma sequência exata curta f será um monomorfismo enquanto que g será um epimorfismo.

Abaixo listamos um resultado importante, chamado de Lema dos Cinco, cuja demonstração pode ser vista no primeiro capítulo de Elon[18].

Teorema 1.1 (Lema dos Cinco). *Dado o diagrama comutativo abaixo*

$$\begin{array}{ccccccccc}
 M_1 & \xrightarrow{f_1} & M_2 & \xrightarrow{f_2} & M_3 & \xrightarrow{f_3} & M_4 & \xrightarrow{f_4} & M_5 \\
 \downarrow \phi_1 & & \downarrow \phi_2 & & \downarrow \phi_3 & & \downarrow \phi_4 & & \downarrow \phi_5 \\
 N_1 & \xrightarrow{f_1} & N_2 & \xrightarrow{f_2} & N_3 & \xrightarrow{f_3} & N_4 & \xrightarrow{f_4} & N_5.
 \end{array}$$

Então, se as linhas do diagrama forem formadas por sequências exatas, e $\phi_1, \phi_2, \phi_4, \phi_5$ forem isomorfismos, concluímos que ϕ_3 será um isomorfismo.

1.5 COMPLEXO DE CADEIAS SIMPLICIAIS

Agora vamos induzir uma estrutura algébrica nos complexos simpliciais abstratos, como dito no final da Seção 3. Como fixado anteriormente, chamaremos os complexos simpliciais abstratos de apenas complexos simpliciais, e seus simplexos abstratos de apenas simplexos.

Porém, antes de iniciarmos, precisamos estabelecer uma orientação nos vértices de um simplexo. Portanto, seja $\Phi = (\mathcal{K}, \mathcal{S})$ um complexo simplicial e $s \in \mathcal{S}$ um simplexo de dimensão n , i.e., $\text{card}(s) = n + 1$. Estabeleceremos uma orientação nos vértices de s da seguinte forma, escolhemos, inicialmente, uma ordenação qualquer para s , i.e., fixamos uma permutação $\sigma : I_n \rightarrow s$, com $I_n := \{0, 1, 2, \dots, n\}$, e fixamos, baseados em σ , o simplexo com orientação positiva

$$\mathbf{s} := [a_0, a_1, \dots, a_n], \text{ com } a_i = \sigma(i), \forall i \in I_n.$$

Fixada esta orientação, podemos aplicar permutações na ordem dos vértices do simplexo orientado $[a_0, a_1, \dots, a_n]$ o qual, entretanto, continuarão sendo ainda o mesmo simplexo.

Para isto, sabemos que o grupo simétrico S_n admite um morfismo com o grupo $\{-1, 1\}$, munido da multiplicação dos inteiros. Portanto, seja $\alpha \in S_{n+1}$ uma permutação,

e $[a_0, a_1, \dots, a_n]$ uma orientação para o simplexo s . Então, temos que toda permutação dos vértices do simplexo s , dada por $[a_{\alpha(0)}, a_{\alpha(1)}, \dots, a_{\alpha(n)}]$, valerá

$$[a_{\alpha(0)}, a_{\alpha(1)}, \dots, a_{\alpha(n)}] = \begin{cases} +[a_0, a_1, \dots, a_n] & \text{se a permutação } \alpha \text{ for par;} \\ -[a_0, a_1, \dots, a_n] & \text{se a permutação } \alpha \text{ for ímpar;} \end{cases}$$

Exemplo 1.5.1. Considere o simplexo $s = [a_0, a_1, a_2, a_3]$ e as reordenações nos índices de s dadas por

$$s_1 = [a_1, a_0, a_2, a_3] \quad \text{e} \quad s_2 = [a_1, a_2, a_0, a_3].$$

Note que como conjunto, os simplexos orientados s, s_1 e s_2 são os mesmos, ou seja, tratam do mesmo simplexo, mudando apenas a ordem dos vértices. Pela convenção acima, teremos que

$$s_1 = -[a_0, a_1, a_2, a_3],$$

pois aplicamos uma transposição nos índices de s e, a transposição é uma permutação ímpar. Semelhantemente, s_2 terá uma orientação positiva

$$s_2 = +[a_0, a_1, a_2, a_3],$$

pois resulta da composição de duas transposições em s .

Observação 1.5.1. Fixada uma orientação para um simplexo $s = [a_0, a_1, \dots, a_n]$ estabeleceremos que todo subconjunto de s terá a mesma orientação, induzida por s . Por exemplo, seja o simplexo $s = [a_0, a_1, a_2, a_3]$ e sua face t , resultante da retirada dos vértices a_1 e a_3 . Então, o simplexo t terá a seguinte orientação, induzida por s :

$$t = [a_0, a_2].$$

Agora podemos começar nosso tratamento algébrico em complexos simpliciais. Sejam $\Phi = (\mathcal{K}, \mathcal{S})$ um complexo simplicial, A um anel com unidade e, para todos simplexos em Φ , fixemos uma orientação. Estaremos interessados em estudar a sequência de A -módulos definida da seguinte maneira: para todo $i \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$, definimos o i -ésimo elemento desta sequência como sendo o módulo

$$C_i := \text{módulo livre gerado pelo conjunto } \{s \in \mathcal{S}; \dim(s) = i\},$$

onde, caso para algum índice i_0 tivermos $\{s \in \mathcal{S}; \dim(s) = i\} = \emptyset$, então o i_0 -ésimo elemento da sequência valerá $C_{i_0} = (0)$. Chamaremos esta sequência de: *sequência de mó-*

dulos livres gerados pelo complexo simplicial Φ . A seguir, definiremos homomorfismos $\partial_i : C_i \rightarrow C_{i-1}$ de tal forma a obtermos a sequência do diagrama abaixo

$$\cdots \longrightarrow C_{n+1} \xrightarrow{\partial_{n+1}} C_n \xrightarrow{\partial_n} \cdots \xrightarrow{\partial_3} C_2 \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0 = 0} (0)$$

Estes homomorfismos serão definidos da seguinte maneira, dada uma sequência de módulos livres gerados pelo complexo simplicial $\Phi = (\mathcal{K}, \mathcal{S})$, denotada por $(C_i)_{i \in \mathbb{N}_0}$. Iniciamos definindo para $\partial_0 : C_0 \rightarrow (0)$ o homomorfismo nulo e, em seguida, para um índice $i > 0$ qualquer, definiremos $\partial_i : C_i \rightarrow C_{i-1}$ em duas etapas. Primeiro, denotamos a base do A -módulo C_i como

$$S_i = \{s \in \mathcal{S}; \dim s = i\},$$

onde cada $s \in S_i$ é um simplexo orientado, e definimos a função auxiliar

$$\begin{aligned} \partial_i : S_i &\rightarrow C_{i-1} \\ [a_0, a_1, \dots, a_i] &\mapsto \sum_{j=0}^i (-1)^j [a_0, \dots, \widehat{a_j}, \dots, a_i]. \end{aligned}$$

Consequentemente pela propriedade universal dos módulos livres, sabemos que existe, e é único, um homomorfismo $\tilde{\partial}_i : C_i \rightarrow C_{i-1}$ tal que o diagrama abaixo comuta

$$\begin{array}{ccc} S_i & \xhookrightarrow{i} & C_i \\ & \searrow \partial_i & \downarrow \tilde{\partial}_i \\ & & C_{i-1} \end{array}$$

Sendo assim, definimos nossos operadores ∂_i como sendo os homomorfismos que estendem as funções ∂_i , i.e., definimos

$$\partial_i := \tilde{\partial}_i, \quad \forall i \in \mathbb{N}.$$

Chamaremos a sequência $(\partial_i)_{i \in \mathbb{N}_0}$ desses operadores de *sequência dos operadores bordos*, induzidos pela sequência de módulos livres gerados pelo complexo simplicial Φ . Estas duas sequências serão de grande importância para nós, formando a seguinte estrutura:

Definição 1.5.1 (*complexo de cadeias simpliciais*). Sejam $\Phi = (\mathcal{K}, \mathcal{S})$ um complexo simplicial e A um anel com unidade. Fixada uma orientação para todos simplexes em Φ , dizemos que

$$\mathcal{C} = \{(C_i)_{i \in \mathbb{N}_0}, (\partial_i : C_i \rightarrow C_{i-1})_{i \in \mathbb{N}_0}\}, \quad \text{com } \mathbb{N}_0 = \mathbb{N} \cup \{0\}$$

é um *complexo de cadeias simpliciais* se $(C_i)_{i \in \mathbb{N}_0}$ for uma sequência de módulos livres gerados pelo complexo simplicial Φ e ∂_i , para todo índice $i \in \mathbb{N}_0$, for um operador bordo induzido pela sequência $(C_i)_i$. Por simplicidade, usualmente denotaremos os complexos de cadeias simpliciais como $\mathcal{C} = (C_i, \partial_i)_i$, com $i \in \mathbb{N}_0$.

Agora, fixaremos um pouco de nomenclatura para estas estruturas. Seja $\mathcal{C} = (C_i, \partial_i)_i$ um complexo de cadeias simpliciais, induzido pelo complexo simplicial Φ . Então, chamamos de

- (i) *cadeia de dimensão i* , ou *i -cadeia*, todo elemento $x \in C_i$;
- (ii) *operador bordo*, os homomorfismos $\partial_i : C_i \rightarrow C_{i-1}$;
- (iii) *ciclo de dimensão i* , ou *i -ciclo*, todo elemento $x \in \ker(\partial_i)$;
- (iv) *bordo de dimensão i* , ou *i -bordo*, todo elemento $x \in \text{Im}(\partial_{i+1})$.

Abaixo listamos alguns exemplos de complexos de cadeias simpliciais, começando com um caso simples e, depois, calculando os complexos de cadeias baseados nos Exemplos das Seções anteriores.

Exemplo 1.5.2. Seja $\phi = (\mathcal{K}, \mathcal{S})$ o complexo simplicial cujos vértices são dados por

$$\mathcal{K} = \{a_0, a_1, a_2, a_3\}$$

e munido dos simplexes

$$\mathcal{S} = \{\emptyset, \{a_0\}, \{a_1\}, \{a_2\}, \{a_3\}, \{a_0, a_1\}, \{a_0, a_2\}, \{a_1, a_2\}, \{a_0, a_1, a_2\}\}.$$

Utilizando a orientação induzida pela ordenação dos vértices, iremos calcular, primeiramente, os módulos livres gerados pelos simplexes e, em seguida, os operadores bordos. Para simplificar, utilizaremos como anel o corpo dos reais. Ou seja, nossos módulos serão, de fato, espaços vetoriais. A sequência $(C_i)_{i \in \mathbb{N}_0}$ dos módulos livres será dada por:

$$C_0 = \left\{ \sum_{j=0}^3 \alpha_j [a_j]; \alpha_j \in \mathbb{R}, \forall j = 0, 1, 2, 3 \right\} \cong \mathbb{R}^4;$$

$$C_1 = \{ \alpha_0 [a_0, a_1] + \alpha_1 [a_0, a_2] + \alpha_2 [a_1, a_2]; \alpha_j \in \mathbb{R}, \forall j = 0, 1, 2 \} \cong \mathbb{R}^3;$$

$$C_2 = \{ \alpha [a_0, a_1, a_2]; \alpha \in \mathbb{R} \}.$$

Agora calculemos os operadores bordo. O primeiro operador, ∂_0 , será o operador nulo. Já o próximo operador, ∂_1 , terá os seguintes valores na base $\{[a_0, a_1], [a_0, a_2], [a_1, a_2]\}$:

$$\partial_1([a_0, a_1]) = [a_1] - [a_0]$$

$$\partial_1([a_0, a_2]) = [a_2] - [a_0]$$

$$\partial_1([a_1, a_2]) = [a_2] - [a_1],$$

resultando na matriz de transformação linear

$$(\partial_1) = \begin{matrix} & [a_0, a_1] & [a_0, a_2] & [a_1, a_2] \\ \begin{matrix} [a_0] \\ [a_1] \\ [a_2] \\ [a_3] \end{matrix} & \begin{pmatrix} -1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \end{matrix}$$

de posto 2. Finalmente, calculando o operador ∂_2 na cadeia $[a_0, a_1, a_2]$ obtemos

$$\partial_2([a_0, a_1, a_2]) = [a_1, a_2] - [a_0, a_2] + [a_0, a_1].$$

Deste valor, podemos escrever sua matriz de transformação linear

$$(\partial_2) = \begin{matrix} & [a_0, a_1, a_2] \\ \begin{matrix} [a_0, a_1] \\ [a_0, a_2] \\ [a_1, a_2] \end{matrix} & \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \end{matrix}.$$

Decorre destes valores, que o complexo de cadeia simplicial induzido por $\phi = (\mathcal{K}, \mathcal{S})$ pode ser representado pelo seguinte diagrama:

$$\begin{array}{ccccc} C_2 \cong \mathbb{R} & \xrightarrow{\partial_2} & C_1 \cong \mathbb{R}^3 & \xrightarrow{\partial_1} & C_0 \cong \mathbb{R}^4 & \xrightarrow{\partial_0} & (0) \\ & \text{ker}(\partial_2) \cong (0) & & \text{ker}(\partial_1) \cong \mathbb{R} & & \text{ker}(\partial_0) = \mathbb{R}^4 & \\ & \text{Im}(\partial_2) = \mathbb{R} & & \text{Im}(\partial_1) \cong \mathbb{R}^2 & & \text{Im}(\partial_0) \cong (0) & \end{array}$$

Exemplo 1.5.3. Vamos analisar os grafos da Figura 3, considerando nossos módulos como $\mathbb{R}$ -espaços vetoriais. Como veremos na seção seguinte, as estruturas de interesse nos complexos de cadeia serão os **grupos de homologia** e, estes, têm a propriedade de serem invariantes entre espaços homotopicamente equivalentes. Desta forma, ao invés

de realizarmos nossas contas nos grafos da Figura 3, podemos restringir nossa atenção aos grafos de mesmo tipo de homotopia, como representados na Figura 10.

De cara, já notamos que não há nada a fazer para a árvore binária pois, a mesma, é homotopicamente equivalente ao ponto. Portanto, estudemos o grafo homotopicamente equivalente da Figura 10a. Uma triangularização para ele é a exibida na Figura 11 abaixo. O complexo simplicial $\Phi = (\mathcal{K}, \mathcal{S})$ resultante da triangularização acima é dado

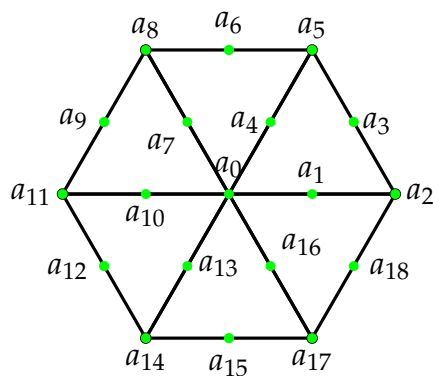


Figura 11: Triangularização do grafo da Figura 10a.

por vértices e arestas, i.e., a dimensão de Φ será 1. O complexo Φ possui os vértices

$$\mathcal{K} = \{a_0, a_1, a_2, a_3, a_4, \dots, a_{18}\}$$

e os seguinte simplexos

$$\begin{aligned} \mathcal{S} = \{ & \emptyset, \{a_0\}, \{a_1\}, \{a_2\}, \{a_3\}, \{a_4\}, \{a_5\}, \{a_6\}, \{a_7\}, \{a_8\}, \{a_9\}, \{a_{10}\}, \\ & \{a_{11}\}, \{a_{12}\}, \{a_{13}\}, \{a_{14}\}, \{a_{15}\}, \{a_{16}\}, \{a_{17}\}, \{a_{18}\}, (a_0, a_1), (a_1, a_2), \\ & \{a_0, a_4\}, \{a_4, a_5\}, \{a_0, a_7\}, \{a_7, a_8\}, \{a_0, a_{10}\}, \{a_{10}, a_{11}\}, \{a_0, a_{13}\}, \{a_{13}, a_{14}\}, \\ & \{a_0, a_{16}\}, \{a_{16}, a_{17}\}, \{a_2, a_3\}, \{a_3, a_5\}, \{a_5, a_6\}, \{a_6, a_8\}, \{a_8, a_9\}, \{a_9, a_{11}\}, \\ & \{a_{11}, a_{12}\}, \{a_{12}, a_{14}\}, \{a_{14}, a_{15}\}, \{a_{15}, a_{17}\}, \{a_{17}, a_{18}\}, \{a_{18}, a_2\} \}. \end{aligned}$$

Orientando os simplexos em função da própria ordenação dos vértices, vamos calcular os operadores bordo. Precisamos analisar apenas o operador ∂_1 , pois o resto são apenas

operadores nulos. Antes de calcularmos a matriz da transformação (∂_1), denotemos por $v_1, v_2, v_3, \dots, v_{24}$ os simplexos orientados

$$\begin{aligned} v_1 &= [a_0, a_1], & v_2 &= [a_1, a_2], & v_3 &= [a_0, a_4], & v_4 &= [a_4, a_5], & v_5 &= [a_0, a_7], \\ v_6 &= [a_7, a_8], & v_7 &= [a_0, a_{10}], & v_8 &= [a_{10}, a_{11}], & v_9 &= [a_0, a_{13}], & v_{10} &= [a_{13}, a_{14}], \\ v_{11} &= [a_0, a_{16}], & v_{12} &= [a_{16}, a_{17}], & v_{13} &= [a_2, a_3], & v_{14} &= [a_3, a_5], & v_{15} &= [a_5, a_6], \\ v_{16} &= [a_6, a_8], & v_{17} &= [a_8, a_9], & v_{18} &= [a_9, a_{11}], & v_{19} &= [a_{11}, a_{12}], & v_{20} &= [a_{12}, a_{14}], \\ v_{21} &= [a_{14}, a_{15}], & v_{22} &= [a_{15}, a_{17}], & v_{23} &= [a_{17}, a_{18}], & v_{24} &= [a_{18}, a_2]. \end{aligned}$$

Sendo assim, a transformação ∂_1 será dada pela seguinte matriz de transformação linear

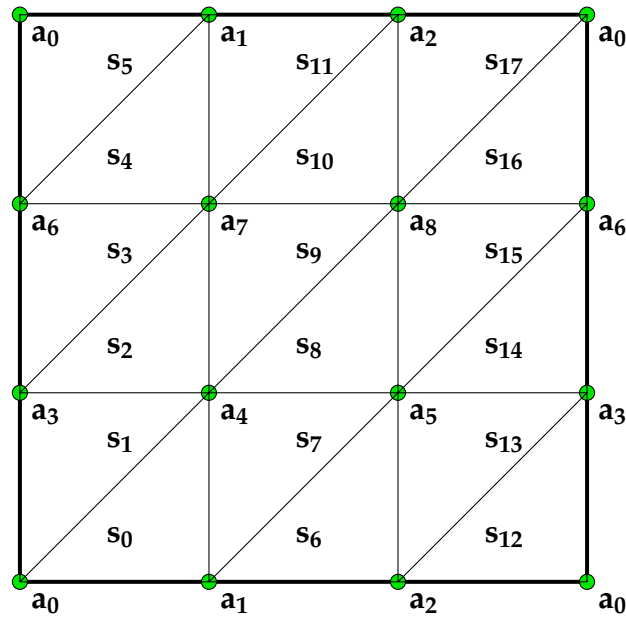
$$\begin{array}{c} \begin{matrix} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 & v_9 & v_{10} & v_{11} & v_{12} & v_{13} & v_{14} & v_{15} & v_{16} & v_{17} & v_{18} & v_{19} & v_{20} & v_{21} & v_{22} & v_{23} & v_{24} \end{matrix} \\ \begin{matrix} [a_0] \\ [a_1] \\ [a_2] \\ [a_3] \\ [a_4] \\ [a_5] \\ [a_6] \\ [a_7] \\ [a_8] \\ [a_9] \\ [a_{10}] \\ [a_{11}] \\ [a_{12}] \\ [a_{13}] \\ [a_{14}] \\ [a_{15}] \\ [a_{16}] \\ [a_{17}] \\ [a_{18}] \end{matrix} \begin{pmatrix} -1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \end{pmatrix} \end{matrix}$$

onde

$$\text{Im}(\partial_1) \cong \mathbb{R}^{18} \quad \text{e} \quad \ker(\partial_1) \cong \mathbb{R}^6.$$

Exemplo 1.5.4. Nosso último exemplo em complexos de cadeias simpliciais será o complexo induzido pela triangularização do toro, exibido abaixo. (vide Figuras 4 e 5.) Desta triangularização, temos o complexo de cadeias $\Phi = (\mathcal{K}, \mathcal{S})$, formado por

$$\mathcal{K} = \{a_0, a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8\}$$



e pelos simplexos

$$\begin{aligned} \mathcal{S} = & \{\emptyset, \{a_0\}, \{a_1\}, \{a_2\}, \{a_3\}, \{a_4\}, \{a_5\}, \{a_6\}, \{a_7\}, \{a_8\}, \\ & \{a_0, a_1\}, \{a_1, a_2\}, \{a_0, a_2\}, \{a_3, a_4\}, \{a_4, a_5\}, \{a_3, a_5\}, \{a_6, a_7\}, \{a_7, a_8\}, \{a_6, a_8\}, \\ & \{a_0, a_3\}, \{a_3, a_6\}, \{a_0, a_6\}, \{a_1, a_4\}, \{a_4, a_7\}, \{a_1, a_7\}, \{a_2, a_5\}, \{a_5, a_8\}, \{a_2, a_8\}, \\ & \{a_0, a_4\}, \{a_1, a_5\}, \{a_2, a_3\}, \{a_3, a_7\}, \{a_4, a_8\}, \{a_5, a_6\}, \{a_1, a_6\}, \{a_2, a_7\}, \{a_0, a_8\}, \\ & s_0, s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9, s_{10}, s_{11}, s_{12}, s_{13}, s_{14}, s_{15}, s_{16}, s_{17}\} \end{aligned}$$

orientados de acordo com a ordenação dos vértices. Calculemos apenas os operadores ∂_1 e ∂_2 , já que os demais são homomorfismos nulos. Porém, antes de iniciarmos as contas, estabeleceremos a seguinte notação para os simplexos ordenados de dimensão 1:

$$\begin{aligned} v_1 &= [a_0, a_1], & v_2 &= [a_1, a_2], & v_3 &= [a_0, a_2], & v_4 &= [a_3, a_4], & v_5 &= [a_4, a_5], \\ v_6 &= [a_3, a_5], & v_7 &= [a_6, a_7], & v_8 &= [a_7, a_8], & v_9 &= [a_6, a_8], & v_{10} &= [a_0, a_3], \\ v_{11} &= [a_3, a_6], & v_{12} &= [a_0, a_6], & v_{13} &= [a_1, a_4], & v_{14} &= [a_4, a_7], & v_{15} &= [a_1, a_7], \\ v_{16} &= [a_2, a_5], & v_{17} &= [a_5, a_8], & v_{18} &= [a_2, a_8], & v_{19} &= [a_0, a_4], & v_{20} &= [a_1, a_5], \\ v_{21} &= [a_2, a_3], & v_{22} &= [a_3, a_7], & v_{23} &= [a_4, a_8], & v_{24} &= [a_5, a_6], & v_{25} &= [a_1, a_6], \\ v_{26} &= [a_2, a_7], & v_{27} &= [a_0, a_8]. \end{aligned}$$

Portanto, a matriz de transformação linear do operador bordo ∂_1 será

$$\begin{matrix} & v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 & v_9 & v_{10} & v_{11} & v_{12} & v_{13} & v_{14} & v_{15} & v_{16} & v_{17} & v_{18} & v_{19} & v_{20} & v_{21} & v_{22} & v_{23} & v_{24} & v_{25} & v_{26} & v_{27} \\ \begin{matrix} [a_0] \\ [a_1] \\ [a_2] \\ [a_3] \\ [a_4] \\ [a_5] \\ [a_6] \\ [a_7] \\ [a_8] \end{matrix} & \begin{pmatrix} -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \end{pmatrix}$$

valendo

$$\text{Im}(\partial_1) \cong \mathbb{R}^8 \quad \text{e} \quad \ker(\partial_1) \cong \mathbb{R}^{19}.$$

Já, para o operador bordo ∂_2 temos a seguinte matriz de transformação linear

$$\begin{matrix} & s_0 & s_1 & s_2 & s_3 & s_4 & s_5 & s_6 & s_7 & s_8 & s_9 & s_{10} & s_{11} & s_{12} & s_{13} & s_{14} & s_{15} & s_{16} & s_{17} \\ \begin{matrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \\ v_9 \\ v_{10} \\ v_{11} \\ v_{12} \\ v_{13} \\ v_{14} \\ v_{15} \\ v_{16} \\ v_{17} \\ v_{18} \\ v_{19} \\ v_{20} \\ v_{21} \\ v_{22} \\ v_{23} \\ v_{24} \\ v_{25} \\ v_{26} \\ v_{27} \end{matrix} & \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 \end{pmatrix} \end{pmatrix}$$

com

$$\text{Im}(\partial_2) \cong \mathbb{R}^{17} \quad \text{e} \quad \ker(\partial_2) \cong \mathbb{R}^1.$$

Para complexos de cadeias simpliciais, vale o resultado sobre os operadores bordo:

Proposição 1.5.1. *Seja $\Phi = (\mathcal{K}, \mathcal{S})$ um complexo simplicial e $C = (C_i, \partial_i)_i$ o complexo de cadeias simplicial induzido por Φ . Então, para todo índice $i \in \mathbb{N}_0$, vale*

$$\partial_i \circ \partial_{i+1} = 0.$$

Demonstração. Vide [21], página 105, Lema 2.1. □

Agora vamos estabelecer algumas estruturas para os complexos de cadeias simpliciais, a fim de que a classe formada por eles tenha estrutura de uma categoria. Antes, definimos os **complexos de cadeia**, uma generalização dos complexos de cadeia simpliciais.

Definição 1.5.2 (complexos de cadeia). Dizemos que a sequência $C = (C_i, \partial_i)_{i \in \mathbb{N}_0}$, onde cada C_i é um A -módulo (A um anel com unidade) e cada $\partial : C_{i+1} \rightarrow C_i$ um homomorfismo de módulos, é um complexo de cadeias sempre quando

$$\partial_i \circ \partial_{i+1} = 0, \forall i \in \mathbb{N}_0$$

for satisfeita.

Observação 1.5.2. Denotaremos sempre os operadores bordo de complexos de cadeias como ∂ . Sendo assim, mesmo estando trabalhando com complexos de cadeias distintos, denotaremos seus operadores bordos pelo mesmo símbolo, sempre que não houver possibilidade de confusão.

Agora, consideremos os complexos de cadeia $C = (C_i, \partial_i)_i$, $D = (D_i, \partial_i)_i$ e $E = (E_i, \partial_i)_i$. A primeira coisa que devemos definir são os morfismos entre os complexos de cadeias. Isto é feito da seguinte maneira: dizemos que uma sequência de morfismos

$$f = (f_i : C_i \rightarrow D_i)_{i \in \mathbb{N}_0}$$

é um *morfismo* entre o complexo de cadeia C e o complexo D se, para todo índice $i \geq 0$, tivermos

$$f_i \circ \partial_{i+1} = \partial_{i+1} \circ f_{i+1}.$$

Em outras palavras, $f = (f_i)_i$ é um morfismo entre os complexos C e D , se o diagrama abaixo comutar.

$$\begin{array}{ccccccccccccccc} \cdots & \longrightarrow & C_{n+1} & \xrightarrow{\partial_{n+1}} & C_n & \xrightarrow{\partial_n} & \cdots & \xrightarrow{\partial_3} & C_2 & \xrightarrow{\partial_2} & C_1 & \xrightarrow{\partial_1} & C_0 & \xrightarrow{\partial_0 = 0} & (0) \\ \cdots & & \downarrow f_{n+1} & & \downarrow f_n & & \cdots & & \downarrow f_2 & & \downarrow f_1 & & \downarrow f_0 & & \\ \cdots & \longrightarrow & D_{n+1} & \xrightarrow{\partial_{n+1}} & D_n & \xrightarrow{\partial_n} & \cdots & \xrightarrow{\partial_3} & D_2 & \xrightarrow{\partial_2} & D_1 & \xrightarrow{\partial_1} & D_0 & \xrightarrow{\partial_0 = 0} & (0) \end{array}$$

Além do mais, dado o morfismo $g = (g_i : D_i \rightarrow E_i)_{i \in \mathbb{N}_0}$, podemos considerar a composta

$$f \circ g := (g_i \circ f_i)_{i \in \mathbb{N}_0}$$

que o diagrama abaixo, ainda, continua comutativo.

$$\begin{array}{ccccccccccccccc}
 \cdots & \longrightarrow & C_{n+1} & \xrightarrow{\partial_{n+1}} & C_n & \xrightarrow{\partial_n} & \cdots & \xrightarrow{\partial_3} & C_2 & \xrightarrow{\partial_2} & C_1 & \xrightarrow{\partial_1} & C_0 & \xrightarrow{\partial_0 = 0} & (0) \\
 & & \downarrow f_{n+1} & & \downarrow f_n & & \cdots & & \downarrow f_2 & & \downarrow f_1 & & \downarrow f_0 & & \\
 \cdots & \longrightarrow & D_{n+1} & \xrightarrow{\partial_{n+1}} & D_n & \xrightarrow{\partial_n} & \cdots & \xrightarrow{\partial_3} & D_2 & \xrightarrow{\partial_2} & D_1 & \xrightarrow{\partial_1} & D_0 & \xrightarrow{\partial_0 = 0} & (0) \\
 & & \downarrow g_{n+1} & & \downarrow g_n & & \cdots & & \downarrow g_2 & & \downarrow g_1 & & \downarrow g_0 & & \\
 \cdots & \longrightarrow & E_{n+1} & \xrightarrow{\partial_{n+1}} & E_n & \xrightarrow{\partial_n} & \cdots & \xrightarrow{\partial_3} & E_2 & \xrightarrow{\partial_2} & E_1 & \xrightarrow{\partial_1} & E_0 & \xrightarrow{\partial_0 = 0} & (0)
 \end{array}$$

Note que a identidade entre complexos de cadeias é a trivial:

$$\mathbb{1}_C = (\mathbb{1}_{C_i} : C_i \rightarrow C_i)_{i \in \mathbb{N}_0}.$$

Com isto, temos a *categoria dos complexos de cadeias*, denotada por $\mathbf{Cx}$, e descrita como

$\mathbf{Cx}$:

Obj ($\mathbf{Cx}$) $:= \{C = (C_i, \partial_i)_i; C \text{ é um complexo de cadeia}\}$

Hom $((C_i, \partial_i), (D_i, \partial_i)_i) := \{f = (f_i : C_i \rightarrow D_i)_i; f \text{ um morfismo entre os complexos}\}$

Composta entre setas $:= (f_i : C_i \rightarrow D_i)_i \circ (g_i : D_i \rightarrow E_i)_i = (g_i \circ f_i)_i.$

A seguir, mostraremos como construir complexos de cadeias a partir de subconjuntos dos módulos compondo esta cadeia, e exibiremos seus morfismos. Estas construções, de certa forma, se relacionam com os homomorfismos de inclusão e de projeção canônica.

Seja $C = (C_i, \partial_i)_{i \in \mathbb{N}_0}$ um complexo de cadeias e $(D_i)_{i \in \mathbb{N}_0}$ uma sequência de R -módulos satisfazendo

$$D_i \subseteq C_i \quad \text{e} \quad \partial_{i+1}(D_{i+1}) \subseteq D_i,$$

$\forall i \in \mathbb{N}_0$. Logo, a estrutura $D = (D_i, \partial_i|_{D_i})_{i \in \mathbb{N}_0}$ será um complexo de cadeias e

$$j = (j_i : D_i \hookrightarrow C_i)_{i \in \mathbb{N}_0},$$

será um homomorfismo entre cadeias, onde cada j_i representa a inclusão do R -módulos $D_i \subseteq C_i$. Segue abaixo o diagrama comutativo representando o morfismo entre os

complexos de cadeia. Nestas condições, chamamos o complexo $D = (D_i, \partial_i|_{D_i})_{i \in \mathbb{N}_0}$ de *subcomplexo* do complexo $C = (C_i, \partial_i)_{i \in \mathbb{N}_0}$.

$$\begin{array}{ccccccccccccccc} \cdots & \longrightarrow & D_{n+1} & \xrightarrow{\partial_{n+1}} & D_n & \xrightarrow{\partial_n} & \cdots & \xrightarrow{\partial_3} & D_2 & \xrightarrow{\partial_2} & D_1 & \xrightarrow{\partial_1} & D_0 & \xrightarrow{\partial_0} & (0) \\ & & \downarrow j_{n+1} & & \downarrow j_n & & \cdots & & \downarrow j_2 & & \downarrow j_1 & & \downarrow j_0 & & \\ \cdots & \longrightarrow & C_{n+1} & \xrightarrow{\partial_{n+1}} & C_n & \xrightarrow{\partial_n} & \cdots & \xrightarrow{\partial_3} & C_2 & \xrightarrow{\partial_2} & C_1 & \xrightarrow{\partial_1} & C_0 & \xrightarrow{\partial_0} & (0) \end{array}$$

Agora, seja $C = (C_i, \partial_i)_{i \in \mathbb{N}_0}$ um complexo de cadeias e $D = (D_i, \partial_i)_{i \in \mathbb{N}_0}$ um subcomplexo de C . Então, a estrutura $C/D = (C_i/D_i, \bar{\partial}_i)_{i \in \mathbb{N}_0}$, onde C_i/D_i é o quociente de R -módulos e

$$\bar{\partial}_i(x + D_i) = \partial_i(x), \quad \forall i \in \mathbb{N}_0 \quad \text{e} \quad x \in C_i,$$

será um complexo de cadeias, tal que a sequência das projeções canônicas

$$\pi = (\pi_i : C_i \rightarrow C_i/D_i)_{i \in \mathbb{N}_0}$$

será um morfismo entre os complexos C e C/D , dado pelo diagrama comutativo abaixo.

$$\begin{array}{ccccccccccccccc} \cdots & \longrightarrow & C_{n+1} & \xrightarrow{\partial_{n+1}} & C_n & \xrightarrow{\partial_n} & \cdots & \xrightarrow{\partial_3} & C_2 & \xrightarrow{\partial_2} & C_1 & \xrightarrow{\partial_1} & C_0 & \xrightarrow{\partial_0} & (0) \\ & & \downarrow \pi_{n+1} & & \downarrow \pi_n & & \cdots & & \downarrow \pi_2 & & \downarrow \pi_1 & & \downarrow \pi_0 & & \\ \cdots & \longrightarrow & C_{n+1}/D_{n+1} & \xrightarrow{\bar{\partial}_{n+1}} & C_n/D_n & \xrightarrow{\bar{\partial}_n} & \cdots & \xrightarrow{\bar{\partial}_3} & C_2/D_2 & \xrightarrow{\bar{\partial}_2} & C_1/D_1 & \xrightarrow{\bar{\partial}_1} & C_0/D_0 & \xrightarrow{\bar{\partial}_0} & (0) \end{array}$$

Finalmente, estabelecemos o conceito de sequência exata entre complexos de cadeias. Sejam $C = (C_i, \partial_i)_{i \in \mathbb{N}_0}$, $D = (D_i, \partial_i)_{i \in \mathbb{N}_0}$ e $E = (E_i, \partial_i)_{i \in \mathbb{N}_0}$ complexos de cadeias e

$$f = (f_i : C_i \rightarrow D_i)_{i \in \mathbb{N}_0} \quad \text{e} \quad g = (g_i : D_i \rightarrow E_i)_{i \in \mathbb{N}_0}$$

morfismos entre estas cadeias. Então, dizemos que a sequência de complexos

$$(0) \xrightarrow{0} C \xrightarrow{f} D \xrightarrow{g} E \xrightarrow{0} (0)$$

é uma sequência exata se, para cada $i \in \mathbb{N}_0$, valer

- (i) f_i forem monomorfismos;

(ii) g_i forem epimorfismos;

e o diagrama abaixo comutar

$$\begin{array}{ccccccccc}
 \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
 \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\
 (0) & \xrightarrow{0} & C_{n+1} & \xrightarrow{f_{n+1}} & D_{n+1} & \xrightarrow{g_{n+1}} & E_{n+1} & \xrightarrow{0} & (0) \\
 \partial_{n+1} \downarrow & & \partial_{n+1} \downarrow & & \partial_{n+1} \downarrow & & \partial_{n+1} \downarrow & & \partial_{n+1} \downarrow \\
 (0) & \xrightarrow{0} & C_n & \xrightarrow{f_n} & D_n & \xrightarrow{g_n} & E_n & \xrightarrow{0} & (0) \\
 \partial_n \downarrow & & \partial_n \downarrow & & \partial_n \downarrow & & \partial_n \downarrow & & \partial_n \downarrow \\
 \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
 \partial_3 \downarrow & & \partial_3 \downarrow & & \partial_3 \downarrow & & \partial_3 \downarrow & & \partial_3 \downarrow \\
 (0) & \xrightarrow{0} & C_2 & \xrightarrow{f_2} & D_2 & \xrightarrow{g_2} & E_2 & \xrightarrow{0} & (0) \\
 \partial_2 \downarrow & & \partial_2 \downarrow & & \partial_2 \downarrow & & \partial_2 \downarrow & & \partial_2 \downarrow \\
 (0) & \xrightarrow{0} & C_1 & \xrightarrow{f_1} & D_1 & \xrightarrow{g_1} & E_1 & \xrightarrow{0} & (0) \\
 \partial_1 \downarrow & & \partial_1 \downarrow & & \partial_1 \downarrow & & \partial_1 \downarrow & & \partial_1 \downarrow \\
 (0) & \xrightarrow{0} & C_0 & \xrightarrow{f_0} & D_0 & \xrightarrow{g_0} & E_0 & \xrightarrow{0} & (0) \\
 \partial_0 \downarrow & & \partial_0 \downarrow & & \partial_0 \downarrow & & \partial_0 \downarrow & & \partial_0 \downarrow \\
 (0) & & (0) & & (0) & & (0) & & (0)
 \end{array}$$

1.6 HOMOLOGIA SIMPLICIAL

Após o estudo dos complexos simpliciais e os complexos de cadeias, estamos muito próximos de trabalhar com a estrutura central de nossa análise, o qual, de certa forma, nos permite caracterizar algumas propriedades topológicas. Estas propriedades topológicas - componentes conexas, buracos, ... - aparecerão quando estudarmos os bordos e os ciclos do complexo de cadeias. Um bom Exemplo, para criar intuição, pode ser visto

na Introdução do capítulo 2 do livro do Hatcher [21], mais precisamente, quando ele inicia a discussão sobre a ideia de Homologia. (aconselha-se fortemente a leitura desta parte da introdução - “The idea of Homology”.)

Ao estudarmos um complexo de cadeia $C = (C_i, \partial_i)_{i \in \mathbb{N}_0}$, estabeleceremos a seguinte notação usual:

(i) $Z(C) = (Z_i(C))_{i \in \mathbb{N}_0}$ a sequência dos R -módulos onde

$$Z_i(C) = \ker(\partial_i), \quad \forall i \in \mathbb{N}_0,$$

i.e., $Z_i(C)$ é o conjunto formado pelos de i -ésimos ciclos;

(ii) $B(C) = (B_i(C))_{i \in \mathbb{N}_0}$ a sequência dos R -módulos onde

$$B_i(C) = \text{Im}(\partial_{i+1}), \quad \forall i \in \mathbb{N}_0,$$

i.e., $B_i(C)$ é o conjunto formado pelos de i -ésimos bordos.

Dos ciclos e bordos vale a seguinte inclusão:

$$B_i(C) \subseteq Z_i(C), \quad \forall i \in \mathbb{N}_0,$$

que é resultado imediato de $\partial_i \circ \partial_{i+1} = 0$. Agora, dado que $\partial(Z_i(C)) = (0)$, podemos estabelecer o complexo de cadeia

$$Z(C) = (Z_i(C), \partial_i|_{Z_i(C)})_{i \in \mathbb{N}_0}$$

como vimos no final da Seção anterior. Além do mais, dado que o conjunto dos i -ésimos bordos é submódulo do conjunto formado pelos i -ésimos ciclos, podemos quocientar $Z_i(C)$ por $B_i(C)$, gerando o complexo de cadeia

$$H(C) = (H_i(C), \partial_i)_{i \in \mathbb{N}_0},$$

onde

$$H_i(C) = Z_i(C)/B_i(C) \quad \text{e} \quad \partial_i(x + B_i(C)) = \partial_i(x),$$

o qual sabemos, pela seção anterior, que de fato é um complexo de cadeia. Este complexo de cadeia será chamado de *grupo de homologia simplicial* do complexo C .

Observe que estamos trabalhando com complexos de cadeias induzidas por algum complexo simplicial $\Phi = (\mathcal{K}, S)$. A seguir, estabeleceremos outro tipo importante de grupo de homologia, que será de grande valia para nosso estudo, que são os *grupos de*

homologia singular. Para isto, começamos definindo os *complexos de cadeia singular*, que será definido, no final das contas, de forma similar ao caso simplicial.

Seja X um espaço topológico fixado. Dizemos que toda função

$$\sigma : \Delta^n \rightarrow X,$$

com Δ^n o simplexo canônico da Definição 1.1.2, será um *simplexo singular de dimensão n* quando σ for uma função contínua. Desta forma, podemos definir para todo $i \in \mathbb{N}_0$ e para todo anel com unidade R , os R -módulos

$S_i :=$ módulo livre gerado por $\{\sigma : \Delta^i \rightarrow X; \sigma \text{ é um simplexo singular de dimensão } i\}$,

que nos permite definir a sequência de R -módulos $(S_i)_{i \in \mathbb{N}_0}$, chamada de *sequência de módulos livres gerados pelo simplexos singulares*. Agora vamos definir homomorfismos entre estes módulos, da mesma forma que fizemos para os complexos de cadeia simpliciais, e que serão chamados também de *operadores bordos*.

Sabemos, pela propriedade universal dos módulos livres, que para definirmos estes os homomorfismos basta definirmos seu comportamento na base dos módulos livres. Sendo assim, para cada índice $i \in \mathbb{N}_0$ definimos os homomorfismos

$$\partial_0 : S_0 \rightarrow (0), x \mapsto 0$$

e, para $i > 0$,

$$\partial_i : S_i \rightarrow S_{i-1}$$

vale, para todo elemento $\sigma : \Delta^i \rightarrow X$ da base de S_i ,

$$\partial_i(\sigma) = \sum_{j=0}^i (-1)^j \sigma|_{[e_0, e_1, \dots, \widehat{e}_j, \dots, e_i]},$$

onde $e_0, e_1, \dots, e_i$ são elementos da base canônica de $\mathbb{R}^{i+1}$, $[e_0, e_1, \dots, \widehat{e}_j, \dots, e_i]$ é a face do simplexo canônico Δ^i , com zero na j -ésima coordenada, e $\sigma|_{[e_0, e_1, \dots, \widehat{e}_j, \dots, e_i]}$ é a função σ restrita a esta face.

Observação 1.6.1. Dado o simplexo singular de dimensão n e $[e_0, e_1, \dots, \widehat{e}_j, \dots, e_n]$ uma face de Δ^n . Note que $\sigma|_{[e_0, e_1, \dots, \widehat{e}_j, \dots, e_n]}$ não pertence ao módulo S_{n-1} . De fato, o que fazemos aqui é um abuso de notação, permitido pelo fato de existir um homomorfismo $\xi : \Delta^{n-1} \rightarrow [e_0, e_1, \dots, \widehat{e}_j, \dots, e_n]$ tal que

$$\sigma|_{[e_0, e_1, \dots, \widehat{e}_j, \dots, e_n]} = \sigma \circ \xi \in S_{n-1}.$$

Ou seja, sempre quando escrevermos $\sigma|[e_0, e_1, \dots, \hat{e}_j, \dots, e_n]$, estaremos, de fato, nos referindo a função $\sigma \circ \xi$. Tal abuso de notação tem a utilidade de preservar a simplicidade nas contas.

Sendo assim, dado um espaço topológico X , podemos definir

Definição 1.6.1 (*complexo de cadeia singular*). Seja X um espaço topológico. Chamamos toda tupla $S = (S_i, \partial_i)_{i \in \mathbb{N}_0}$ de complexo de cadeia singular se, $(C_i)_{i \in \mathbb{N}_0}$ for uma sequência de módulos livres gerados pelos simplexes singulares, e $(\partial_i)_{i \in \mathbb{N}_0}$ for uma sequência de operadores bordos

$$\partial_{i+1} : S_{i+1} \rightarrow S_i,$$

tal qual definida acima.

Acontece que complexos de cadeia singulares serão, de fato, complexos de cadeias, i.e., para todo $i \geq 1$ vale

$$\partial_i \circ \partial_{i+1} = 0.$$

A demonstração desta propriedade é semelhante a de se mostrar que complexos de cadeia simpliciais são complexos de cadeia.

Sendo assim, chegamos a dois tipos de complexos de cadeia:

(i) *complexos de cadeia simpliciais*;

(ii) *complexos de cadeia singulares*.

Note que, para complexos de cadeia singulares, definimos os *grupos de homologia singulares* da mesma forma com o qual faríamos para os complexos de cadeia simpliciais, i.e., o grupo de homologia singular de um espaço topológico X é o complexo de cadeia $H(X) = (H_i(X), \partial_i)_{i \in \mathbb{N}_0}$ tal que, para todo $i \geq 0$, temos os grupos de homologia sendo os R -módulos (R um anel com unidade)

$$H_i(X) = Z_i(X)/B_i(X)$$

com

$$Z_i(X) = \ker(\partial_i),$$

o conjunto dos ciclos de dimensão n e

$$B_i(X) = \text{Im}(\partial_{i+1})$$

o conjunto dos bordos de dimensão n . Além do mais, os operadores bordos serão dados pelas projeções canônicas:

$$\bar{\partial}_n : H_n(X) \rightarrow H_{n-1}(X), \quad x + B_n(X) \mapsto \partial_n(X).$$

Em geral, denotaremos os operadores bordo dos grupos de homologia singulares como apenas ∂ , para facilitar a notação.

Antes de partirmos para os exemplos, enunciaremos três resultados sobre grupos de homologia. O primeiro resultado nos garante compararmos grupos de homologias singulares entre espaços topológicos com o mesmo tipo de homotopia:

Proposição 1.6.1. *Sejam X e Y espaços topológicos. Então, se X e Y forem homotopicamente equivalentes teremos que seus grupos de homologia singulares serão isomorfos. Em outras palavras,*

$$X \cong Y \implies H_n(X) \cong_A H_n(Y),$$

onde $\cong$ representa que X e Y são espaços com o mesmo tipo de homotopia e $\cong_A$ representa o isomorfismo algébrico de R -módulos.

Demonstração. Vide Hatcher[21], pg. 111, Corolário 2.11. □

Em seguida, temos o resultado que estabelece a conexão entre grupos de homologia singular e simplicial:

Proposição 1.6.2. *Seja X um espaço topológico triangularizável, i.e., existe um complexo singular $\Phi = (K, S)$ tal que X é homotopicamente equivalente a realização geométrica do complexo simplicial Φ . Então, considerando $H(X)$ como o complexo de cadeias dos grupos de homologia singular e $H(\Phi)$ o complexo de cadeias dos grupos de homologia simpliciais, temos que*

$$H_n(X) \cong_A H_n(\Phi),$$

para todo $n \in \mathbb{N}_0$, com $\cong_A$ representando o isomorfismo algébrico entre módulos.

Demonstração. A demonstração deste fato pode ser vista em Hatcher[21], pg 128, Teorema 2.27. Porém, para se chegar a este resultado é necessário de uma série de resultados preliminares, envolvendo a ferramenta chamada de **divisão baricêntrica**, a qual pode ser encontrada na Seção sobre ecisões, pg 119 do mesmo autor. Ou, se preferir, na Seção 4 do Capítulo 3 do livro de Elon [18], pode-se encontrar o tratamento sobre divisão baricêntrica. □

Estes dois resultados nos dão a liberdade de estudarmos os grupos de homologia sobre duas perspectivas, sempre quando o espaço topológico for triangularizável. Além do mais, utilizando ferramentas como contração, ou retratos por deformação, podemos simplificar, e muito, as contas para calcularmos os grupos de homologia.

Finalmente, enunciamos o resultado que nos garante, de certa forma, caracterizar grupos de homologia de dimensão 0 como a quantidade de componentes conexas do espaço em estudo:

Proposição 1.6.3. *Seja X um espaço topológico. Se X for conexo por caminhos, então seu grupo de homologia singular e de dimensão 0 será isomorfo ao anel dos escalares. Anel este associado aos módulos dos grupos de homologia. Ou seja, dado o complexo de cadeia $H(X) = (H_n(X), \partial_n)_{n \in \mathbb{N}_0}$, com $H_n(X)$ o grupo de homologia singular e de dimensão n , que nada mais são que R -módulos (R um anel com unidade), então*

$$H_0(X) \cong_A R,$$

onde o anel R é visto como R -módulo e $\cong_A$ o isomorfismo algébrico entre módulos. Além do mais, escrevendo o espaço topológico como a união disjunta de suas componentes conexas

$$X = \bigcup_{i \in L} X_i,$$

com L um conjunto de índices e $X_i \subseteq X$ suas componentes conexas, vale que o grupo de homologia singular de dimensão 0 satisfaz

$$H_0(X) \cong_A \bigoplus_{i \in L} R_i, \quad R_i \cong_A R.$$

Sabemos que ao estudarmos os grupos de homologia singulares de dimensão 0 de algum espaço topológico X estamos, de fato, contando a quantidade de componentes conexas deste espaço. Vejamos, agora, como os grupos de homologia se relacionam com os “buracos” do espaço X . Para isto, utilizaremos o seguinte exemplo:

Exemplo 1.6.1 (obs: aqui consideramos R um anel com unidade). Consideremos o conjunto formado pelas palavras do alfabeto, descrito no Exemplo 1.3.4. Como visto, este conjunto pode ser decomposto em três classes de equivalência, com a relação de equivalência dada pela equivalência homotópica. Destas três classes de equivalência, destacamos as três palavras homotopicamente distintas:

$$O, \quad B, \quad I.$$

Note que para estas palavras podemos destacar as seguintes características:

$$O \rightarrow \begin{cases} \text{Uma componente conexa;} \\ \text{Um "buraco";} \end{cases}$$

$$B \rightarrow \begin{cases} \text{Uma componente conexa;} \\ \text{Dois "buracos";} \end{cases}$$

$$I \rightarrow \begin{cases} \text{Uma componente conexa;} \\ \text{Nenhum "buraco"}. \end{cases}$$

Calculemos os grupos de homologia destas letras. Note que a letra I é uma contração, i.e., tem o mesmo tipo de homotopia de um ponto. Ou seja, temos

$$I \cong \{x\}$$

com x um ponto qualquer. Os grupos de homologia singular do conjunto $\{x\}$ são os triviais

$$H_n(\{x\}) \cong_A \begin{cases} R & n = 0; \\ (0) & \text{caso contrário} \end{cases} \xrightarrow{\text{por } I \cong \{x\}} H_n(I) \cong_A \begin{cases} R & n = 0; \\ (0) & \text{caso contrário}. \end{cases}$$

Para a letra O , vamos triangularizá-la pelo complexo simplicial $\Phi_1 = (\mathcal{K}_1, \mathcal{S}_1)$, o qual satisfaz

$$\mathcal{K}_1 = \{x, y, z\}$$

e

$$\mathcal{S}_1 = \{\emptyset, \{x\}, \{y\}, \{z\}, \{x, y\}, \{x, z\}, \{y, z\}\}.$$

Note que a representação geométrica K_{Φ_1} terá o mesmo tipo de homotopia da letra O no $\mathbb{R}^2$. De fato, temos que K_{Φ_1} será o simplexo formado pelas faces do simplexo canônico $\Delta^2 \subseteq \mathbb{R}^3$ o qual, por sua vez, é homotopicamente equivalente ao complexo simplicial K'_{Φ_1} formado por simplexos de dimensão 0, 1 do $\mathbb{R}^2$, como pode ser visto no triângulo circunscrito pela letra O , abaixo na Figura 12.

Calculemos os grupos de homologia da triangularização da letra O . Para isto, teremos que estudar os operadores bordo do complexo de cadeias simplicial induzido por Φ_1 . Começamos com os ciclos de dimensão 1, dados $\alpha, \beta, \gamma \in R$, vale

$$\begin{aligned} \partial_1(\alpha[x, y] + \beta[x, z] + \gamma[y, z]) &= \alpha([y] - [x]) + \beta([z] - [x]) + \gamma([z] - [y]) \\ &= [x](-\alpha - \beta) + [y](\alpha - \gamma) + [z](\beta + \gamma). \end{aligned}$$

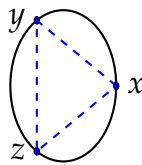


Figura 12: Triangularização da letra O.

Portanto,

$$\partial_1(\alpha[x, y] + \beta[x, z] + \gamma[y, z]) = 0 \iff \alpha = -\beta = \gamma$$

ou seja,

$$H_1(\Phi_1) = Z_1(\Phi_1)/B_1(\Phi_1) \cong_A R.$$

Como sabemos que a homologia singular da letra O é isomorfa a homologia simplicial da triangularização K_{Φ_1} , concluímos que

$$H_0(O) \cong_A R, \quad H_1(O) \cong_A R, \quad \text{e} \quad H_n(O) \cong_A (0), \quad \forall n \geq 2.$$

(obs, a homologia de dimensão 0 não calculamos pois, sabemos que a homologia singular de dimensão zero nada mais é que a contagem de componentes conexas, e a letra O tem apenas uma componente conexa.)

Finalmente, calculemos a homologia da letra B, o qual é triangularizada pelo complexo simplicial $\Phi_2 = (\mathcal{K}_2, \mathcal{S}_2)$, composto pelos vértices

$$\mathcal{K}_2 = \{v, x, y, z, w\}$$

e simplexes

$$\mathcal{S}_2 = \{\emptyset, \{v\}, \{x\}, \{y\}, \{z\}, \{w\}, \{v, x\}, \{v, y\}, \{x, y\}, \{y, z\}, \{y, w\}, \{z, w\}\}$$

cuja representação geométrica pode ser vista na Figura 13. Calculemos o grupo de

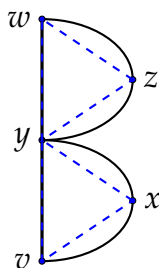


Figura 13: Triangularização da letra B.

homologia de dimensão 1. Dados os escaleres $\alpha_1, \alpha_2, \dots, \alpha_6$ e o ponto

$$p = \alpha_1[v, x] + \alpha_2[v, y] + \alpha_3[x, y] + \alpha_4[y, z] + \alpha_5[y, w] + \alpha_6[z, w],$$

temos que

$$\begin{aligned} \partial_1(p) &= \alpha_1([x] - [v]) + \alpha_2([y] - [v]) + \alpha_3([y] - [x]) + \alpha_4([z] - [y]) + \alpha_5([w] - [y]) + \\ &\quad + \alpha_6([w] - [z]) \\ &= [v](-\alpha_1 - \alpha_2) + [x](\alpha_1 - \alpha_3) + [y](\alpha_2 + \alpha_3 - \alpha_4 - \alpha_5) + [z](\alpha_4 - \alpha_6) + [w](\alpha_5 + \alpha_6). \end{aligned}$$

Logo,

$$\partial_1(p) = 0 \iff \alpha_1 = -\alpha_2 = \alpha_3, \quad \text{e} \quad \alpha_6 = -\alpha_5 = \alpha_4.$$

Desta forma temos que $Z_1(\Phi_2) \cong_A R \oplus R$ e $B_1(\Phi) = (0)$. Portanto, os grupos de homologia da letra B são

$$H_0(B) \cong_A R, \quad H_1(B) \cong_A R \oplus R, \quad \text{e} \quad H_n(B) \cong_A (0), \quad \forall n \geq 2.$$

O Exemplo 1.6.1 nos mostra como o grupo de homologia de dimensão 1 conta a quantidade de buracos contornados por ciclos de dimensão 1, de um dado espaço topológico. Da letra l seguimos para a letra O, que possui um buraco em seu centro, e terminamos com a letra B, o qual adiciona-se um buraco a mais em relação a letra O.

Agora vamos calcular os grupos de homologia dos grafos apresentados na Figura 3. Sabemos que estes grafos são homotopicamente equivalentes aos apresentados na Figura 10, o qual são mais fáceis de calcular os operadores bordo. Como mostrado na Figura 10, a árvore binária é uma contração, ou seja, seus grupos de homologia singulares são isomorfos ao espaço topológico definido por apenas um ponto, o qual admite homologia trivial, como já mostrado. Calculemos então os grupos de homologia do grafo da Figura 10a, que denotaremos por G . Utilizando nossa intuição, diríamos que este grafo tem os seguintes grupos de homologia (neste exemplo utilizamos os reais como anel):

$$H_0(G) \cong \mathbb{R}, \quad H_1(G) \cong \mathbb{R}^6, \quad H_n(G) \cong (0), \quad \forall n \geq 2.$$

De fato, esta intuição está correta, pois como já calculado no Exemplo 1.5.3, temos que

$$Z_1(G) \cong \mathbb{R}^6, \quad B_1(G) = (0) \implies H_1(G) \cong \mathbb{R}^6.$$

Esta intuição geométrica dos grupos de homologia pode continuar. Por exemplo, em dimensões maiores, grupos de homologia podem contar o “vazio” coberto por

alguma superfície. Para fixarmos ideia, pensemos no Toro da Figura 4. Pelo exposto no Exemplo 1.5.4, temos que os grupos de homologia do Toro de dimensão 0, 1 serão

$$H_0(S_1 \times S_1) \cong_A \mathbb{R} \quad \text{e} \quad H_1(S_1 \times S_1) \cong_A \mathbb{R} \oplus \mathbb{R},$$

com $S_1 \times S_1$ o Toro, pois o Toro possui uma componente conexa e dois buracos, um em seu centro e outro sendo o resultado da rotação de uma circunferência em relação ao eixo z . Além do mais, decorre desta rotação que um “vazio” se formará dentro desta superfície, sendo captado pelo grupo de homologia de dimensão 2:

$$Z_2(S_1 \times S_1) \cong \mathbb{R} \implies H_2(S_1 \times S_1) \cong_A \mathbb{R}.$$

Para dimensões maiores os grupos de homologia do Toro são os espaços vetoriais nulos.

Os grupos de homologia desempenharão um papel fundamental em nossa teoria, já que eles serão os responsáveis em capturar as informações topológicas relevantes para todo conjunto de amostra que formos avaliar mais a frente. Por isso, é bom ter em mente estas intuições topológicas dos grupos de homologia.

A seguir, listaremos alguns resultados auxiliares sobre grupos de homologia, que podem facilitar nossos cálculos.

1.7 INDUZINDO MORFISMOS NA CATEGORIA DOS COMPLEXOS DE CADEIAS

Nesta Seção apresentaremos alguns morfismos importantes na Categoria dos Complexos Simpliciais e Complexos Singulares. Sejam $K^1 = (\mathcal{K}_1, \mathcal{S}_1)$ e $K^2 = (\mathcal{K}_2, \mathcal{S}_2)$ complexos simpliciais e $f : K^1 \rightarrow K^2$ um morfismo entre estes complexos simpliciais, i.e.,

$$f : K^1 \rightarrow K^2,$$

com $f(s) = \langle f(a_0), f(a_1), \dots, f(a_n) \rangle \in \mathcal{S}_2$, para todo $s = \langle a_0, a_1, \dots, a_n \rangle \in \mathcal{S}_1$. Sabemos que estes complexos simpliciais induzem os complexos de cadeias simpliciais $C(K^1) = (C_i, \partial_i)_i$, $C(K^2) = (D_i, \partial_i)_i$. Logo, para todo natural $i \in \mathbb{N}_0$, definimos os homomorfismos de módulos

$$f_{\#,i} : C_i \rightarrow D_i, \quad f_{\#,i}(s) = f(s), \quad \forall s \in \mathcal{S}_1.$$

(Note que o homomorfismo acima existe pois o definimos na base dos módulos C_i , i.e., estão definidos nos i -simplexos.) Agora, tome um n -simplexo $s = \langle a_0, a_1, \dots, a_n \rangle \in \mathcal{S}_1$, com $n > 0$. Temos que

$$\begin{aligned} \partial(f_{\#,n}(s)) &= \sum_{i=0}^n (-1)^i [f(a_0), f(a_1), \dots, \widehat{f(a_i)}, \dots, f(a_n)] = \sum_{i=0}^n (-1)^i (f_{\#,n-1}[a_0, a_1, \dots, \widehat{a_i}, \dots, a_n]) = \\ &= f_{\#,n-1}(\sum_{i=0}^n (-1)^i [a_0, a_1, \dots, \widehat{a_i}, \dots, a_n]) = f_{\#,n-1}(\partial(s)). \end{aligned}$$

Ou seja,

$$\partial \circ f_{\#,n} = f_{\#,n-1} \circ \partial$$

e, portanto, temos que o morfismo $f_{\#} = (f_{\#,i})_{i \in \mathbb{N}_0}$ é um morfismo entre as cadeias de complexo $C(K^1)$ e $C(K^2)$. Além disso, de maneira semelhante, podemos induzir um morfismo entre os grupos de homologia simplicial:

$$f_* = (f_{*,i} : H_i(K^1) \rightarrow H_i(K^2))_{i \in \mathbb{N}_0}.$$

Em diagramas, isto equivale ao seguinte

$$\begin{array}{c} K^1 \\ \downarrow f \\ K^2 \end{array} \quad \Longrightarrow \quad \begin{array}{ccccccc} \cdots & \longrightarrow & C_2 & \xrightarrow{\partial} & C_1 & \xrightarrow{\partial} & C_0 \xrightarrow{\partial} (0) \\ & & \downarrow f_{\#,2} & & \downarrow f_{\#,1} & & \downarrow f_{\#,0} \\ \cdots & \longrightarrow & D_2 & \xrightarrow{\partial} & D_1 & \xrightarrow{\partial} & D_0 \xrightarrow{\partial} (0) \end{array}$$

e

$$\begin{array}{c} K^1 \\ \downarrow f \\ K^2 \end{array} \quad \Longrightarrow \quad \begin{array}{ccccccc} \cdots & \longrightarrow & H_2(K^1) & \xrightarrow{\partial} & H_1(K^1) & \xrightarrow{\partial} & H_0(K^1) \xrightarrow{\partial} (0) \\ & & \downarrow f_{*,2} & & \downarrow f_{*,1} & & \downarrow f_{*,0} \\ \cdots & \longrightarrow & H_2(K^2) & \xrightarrow{\partial} & H_1(K^2) & \xrightarrow{\partial} & H_0(K^2) \xrightarrow{\partial} (0) \end{array}$$

De forma semelhante, dados dois espaços topológicos X e Y e uma função contínua $g : X \rightarrow Y$, podemos induzir morfismos entre os complexos de cadeias singulares $C(X)$ e $C(Y)$ e podemos induzir morfismos entre os grupos de homologia singulares destas cadeias. Denotaremos sempre os morfismos entre cadeias complexas singulares induzidos pela função g como

$$g_{\#} : C(X) \rightarrow C(Y)$$

e denotaremos os morfismos entre os grupos de homologia singulares induzidos por g como

$$g_{*} : H(X) \rightarrow H(Y).$$

Usaremos bastante, no próximo Capítulo, estes morfismos de complexos de cadeias, induzidos por morfismos da categoria dos complexos simpliciais e das topologias.

2

HOMOLOGIA PERSISTENTE

Como visto no Capítulo anterior, os grupos de homologia são uma importante ferramenta para capturar as propriedades de um espaço topológico. Nada mais natural, então, do que utilizarmos este instrumental matemático para estudarmos os padrões topológicos de um conjunto de dados, para assim inferirmos suas características. Em outras palavras, sejam $\mathcal{X}$ um espaço topológico e $X \subseteq \mathcal{X}$ um subconjunto finito, usualmente chamado de *nuvem de pontos de dados*, estaremos interessados em inferir as propriedades topológicas de $\mathcal{X}$ através da nuvem de pontos (i.e., nossa amostra X). Estas propriedades topológicas serão das mais diversas, como a quantidade de componentes conexas, quantidade de ciclos em um grafo, entre outras.

Para esta tarefa, recorreremos ao uso da ferramenta *Homologia Persistente*. A Homologia Persistente, subramo da área de *Análise topológica de Dados*, e abreviada simplesmente por *TDA*, tem como principal objetivo o de utilizar o conceito de Homologia (seja ela singular ou simplicial) para analisar as propriedades topológicas de uma nuvem de pontos de dados. Um importante aspecto da homologia persistente, reside no fato que os grupos de homologia estudados serão parametrizados por uma variável τ , que nos permitirá inferir o quão relevante é certa propriedade topológica, na nuvem de dados em questão.

Neste capítulo, formalizaremos o conceito de homologia persistente, e mostraremos como o mesmo pode ser usado para extrair informações de uma nuvem de dados, resumizando estas informações muito comumente por meio de um diagrama, chamado de *diagrama de persistência*. Em seguida, estabeleceremos uma série de exemplos de diagramas de persistência, e mostraremos como esta técnica pode ser aplicada para uma série de problemas práticos. Finalmente, para uma visão mais ampla deste tópico, tal qual para o estudo mais aprofundado de Análise de Dados Topológicos, recomendamos a leitura dos livros *Persistence Theory: From Quiver Representations to Data Analysis*, do autor Steve Y. Oudot (vide a referência [26] - especialmente os três primeiros capítulos, para uma compreensão mais formal), e *Geometric and Topological Inference*, dos autores Boissonnat, Chazal e Yvinec [27] (capítulo 11 - este livro tem enfoque maior no estudo

dos algoritmos de TDA). Estes livros serviram de base para o desenvolvimento deste capítulo. Além disso, recomendamos, também, a leitura de [28] e as notas de aula [29].

2.1 MOTIVAÇÃO

Antes de apresentarmos os conceitos de homologia persistente, comecemos discutindo um problema prático de classificação, apresentando sua forma usual de resolvê-lo. Em seguida, apontaremos uma outra perspectiva para resolver o mesmo problema, baseada na homologia, e servindo de motivação para o uso das propriedades topológicas para o problema de análise de dados.

O problema que utilizaremos como ponto de partida consiste em uma breve avaliação do banco de dados ¹ coletado pelo Instituto de Estudos do Coração de Framingham ². A informação contida neste banco de dados resume-se, basicamente, na avaliação do risco de pacientes desenvolverem, em um período de 10 anos, doenças na artéria coronária, tomando como base valores como: idade, sexo, pressão arterial, taxa de colesterol, e outras (totalizando 16 variáveis).

Para simplificarmos nossa discussão, consideraremos apenas uma amostra de 50 indivíduos, obtida a partir do banco de dados acima. Neste banco de dados, as variáveis que observaremos serão: o total de gordura no sangue; a taxa de glicose no sangue; e o risco do indivíduo adquirir, em até 10 anos, doenças arteriais coronárias, denotada por $CHD10$. (Por convenção, $CHD10 = 0$ significará que o indivíduo apresenta baixo risco de adquirir doenças arteriais, enquanto que $CHD10 = 1$ significará alto risco.) Na Figura 14(a) podemos ver o padrão de nossa amostra de dados.

Temos agora o seguinte problema: dada a amostra

$$X = \{(x_i, y_i); i = 1, 2, \dots, 50\},$$

com x_i o vetor aleatório

$$x_i = (\text{Colesterol Total, Level de Glicose})$$

e

$$y_i = CHD10$$

¹ obtido no site <https://www.kaggle.com/>.

² sítio do projeto <https://framinghamheartstudy.org/>.

a variável aleatória referente ao risco do indivíduo adquirir, ou não, doenças arteriais (Aqui, ambas as variáveis x_i e y_i são indexadas pelo i -ésimo indivíduo da amostra). Gostaríamos, então, de encontrar uma função estimadora,

$$\hat{h} : \mathbb{R} \times \mathbb{R} \rightarrow \{0, 1\},$$

contida em um conjunto $\mathcal{H}$ (conjunto este formado pelas funções que assumimos descreverem bem o fenômeno em estudo), tal que o estimador goze da seguinte propriedade

$$\hat{h} = \arg \min_{h \in \mathcal{H}} \left\{ \sum_{i=1}^{50} l(y_i, h(x_i)) \right\},$$

com $l : \{0, 1\}^2 \rightarrow \mathbb{R}_{>0}$ uma função, definida a priori, e que serve para mensurar o erro induzido pelas predições das funções de classificação em $\mathcal{H}$.

Uma forma de resolver este problema, e usualmente aplicada em Aprendizado de Máquinas, é aplicar a técnica de Regressão Logística a estes dados, que nada mais é que encontrar um hiperplano que separe os pontos suficientemente bem, minimizando a função erro

$$\text{Erro}(w, b) = \frac{\sum_{i=1}^{50} \log(1 + \exp(\langle w, x_i \rangle + b))}{50}, \quad (\hat{w}, \hat{b}) = \arg \min_{(w, b)} \{\text{Erro}(w, b)\}, \quad (1)$$

com $w \in \mathbb{R}^2$ e $b \in \mathbb{R}$. Na Figura 14(b), temos um exemplo de um hiperplano separando os pontos amostrais. Note que, dado um indivíduo qualquer, cujas características pertinentes ao problema possam ser representadas pelo vetor

$$(\text{Colesterol Total, Level de Glicose}) = x \in \mathbb{R}^2,$$

vale que

$$\hat{h}(x) = \begin{cases} \text{indivíduo mais predisposto a doenças do coração,} & \text{se } \langle \hat{w}, x \rangle + \hat{b} \geq 0 \\ \text{indivíduo menos predisposto a doenças do coração,} & \text{se } \langle \hat{w}, x \rangle + \hat{b} < 0, \end{cases}$$

com $\hat{h}$ a função de classificação escolhida pelo algoritmo de Regressão Logística. Neste problema em específico, um indivíduo será classificado como alguém com alto, ou baixo, risco de desenvolver problemas cardíacos, dependendo da orientação de x em relação ao hiperplano escolhido. Para mais sobre Regressão Logística, vide o Capítulo 9, do livro Understanding Machine Learning [30]).

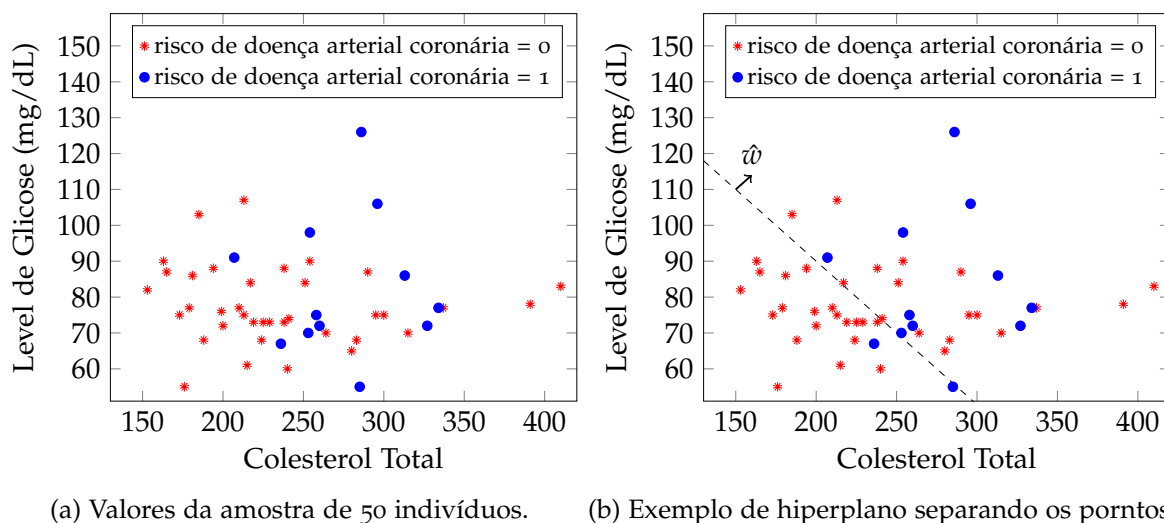


Figura 14: Gráficos (a): Level de Glicose em função do Colesterol total. Cada ponto é categorizado pelo fator que define o risco do indivíduo de desenvolver doenças arteriais coronárias em 10 anos ou mais. Gráfico (b): mesmo Gráfico do item (a), porém adicionado um hiperplano separando os pontos.

Observação 2.1.1. Na Regressão Logística estamos trabalhando com o problema de encontrarmos um hiperplano tal que minimizamos a função erro da equação 1, ou seja, estamos considerando como informação apenas a distância de nossa amostra para o hiperplano, desprezando informações referentes a, por exemplo, distância entre os pontos da amostra, ou, como os pontos se distribuem no espaço, destacando padrões geométricos.

Avaliemos novamente os dados da Figura 14(a), só que agora sob uma nova perspectiva. Ao invés de tentarmos separar os pontos em dois conjuntos disjuntos, iremos focar no comportamento local dos pontos.

Dado $\varepsilon > 0$ um real positivo qualquer e fixado. Estaremos interessados no comportamento dos grafos não orientados

$$G_\varepsilon^0 = (X^0, A^0) \quad \text{e} \quad G_\varepsilon^1 = (X^1, A^1)$$

que satisfzem a seguinte condição:

$$x_i \in X^0 \subseteq X \iff \begin{cases} \text{o } i\text{-ésimo indivíduo não apresenta risco} \\ \text{de desenvolver doenças da artéria coronária} \end{cases}$$

e

$$x_i \in X^1 \subseteq X \iff \begin{cases} \text{o } i\text{-ésimo indivíduo apresentar alto risco} \\ \text{de desenvolver doenças da artéria coronária,} \end{cases}$$

juntamente com a restrição conjunta para as arestas

$$\begin{aligned} \{x_i, x_j\} \in A^0 &\iff x_i, x_j \in X^0 \text{ e } d(x_i, x_j) < \varepsilon, \\ \{x_i, x_j\} \in A^1 &\iff x_i, x_j \in X^1 \text{ e } d(x_i, x_j) < \varepsilon, \end{aligned}$$

com $i, j = 1, 2, \dots, 50$ e $d(\cdot, \cdot)$ a distância euclidiana. Ou seja, estamos interessados em grafos cujas arestas $\{x_i, x_j\}$ satisfazem a seguinte condição: bolas abertas de raio ε , e centradas em x_i e x_j , não podem ter interseção vazia, i.e.,

$$B(x_i, \varepsilon) \cap B(x_j, \varepsilon) \neq \emptyset.$$

Destas estruturas, ganhamos a vantagem de podermos avaliar a informação do comportamento local dos pontos amostrais. Para ilustrarmos o comportamento local destes pontos, fixemos um caso concreto. Consideremos ε um real positivo tal que, para todo os pontos $x_i, x_j \in X^0$ (ou $x_i, x_j \in X^1$), vale a seguinte condição

$$d(x_i, x_j) < \varepsilon \iff \begin{cases} \text{Colesterol Total}_i - \text{Colesterol Total}_j < 50 \\ \text{Level de Glicose}_i - \text{Level de Glicose}_j < 10\text{mg/dL}. \end{cases}$$

Na Figura 15 temos a representação gráfica deste caso exemplo.

Um fato curioso dos grafos $G_\varepsilon^0, G_\varepsilon^1$ são os padrões que eles apresentam. No gráfico da Figura 15(b), referente aos indivíduos com baixo risco de desenvolverem doenças da artéria coronária, percebemos a presença de uma componente conexa, que se destaca por uma quantidade significativa de pontos. Da mesma forma, é possível destacarmos pelo menos 5 ciclos de dimensão 1 que chamam a atenção. Enquanto isso, para o grupo de indivíduos com alto risco de doenças arteriais coronárias, percebemos pontos muito distantes entre si, com apenas um ciclo de dimensão 1.

Note que em nosso exemplo, trabalhamos com um ε fixo, o que está longe de ser o ideal. A partir deste ponto que a ferramenta de *Homologia Persistente* surgirá, já que com esta ferramenta não ficaremos mais presos a um único ε , escolhido arbitrariamente. Ao contrário, dela teremos uma visão global das possíveis conexões, nos permitindo estudar as cadeias do tipo

$$\mathcal{F}^0 = (X_\varepsilon^0)_{\varepsilon \in P \subseteq \mathbb{R}},$$

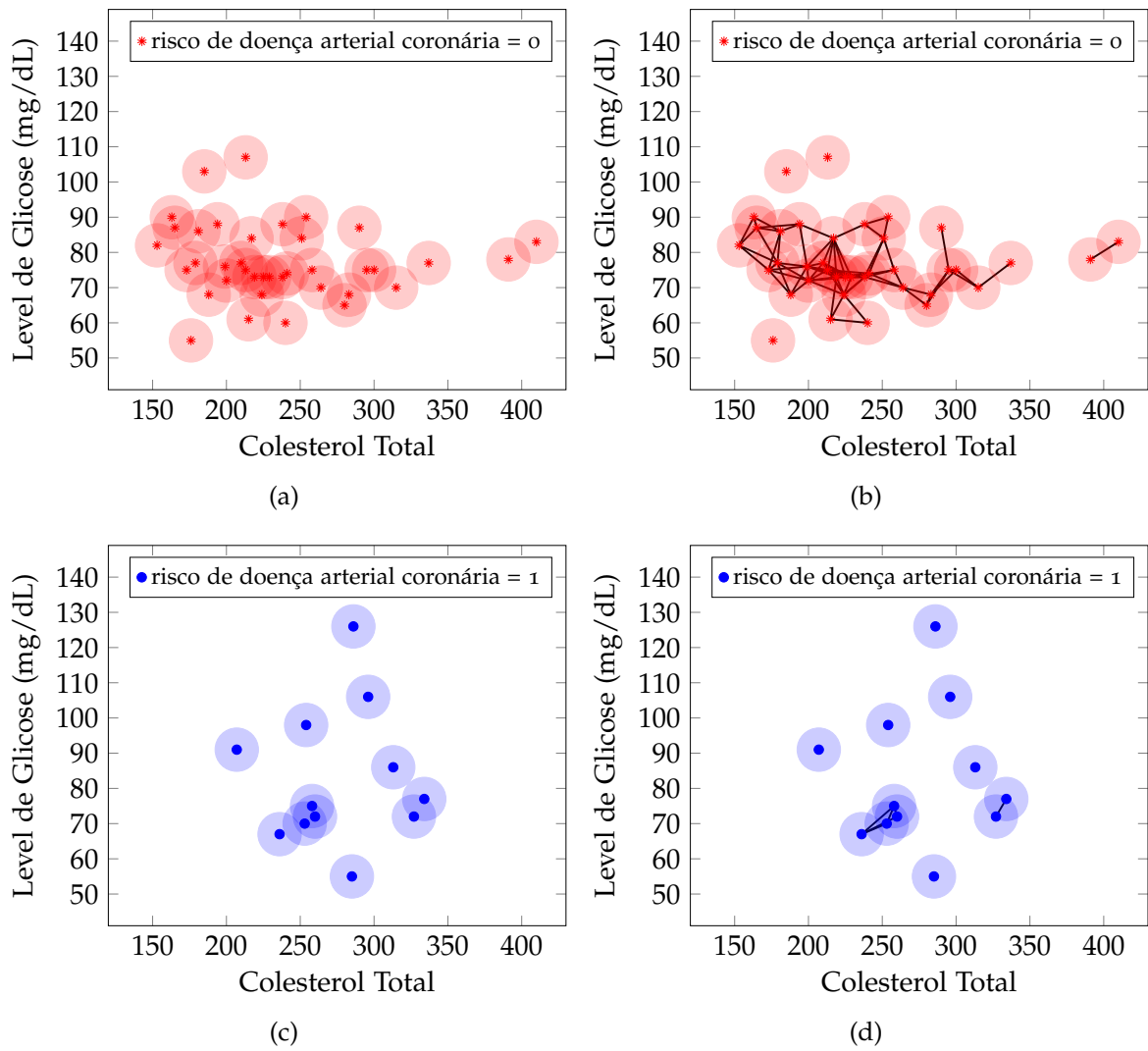


Figura 15: Representação dos grafos $G_\epsilon^0, G_\epsilon^1$, acompanhados das bolas abertas de raio ϵ e centradas em cada ponto amostral.

onde, para todo $a \leq b \leq c \in P$, valem as inclusões

$$X_a^0 \subseteq X_b^0 \subseteq X_c^0 \subseteq X, \quad \text{com } X \text{ um conjunto fixado a priori.}$$

Desta parametrização, o foco da *homologia persistente* será avaliar como os grupos de homologia deste espaço variam com a parametrização P , nos informando, por exemplo, o quanto que propriedades topológicas, que aparecem nestes conjuntos X_ϵ , tem influência no comportamento geral.

Com esta visão inicial do problema, daremos início a formalização da homologia persistente.

2.2 FORMALIZAÇÃO

Iniciemos definindo o que são filtrações, tanto para espaços topológicos triangularizáveis como para complexos simpliciais:

Definição 2.2.1 (Filtrações). Seja $T \subseteq \mathbb{R}$ um sub-conjunto não nulo da reta real. Dizemos que as sequências

$$\mathcal{F}_1 = (X_i)_{i \in T} \quad \text{e} \quad \mathcal{F}_2 = (\Phi_i)_{i \in T},$$

com X_i um espaço topológico triangularizável e Φ_i um complexo simplicial, são *filtrações* se, para todo $i, j \in T$, valer

$$i \leq j \implies \begin{cases} X_i \subseteq X_j \\ \Phi_i \subseteq \Phi_j. \end{cases}$$

No caso de $\mathcal{F}_2$, é usual chamar esta filtração de *filtração simplicial*.

Observação 2.2.1. Sejam $\mathcal{K}_1, \mathcal{K}_2$ conjuntos não nulos e $\mathcal{S}_1$ e $\mathcal{S}_2$ famílias de subconjuntos de $\mathcal{K}_1$ e $\mathcal{K}_2$, respectivamente. Dizemos que um complexo simplicial $\Phi_1 = (\mathcal{K}_1, \mathcal{S}_1)$ está contido em um complexo simplicial $\Phi_2 = (\mathcal{K}_2, \mathcal{S}_2)$ se, e somente se,

$$\mathcal{K}_1 \subseteq \mathcal{K}_2 \quad \text{e} \quad \mathcal{S}_1 \subseteq \mathcal{S}_2.$$

Filtrações exercerão um papel fundamental para o que vem a seguir.

Exemplo 2.2.1. Como mostrado na seção anterior, os grafos X_ε , para todo $\varepsilon \geq 0$, formam uma filtração simplicial. De fato, basta recordarmos que grafos podem ser vistos, também, como complexos simpliciais.

Seja $\mathcal{F} = (X_j)_{j \in T}$ uma filtração parametrizada por $T \subseteq \mathbb{R}$, com cada X_j sendo um espaço topológico triangularizável, ou um complexo simplicial. Sabemos, pelo exposto no final do capítulo anterior, que para todo $a, b \in T$, satisfazendo $a \leq b$, vale que o operador inclusão $i^{a,b} : X_a \hookrightarrow X_b$ induz um morfismo

$$i_*^{a,b} : (i_{*,p} : H_p(X_a) \rightarrow H_p(X_b))_{p \in \mathbb{N}_0} \quad (2)$$

na categoria dos complexos de cadeias. Em outras palavras,

$$\forall a \leq b \in T \quad \text{existe} \quad i_*^{a,b} : H(X_a) \rightarrow H(X_b),$$

com $i_*^{a,b} \in \mathbf{Hom}(H(X_a), H(X_b))$ um morfismo na categoria dos complexos de cadeia $\mathbf{Cx}$, e

$$H(X_a) = (H_p(X_a), \partial_p) \quad \text{e} \quad H(X_b) = (H_p(X_b), \partial_p)$$

os complexos de cadeias formados pelos grupos de homologia e seus operadores bordo.

Lembremos, também, que dados espaços topológicos (ou complexos simpliciais) quaisquer e não nulos, X, Y e Z , vale que para todos os morfismos entre estas estruturas,

$$f : X \rightarrow Y \quad \text{e} \quad g : Y \rightarrow Z,$$

ao aplicarmos o funtor responsável por transportar objetos da categoria dos espaços topológicos (ou, respectivamente, da categoria dos complexos simpliciais) para objetos da categoria dos complexos de cadeia, obtemos os morfismos

$$f_* : H(X) \rightarrow H(Y) \quad \text{e} \quad g_* : H(Y) \rightarrow H(Z),$$

cujos domínios e contra-domínios são grupos de homologia, e satisfazem

$$(g \circ f)_* = g_* \circ f_*.$$

Abstraíndo estas condições para o nosso contexto, temos que para toda filtração $\mathcal{F} = (X_j)_{j \in T}$, parametrizada por $T \subseteq \mathbb{R}$ e cujos conjuntos são não-nulos, que

$$\forall a, b, c \in T, \quad a \leq b \leq c \quad \text{vale} \quad i_*^{a,c} = i_*^{b,c} \circ i_*^{a,b} \implies i_*^{a,c} = i_*^{b,c} \circ i_*^{a,b}, \quad (3)$$

com $i_*^{a,b}, i_*^{b,c}$ os operadores inclusão discutidos no parágrafo acima.

Observação 2.2.2. Esta ligação entre a ordem parcial, dos índices de uma filtração, e os morfismos $i_*^{a,c}$, dos respectivos grupos de homologia, pode ser expressa de forma mais condensada se utilizarmos a linguagem dos funtores. Estes funtores serão o pilares da homologia persistente, que serão estudados mais a frente.

Já vimos no capítulo anterior que os complexos de cadeia, e seus morfismos, satisfazem as condições necessárias para que sejam uma categoria. Resta, então, inserirmos as estruturas das filtrações em alguma categoria, de tal forma que a mesma capture a ordem parcial do conjunto dos índices da filtração.

A categoria que nos fornecerá esta estrutura será a Categoria dos **Posets**. Seja T um conjunto não nulo e que admite uma ordem parcial $\preceq$, então a categoria Poset induzida por T e por sua ordem parcial $\preceq$, denotada por $(T, \preceq)$, será a categoria cujos objetos

são os seus próprios elementos e seus morfismos serão, basicamente, os morfismos definidos pela relação de ordem de T .

Dados objetos quaisquer $a, b \in T$, o morfismo entre estes objetos será sempre um conjunto com no máximo um elemento: $\text{hom}(a, b)$ será não nulo se $a \preceq b$, e vazio caso contrário. Ou seja, os morfismos desta categoria são definidos por $\preceq$ e, usualmente, também denotados por $\preceq$. Finalmente, observemos que para os objetos $a, b, c \in T$ satisfazendo

$$a \preceq b \preceq c,$$

existem morfismos $f_{b,c} \in \text{hom}(b, c)$ e $f_{a,b} \in \text{hom}(a, b)$ tal que sua composta está bem definida

$$f_{a,c} := f_{b,c} \circ f_{a,b}.$$

Agora podemos definir Homologia Persistente:

Definição 2.2.2 (módulos de persistência e homologia persistente). Seja $\mathcal{F} = (X_i)_{i \in T}$ uma filtração parametrizada por $T \subseteq \mathbb{R}$ e cujos conjuntos X_i são todos espaços topológicos triangularizáveis, ou são todos complexos simpliciais. Então, fixado um natural $p \in \mathbb{N}_0$, dizemos que $H_p(\mathcal{F})$ é um *módulo de persistência* de dimensão p , da filtração $\mathcal{F}$, se $H_p(\mathcal{F})$ for um funtor dado por

$$\begin{aligned} H_p(\mathcal{F}) : (T, \preceq) &\rightarrow \mathbf{RMod} \\ T \ni i &\mapsto H_p(X_i) \\ \text{hom}(a, b) \ni a \leq b &\mapsto i_{*,p}^{a,b} : H_p(X_a) \rightarrow H_p(X_b), \end{aligned}$$

com $H_p(X_i)$ o grupo de homologia de dimensão p de X_i , i_* o operador induzido pela inclusão, discutido na equação 2, e $\mathbf{RMod}$ a categoria dos R -módulos à esquerda, com R um anel com unidade. Além do mais, chamamos de *homologia persistente* da filtração $\mathcal{F}$ a família

$$H_*(\mathcal{F}) = (H_p(\mathcal{F}))_{p \in \mathbb{N}_0}$$

de todos módulos de persistência de $\mathcal{F}$.

Ilustremos como calcular a homologia persistente de uma filtração:

Exemplo 2.2.2. Seja $G = (V, A)$ o grafo cíclico, e não orientado, representado pela Figura 16. Calculemos a homologia persistente da seguinte filtração

$$\mathcal{F} = (\Phi_\varepsilon)_{\varepsilon \in \mathbb{R}_{\geq 0}},$$

onde o complexo simplicial Φ_ε é dado por

$$\Phi_\varepsilon = (V, A_\varepsilon), \quad \text{tal que } \forall x, y \in V, \{x, y\} \in A_\varepsilon \iff d(x, y) \leq \varepsilon,$$

e $d(\cdot, \cdot)$ a distância que calcula o menor caminho possível entre os vértices $x, y \in V$ do grafo. Por exemplo, no nosso grafo, temos

$$d(0, 1) = 1, d(0, 2) = 2, d(0, 3) = 2, d(0, 4) = 1, \dots$$

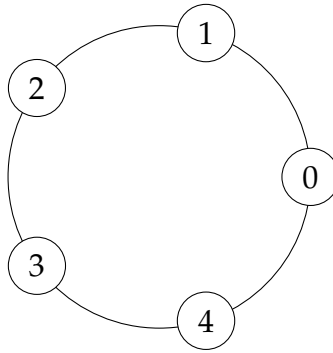


Figura 16

Portanto, a filtração $\mathcal{F}$ será a seguinte

$$\varepsilon \in [0, 1) \implies \Phi_\varepsilon = (V, A_\varepsilon), \text{ com } A_\varepsilon = \{\emptyset, \{0\}, \{1\}, \{2\}, \{3\}, \{4\}\}$$

$$\varepsilon \in [1, 2) \implies \Phi_\varepsilon = (V, A_\varepsilon), \text{ com}$$

$$A_\varepsilon = \{\emptyset, \{0\}, \{1\}, \{2\}, \{3\}, \{4\}, \{0, 1\}, \{1, 2\}, \{2, 3\}, \{3, 4\}, \{0, 4\}\}$$

e

$$\varepsilon \in [2, \infty) \implies \Phi_\varepsilon = (V, A_\varepsilon), \text{ onde}$$

$$\begin{aligned} A_\varepsilon = \{ & \emptyset, \{0\}, \{1\}, \{2\}, \{3\}, \{4\}, \{0, 1\}, \{1, 2\}, \{2, 3\}, \{3, 4\}, \{0, 4\}, \\ & \{0, 2\}, \{0, 3\}, \{1, 3\}, \{1, 4\}, \{2, 4\}, \\ & \{0, 1, 2\}, \{0, 1, 3\}, \{0, 1, 4\}, \{0, 2, 3\}, \{0, 2, 4\}, \{0, 3, 4\}, \{1, 2, 3\}, \\ & \{1, 2, 4\}, \{1, 3, 4\}, \{2, 3, 4\}, \\ & \{0, 1, 2, 3\}, \{0, 1, 2, 4\}, \{0, 1, 3, 4\}, \{0, 2, 3, 4\}, \{1, 2, 3, 4\}, \\ & \{0, 1, 2, 3, 4\} \} \end{aligned}$$

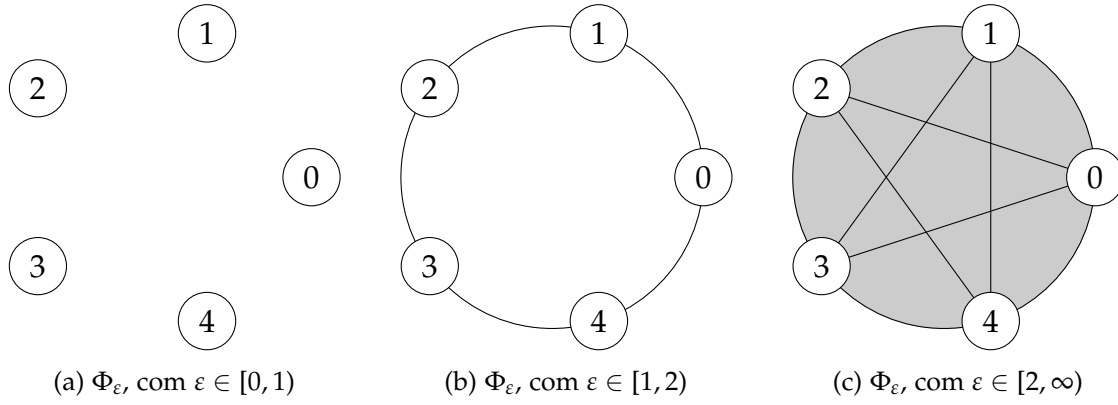


Figura 17: Complexos simpliciais da filtração do grafo da Figura 16

cujas representação gráfica pode ser vista na Figura 17.

Agora calculemos os módulos de persistência para esta filtração, onde os grupos de homologia serão analisados como $\mathbb{R}$ -espaços vetoriais:

- Para todo $\varepsilon \in [0, 1)$ temos

$$H_0(\Phi_\varepsilon) \cong \mathbb{R}^5, \quad H_p(\Phi_\varepsilon) = (0), \quad \forall p \geq 1,$$

- Para todo $\varepsilon \in [1, 2)$ vale

$$H_0(\Phi_\varepsilon) \cong \mathbb{R}, \quad H_1(\Phi_\varepsilon) \cong \mathbb{R}, \quad H_p(\Phi_\varepsilon) = (0), \quad \forall p \geq 2$$

- E, para todo $\varepsilon \in [2, \infty)$, temos as matrizes das transformações lineares, referentes aos operadores bordos de dimensão 1, 2, 3, dadas por ³:

$$\partial_1 = \begin{matrix} & \begin{matrix} [0,1] & [1,2] & [2,3] & [3,4] & [0,4] & [0,2] & [0,3] & [1,3] & [1,4] & [2,4] \end{matrix} \\ \begin{matrix} [0] \\ [1] \\ [2] \\ [3] \\ [4] \end{matrix} & \left(\begin{array}{cccccccccc} -1 & 0 & 0 & 0 & -1 & -1 & -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \end{array} \right) \end{matrix},$$

³ O operador bordo de dimensão 4 não é calculado, pois ele é imediato.

$$\partial_2 = \begin{matrix} & \begin{matrix} [0,1,2] & [0,1,3] & [0,1,4] & [0,2,3] & [0,2,4] & [0,3,4] & [1,2,3] & [1,2,4] & [1,3,4] & [2,3,4] \end{matrix} \\ \begin{matrix} [0,1] \\ [1,2] \\ [2,3] \\ [3,4] \\ [0,4] \\ [0,2] \\ [0,3] \\ [1,3] \\ [1,4] \\ [2,4] \end{matrix} & \left(\begin{array}{cccccccccccc} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & -1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & -1 \end{array} \right) \end{matrix},$$

$$\partial_3 = \begin{matrix} & \begin{matrix} [0,1,2,3] & [0,1,2,4] & [0,1,3,4] & [0,2,3,4] & [1,2,3,4] \end{matrix} \\ \begin{matrix} [0,1,2] \\ [0,1,3] \\ [0,1,4] \\ [0,2,3] \\ [0,2,4] \\ [0,3,4] \\ [1,2,3] \\ [1,2,4] \\ [1,3,4] \\ [2,3,4] \end{matrix} & \left(\begin{array}{ccccc} -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ -1 & 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & -1 & 0 \\ 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{array} \right) \end{matrix}.$$

Destas matrizes dos operadores bordos, observamos que:

$$\text{Im}(\partial_1) \cong \mathbb{R}^4 \quad \text{e} \quad \ker(\partial_1) \cong \mathbb{R}^6,$$

$$\text{Im}(\partial_2) \cong \mathbb{R}^6 \quad \text{e} \quad \ker(\partial_2) \cong \mathbb{R}^4,$$

$$\text{Im}(\partial_3) \cong \mathbb{R}^4 \quad \text{e} \quad \ker(\partial_3) \cong \mathbb{R},$$

$$\text{Im}(\partial_4) \cong \mathbb{R} \quad \text{e} \quad \ker(\partial_4) \cong (0),$$

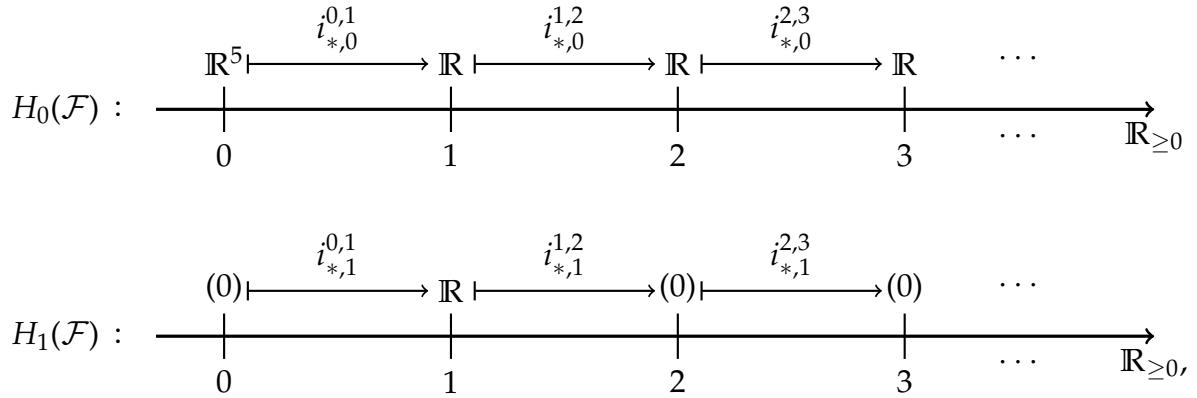
onde, por estarmos trabalhando com espaços vetoriais de dimensão finita, temos que a propriedade

$$\dim(V/U) = \dim(V) - \dim(U)$$

nos garante que

$$H_0(\Phi_\varepsilon) \cong \mathbb{R}, \quad \text{e} \quad H_p(\Phi_\varepsilon) = (0), \quad \forall p \geq 1$$

Do Exemplo 2.2.2, podemos visualizar os módulos de persistência, de dimensão 0 e 1 (que são os casos interessantes), pela seguinte representação:



onde os grupos de homologia estão indexados pelos pontos do segmento de reta $\mathbb{R}_{\geq 0}$, e as transformações lineares $H_p(\mathcal{F})(a \leq b)$, para $a, b \in \mathbb{R}$ e $p \in \mathbb{N}_0$, estão representadas pelas transformações lineares $i_{*,p}^{a,b}$.

Observação 2.2.3 (Interpretação dos módulos de persistência). Da homologia persistente que calculamos do Exemplo 2.2.2 notemos que: no instante de tempo $t = 0$, o complexo simplicial da filtração $\mathcal{F}$ possui 5 componentes conexas, permanecendo-se desta maneira até o instante de tempo $t = 1$, quando conectamos todos os vértices e formamos uma única componente conexa, mantendo-se desta maneira indefinidamente; já para a dimensão de homologia 1, apenas um ciclo aparece no instante de tempo $t = 1$, porém o mesmo desaparece logo em seguida, no instante de tempo $t = 2$.

Observação 2.2.4. Sempre quando estivermos trabalhando com a homologia persistente será aconselhável pensarmos nela como um mapeamento de instantes de tempos, em que propriedades topológicas surgem e desaparecem, de acordo com o conjunto que indexa a filtração em questão, como ilustrado no exemplo acima.

Com o objetivo de extrairmos estas informações dos módulos de persistência em análise, estudaremos o conceito de *Diagrama de Persistência*. Mas antes, precisaremos de alguns resultados auxiliares.

2.3 DIAGRAMAS DE PERSISTÊNCIA

Nesta seção apresentaremos o importante conceito dos *diagramas de persistência*, e seu variante, os *códigos de barra*. Estas estruturas vêm a complementar o visto em

homologia persistente, de tal forma a podermos classificar os diagramas de persistência como uma assinatura dos conjuntos em estudo.

Para construirmos os diagramas de persistência, iremos utilizar um resultado famoso da teoria das representações de quivers: o Teorema de Gabriel. Para mais informação sobre este teorema, e suas aplicações em homologia persistente, vide o Capítulo 1 e o Apêndice A da referência [26].

Iniciemos esta Seção com a definição de transformação natural entre funtores:

Definição 2.3.1. Sejam $F_1, F_2 : \mathcal{X} \rightarrow \mathcal{Y}$ dois funtores covariantes nas categorias $\mathcal{X}, \mathcal{Y}$. Então, dizemos que uma função $\tau : F_1 \rightarrow F_2$ é uma *transformação natural* entre os funtores F_1, F_2 se, para cada objeto $a, b \in \mathbf{Obj}(\mathcal{X})$, a função associar as setas

$$F_1(a) \xrightarrow{\tau_a} F_2(a) \in \text{hom}_{\mathcal{Y}}(F_1(a), F_2(a)), \quad F_1(b) \xrightarrow{\tau_b} F_2(b) \in \text{hom}_{\mathcal{Y}}(F_1(b), F_2(b))$$

tal que, para toda seta $a \xrightarrow{f} b \in \text{hom}_{\mathcal{X}}(a, b)$, o diagrama abaixo seja comutativo

$$\begin{array}{ccc} F_1(a) & \xrightarrow{\tau_a} & F_2(a) \\ F_1(f) \downarrow & & \downarrow F_2(f) \\ F_1(b) & \xrightarrow{\tau_b} & F_2(b). \end{array}$$

Além do mais, se cada função τ_a for bijetora, então τ é chamada de um *isomorfismo natural*.

Para exemplos sobre transformações naturais vide a Seção 4, Capítulo 15, do livro [20]. Desta definição, nos interessaremos apenas na classificação de isomorfismos dos módulos de persistência, que são funtores das categorias dos **Poset** (com ordem induzida pela reta real) nas categorias dos $\mathbb{K}$ -espaços vetoriais, denotados por $\mathbf{Vect}_{\mathbb{K}}$, onde $\mathbb{K}$ é um corpo.

Agora, seja $T \subseteq \mathbb{R}$ um subconjunto da reta real e $S \subseteq T$ um intervalo. Definiremos o funtor chamado *módulo intervalar*, fundamental para nosso estudo. Este funtor, denotado por

$$\mathbb{I}_S : (T, \leq) \rightarrow \mathbf{Vect}_{\mathbb{K}}$$

satisfaz a condição, para todo objeto da categoria $(T, \leq)$, de que

$$T \ni i \mapsto \begin{cases} \mathbb{K}, & \text{se } i \in S \\ (0), & \text{caso contrário} \end{cases}$$

e, para toda seta $i \leq j$,

$$i \leq j \mapsto \begin{cases} \mathbb{1} : \mathbb{K} \rightarrow \mathbb{K}, & \text{a transformação identidade, sempre quando } i, j \in S; \\ f, & \text{caso contrário. Onde, } f \text{ é uma transformação linear tal que} \\ & f(v) = 0, \forall v \in \mathbb{I}_S(i). \end{cases}$$

Agora somos capazes de enunciar o principal resultado de homologia persistente, o qual o dividimos em duas partes, (a) e (b), e cuja demonstração pode ser encontrada no Apêndice A de [26]:

Teorema 2.1 (Aplicação do Teorema de Gabriel, Parte (a)). *Seja $T \subseteq \mathbb{R}$ um subconjunto não nulo, $\mathcal{F}$ uma filtração e $H_*(\mathcal{F})$ a homologia persistente referente a filtração. Se tivermos que*

$$\text{card}(T) < \infty,$$

então teremos que para cada dimensão $p \in \mathbb{N}_0$, os módulos de persistência serão naturalmente isomorfos a soma direta de módulos intervalares. Em outras palavras, para cada dimensão $p \in \mathbb{N}_0$, teremos que existe um conjunto L tal que

$$H_p(\mathcal{F}) \cong \bigoplus_{i \in L} \mathbb{I}_{S_i},$$

com $S_i \subseteq T$ intervalos indexados pelo conjunto L . Além do mais, essa soma de módulos intervalares será única, a menos de uma permutação dos índices. (Aqui, o símbolo $\cong$ representa o isomorfismo natural entre funtores definido no início desta Seção.)

Teorema 2.2 (Aplicação do Teorema de Gabriel, Parte (b)). *Seja $T \subseteq \mathbb{R}$ um subconjunto não nulo, $\mathcal{F}$ uma filtração, $H_*(\mathcal{F})$ a homologia persistente referente a filtração $\mathcal{F}$ e $p \in \mathbb{N}_0$ um número natural. Se para cada elemento $i \in T$ tivermos que os espaços vetoriais, induzidos pelo módulo de persistência de dimensão p , forem de dimensão finita, i.e.,*

$$\dim(H_p(\mathcal{F})(i)) < \infty, \forall i \in T,$$

então, teremos que o módulo de persistência de dimensão p será naturalmente isomorfo a soma direta de módulos intervalares. Em outras palavras, o módulo de persistência de dimensão p pode ser escrito como

$$H_p(\mathcal{F}) \cong \bigoplus_{i \in L} \mathbb{I}_{S_i},$$

com $S_i \subseteq T$ intervalos indexados pelo conjunto L . Além do mais, essa soma de módulos intervalares será única, a menos de uma permutação dos índices. (Novamente, o símbolo $\cong$ representa o isomorfismo natural entre funtores definido no início desta Seção.)

Estes dois resultados nos dizem que a homologia persistente, satisfazendo certas hipóteses, pode ser interpretada como o registro da evolução temporal das propriedades topológicas de uma nuvem de dados. Destes registros, teremos sempre os instantes de *nascimento* e *morte* de uma propriedade de dimensão p .

Para clarear as ideias, relembremos o Exemplo 2.2.2, onde obtivemos que os módulos de persistência de dimensão 0 e 1, da filtração do grafo, podiam ser representadas por

$$\begin{array}{l}
 H_0(\mathcal{F}) : \quad \mathbb{R}^5 \xrightarrow{i_{*,0}^{0,1}} \mathbb{R} \xrightarrow{i_{*,0}^{1,2}} \mathbb{R} \xrightarrow{i_{*,0}^{2,3}} \mathbb{R} \cdots \\
 \quad \quad \quad \begin{array}{ccccccc}
 | & & | & & | & & | \\
 0 & & 1 & & 2 & & 3 \cdots
 \end{array}
 \end{array}$$

$$\begin{array}{l}
 H_1(\mathcal{F}) : \quad (0) \xrightarrow{i_{*,1}^{0,1}} \mathbb{R} \xrightarrow{i_{*,1}^{1,2}} (0) \xrightarrow{i_{*,1}^{2,3}} (0) \cdots \\
 \quad \quad \quad \begin{array}{ccccccc}
 | & & | & & | & & | \\
 0 & & 1 & & 2 & & 3 \cdots
 \end{array}
 \end{array}$$

$\mathbb{R}_{\geq 0}$

Pelo exposto acima, temos que

$$H_1(\mathcal{F}) \cong \mathbb{I}_{[1,2)}.$$

De fato, observe que no instante de tempo $t = 1$ ocorre que um ciclo de dimensão 1 surge na filtração, desaparecendo logo em seguida, no instante $t = 2$, quando cobrimos este ciclo (vide Figura 17, que apresenta a filtração do grafo). Para os módulos de persistência de dimensão zero não é difícil notar que

$$H_0(\mathcal{F}) \cong \mathbb{I}_{[0,\infty)} \oplus \mathbb{I}_{[0,1)} \oplus \mathbb{I}_{[0,1)} \oplus \mathbb{I}_{[0,1)} \oplus \mathbb{I}_{[0,1)}.$$

Novamente, o módulo de persistência de dimensão zero expressa exatamente o que nossa intuição sobre a filtração esperava. No início, temos cinco componentes conexas, o qual quatro somem no instante $t = 1$, permanecendo apenas uma componente conexa em todo restante de tempo.

Destes dois teoremas principais temos que para toda homologia persistente, referente a uma filtração $\mathcal{F}$, e satisfazendo as hipóteses destes teoremas, podemos associar um

diagrama, chamado de *diagrama de persistência*. Ou seja, fixada uma dimensão p , onde o módulo de persistência pode ser escrito como

$$H_p(\mathcal{F}) \cong \bigoplus_{i \in L} \mathbb{I}_{S_i},$$

temos que podemos associar a este módulo de persistência o seguinte multiset

$$\text{Dgm}(H_p(\mathcal{F})) = \left\{ (\inf(S_i), \sup(S_i)) \in \overline{\mathbb{R}}^2; i \in L \text{ e } H_p(\mathcal{F}) \cong \bigoplus_{i \in L} \mathbb{I}_{S_i} \right\},$$

com $\overline{\mathbb{R}}$ a reta estendida, $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$, que definiremos como o diagrama de persistência de dimensão p da homologia persistente, denotado por $\text{Dgm}(H_p(\mathcal{F}))$.

Novamente, consideremos a filtração do grafo do Exemplo 2.2.2, cujos módulos de persistência de dimensão 0 e 1 são

$$H_0(\mathcal{F}) \cong \mathbb{I}_{[0, \infty)} \bigoplus \mathbb{I}_{[0, 1)} \bigoplus \mathbb{I}_{[0, 1)} \bigoplus \mathbb{I}_{[0, 1)} \bigoplus \mathbb{I}_{[0, 1)}$$

e

$$H_1(\mathcal{F}) \cong \mathbb{I}_{[1, 2)}.$$

Para estes módulos de persistência, os diagramas de persistência serão os seguintes multisets

$$\text{Dgm}(H_0(\mathcal{F})) = \{(0, \infty), (0, 1), (0, 1), (0, 1), (0, 1)\} \quad \text{e} \quad \text{Dgm}(H_1(\mathcal{F})) = \{(1, 2)\},$$

que podem ser representados como o gráfico da Figura 18 abaixo.

Em geral estaremos interessados em estudar os diagramas de persistência de conjuntos finitos, o que nos garante que os espaços vetoriais induzidos pelos módulos de persistência sejam de dimensão finita e, portanto, pelo que vimos pela aplicação do Teorema de Gabriel, Parte (b), que nossas homologias de persistência admitirão diagramas de persistência, que por sua vez serão únicos.

Observação 2.3.1. Na maior parte das vezes, calcular os diagramas de persistência de uma filtração não será uma das tarefas mais fáceis. Porém, veremos mais a frente que existirão algoritmos que nos auxiliarão no cálculo dos diagramas de persistência destas estruturas, com a única exigência de que estabeleçamos certas restrições sob o corpo que age nos grupos de homologia induzidos da filtração em estudo.

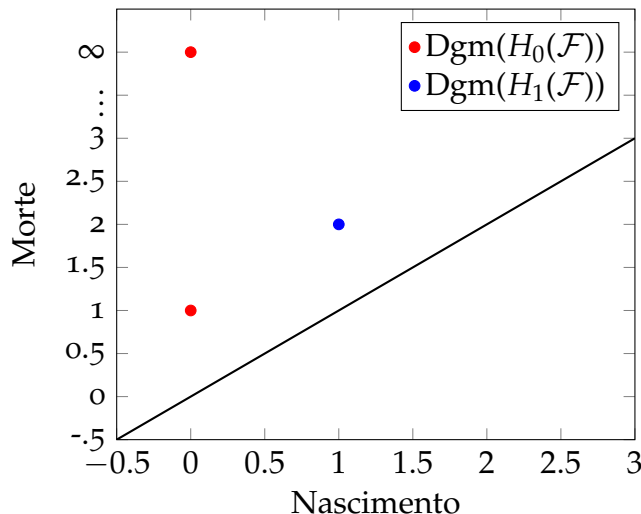


Figura 18: Diagrama de Persistência de dimensão 1 e 2 da nuvem de pontos do Exemplo 2.2.2

2.4 FILTRAÇÕES

Seja $\mathcal{X}$ um conjunto finito, o qual chamaremos de *nuvem de pontos de dados*. Abaixo apresentaremos duas das filtrações mais comuns na área de Topologia de Dados:

Definição 2.4.1 (*filtração simplicial de Vietoris-Rips*). Seja $\mathcal{X} \subseteq M$ uma nuvem de pontos de dados, com M um espaço métrico, e $T \subseteq \mathbb{R}_{\geq 0}$ um conjunto de números reais positivos. Então, a filtração simplicial

$$VR(\mathcal{X}) = (VR(\mathcal{X}, i))_{i \in T}$$

será chamada de *filtração simplicial de Vietoris-Rips* se, para cada índice $i \in T$, $VR(\mathcal{X}, i)$ for um complexo simplicial onde todo simplexo σ satisfaz

$$x, y \in \sigma \iff d(x, y) \leq i,$$

com $d(\cdot, \cdot)$ a métrica do espaço.

Definição 2.4.2 (*filtração simplicial de Čech*). Seja $\mathcal{X} \subseteq M$ uma nuvem de pontos de dados, com M um espaço métrico, e $T \subseteq \mathbb{R}_{\geq 0}$ um conjunto de números reais positivos. Então, a filtração simplicial

$$C(\mathcal{X}) = (C(\mathcal{X}, i))_{i \in T}$$

será chamada de *filtração simplicial de Čech* se, para cada índice $i \in T$, $C(\mathcal{X}, i)$ for um complexo simplicial tal que todo simplexo $\sigma = \{a_0, a_1, \dots, a_n\}$ satisfaz

$$\bigcap_{l=0}^n B[a_l, i] \neq \emptyset,$$

com $B[a_l, i]$ a bola fechada de raio i e centrada em a_l .

Para ilustrarmos estas duas filtrações, consideremos a nuvem de pontos abaixo

$$\mathcal{X} = \{a_0, a_1, a_2, a_3\}.$$

Na Figura 19(a) temos um exemplo de complexo simplicial da filtração de Vietoris-Rips, com parâmetro de filtração igual a 2. Já para a Figura 19(b), apresentamos um complexo simplicial da filtração de Čech, com o parâmetro de filtração igual a um. Note que, neste exemplo, os complexos da filtração de Čech são, de fato, subcomplexos dos complexos simpliciais da filtração de Vietoris-Rips. Isto se exprime mais geralmente, na seguinte proposição:

Proposição 2.4.1. *Seja $\mathcal{X} \subseteq \mathbb{R}^d$ uma nuvem de pontos e*

$$VR(\mathcal{X}) = (VR(\mathcal{X}, i))_{i \geq 0} \quad e \quad C(\mathcal{X}) = (C(\mathcal{X}, i))_{i \geq 0}$$

as respectivas filtrações de Vietoris-Rips e de Čech. Então, para todo $i \geq 0$, vale que

$$VR(\mathcal{X}, i) \subseteq C(\mathcal{X}, i) \subseteq VR(\mathcal{X}, 2i).$$

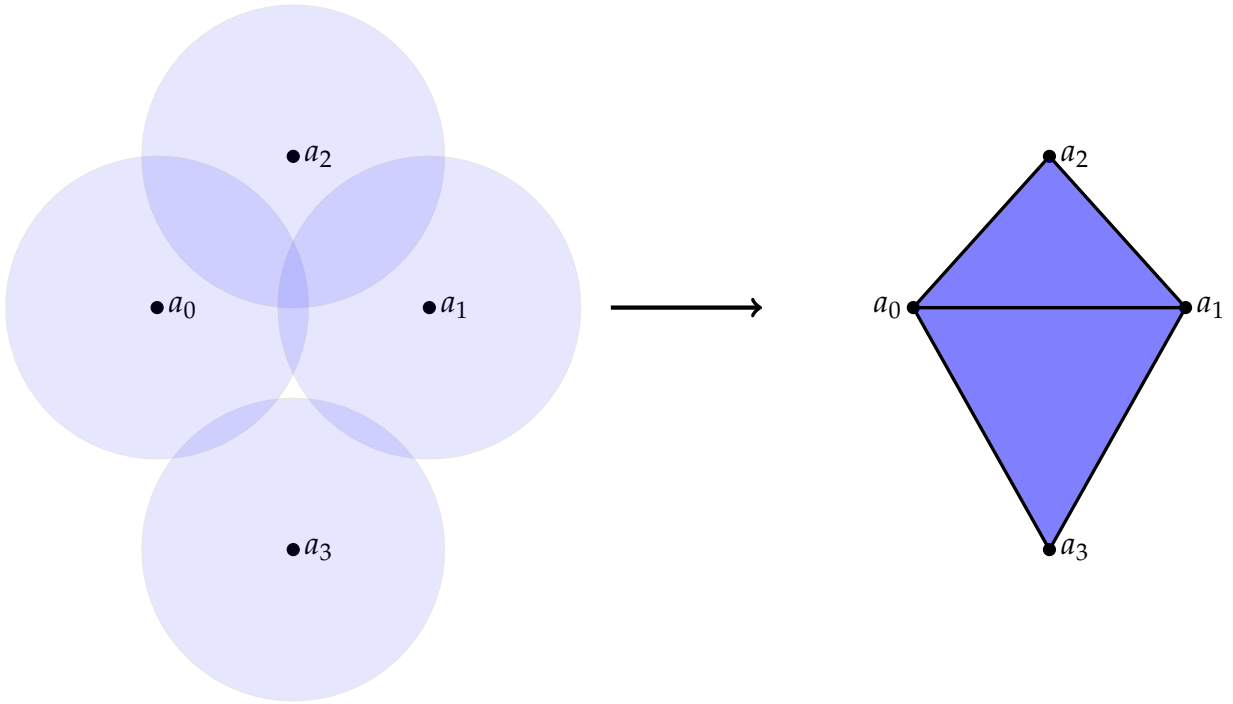
Demonstração. Vide o Capítulo 2 da referência [26]. □

2.5 ALGORITMO PARA CALCULAR O DIAGRAMA DE PERSISTÊNCIA EM $\mathbb{Z}_2$

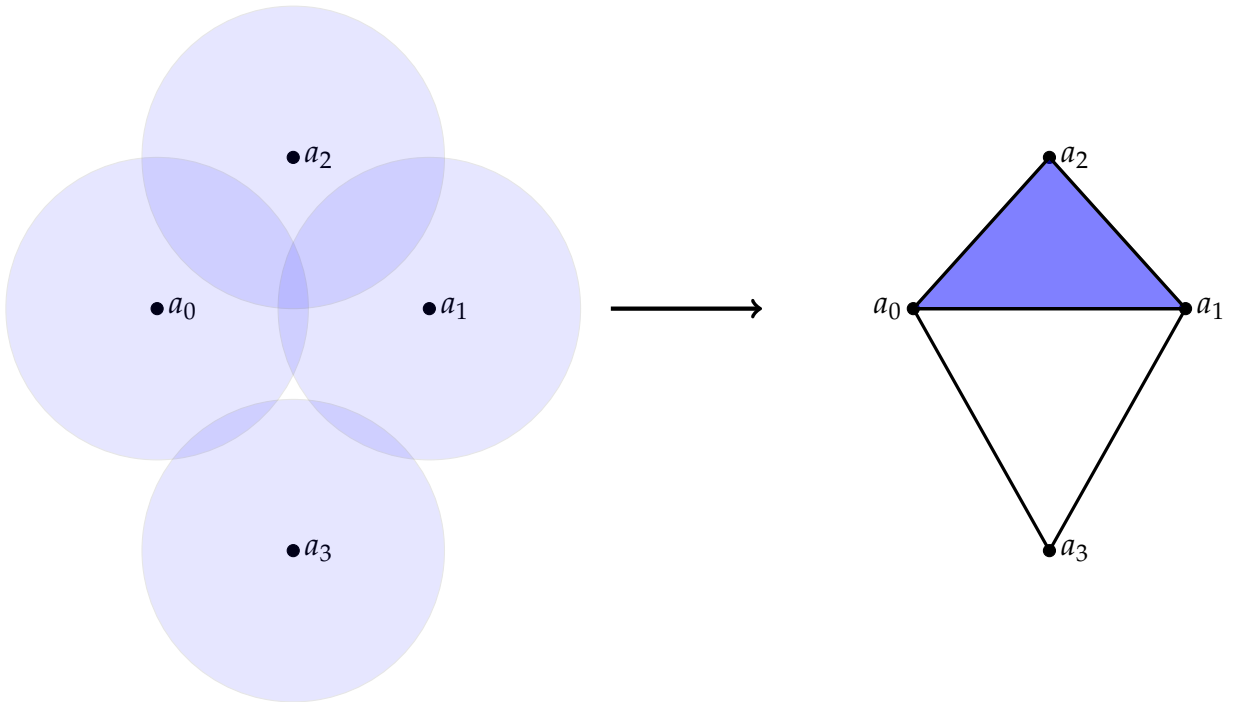
Nesta seção exibiremos um algoritmo que nos permitirá calcular, em $\mathbb{Z}_2$, os diagramas de persistência de uma nuvem de pontos finita. Este algoritmo nos impõe apenas duas restrições:

(A-1) Seja $\mathcal{X}$ uma nuvem de pontos finita e

$$\mathcal{F} = ((\mathcal{K}_i, \mathcal{S}_i))_{i \in I \subseteq \mathbb{N}_0}$$



(a) À esquerda a nuvem de pontos $\mathcal{X} = \{a_0, a_1, a_2, a_3\} \in \mathbb{R}^2$ e as bolas fechadas centradas em cada $a_i, i = 1, 2, 3, 4$, com raio unitário, e, à direita, o complexo simplicial $X_2 = (\mathcal{X}, \mathcal{S}_2)$ da filtração de Vietoris-Rips, com $\mathcal{S}_2 = \{\emptyset, \{a_0\}, \{a_1\}, \{a_2\}, \{a_3\}, \{a_0, a_1\}, \{a_1, a_2\}, \{a_0, a_2\}, \{a_0, a_3\}, \{a_1, a_3\}, \{a_0, a_1, a_2\}, \{a_0, a_1, a_3\}\}$.



(b) À esquerda a nuvem de pontos $\mathcal{X} = \{a_0, a_1, a_2, a_3\} \in \mathbb{R}^2$ e as bolas fechadas centradas em cada $a_i, i = 1, 2, 3, 4$, com raio unitário, e, à direita, o complexo simplicial $X_1 = (\mathcal{X}, \mathcal{S}_1)$ da filtração de Čech, com $\mathcal{S}_1 = \{\emptyset, \{a_0\}, \{a_1\}, \{a_2\}, \{a_3\}, \{a_0, a_1\}, \{a_1, a_2\}, \{a_0, a_2\}, \{a_0, a_3\}, \{a_1, a_3\}, \{a_0, a_1, a_2\}\}$.

Figura 19: Figura (a) fornece um exemplo de complexo simplicial para a filtração de Vietoris-Rips.

Enquanto que a Figura(b), da mesma maneira, apresenta um exemplo de complexo simplicial para uma filtração de Čech. Na figura (a) consideramos como parâmetro da filtração de Rips o raio igual a 2 enquanto que na figura (b) consideramos como parâmetro da filtração de Čech raio igual a 1.

uma filtração simplicial de $\mathcal{X}$, com $I = \{0, 1, 2, \dots, N\}$ um conjunto de índices finito e $\Phi_i = (\mathcal{K}_i, \mathcal{S}_i)$ uma sequência de complexos simpliciais, satisfazendo:

$$\Phi_0 \subseteq \Phi_1 \subseteq \Phi_2 \subseteq \dots \subseteq \Phi_N \quad \text{e} \quad \mathcal{S}_N = \{\sigma^1, \sigma^2, \dots, \sigma^N\},$$

e $\mathcal{S}_i$ uma sequência de conjuntos de simplexos encadeados

$$\mathcal{S}_0 = \{\emptyset\} \subseteq \mathcal{S}_1 = \mathcal{S}_0 \cup \{\sigma^1\} \subseteq \mathcal{S}_2 = \mathcal{S}_1 \cup \{\sigma^2\} \subseteq \dots \subseteq \mathcal{S}_N = \mathcal{S}_{N-1} \cup \{\sigma^N\}.$$

Desta sequência de simplexos, estabeleceremos a primeira restrição importante para o algoritmo a seguir: dado um simplexo qualquer de dimensão n , digamos σ^i , então para toda face deste simplexo, i.e., para todo simplexo $\tau \subseteq \sigma^i$, deveremos ter

$$\min\{j \in I; \tau \in \mathcal{S}_j\} \leq i,$$

ou seja, as faces devem sempre aparecer a frente dos simplexos as quais estão contidas.

(A-2) Finalmente, estabeleceremos que o corpo agindo nos grupos de homologia, dos módulos de persistência $H_p(\mathcal{F})$, deverá ser o $\mathbb{Z}_2$. Ou seja, para todo índice $i \in I$ da filtração, e para toda dimensão $p \in \mathbb{N}_0$, $H_p(i)$ deve ser um $\mathbb{Z}_2$ -espaço vetorial.

Abaixo, na Figura 20 (onde cada quadrado representa um complexo simplicial), segue um exemplo de filtração simplicial $\mathcal{F} = (\Phi_i)_{i=0}^7$ que satisfaz a condição (A-1) acima.

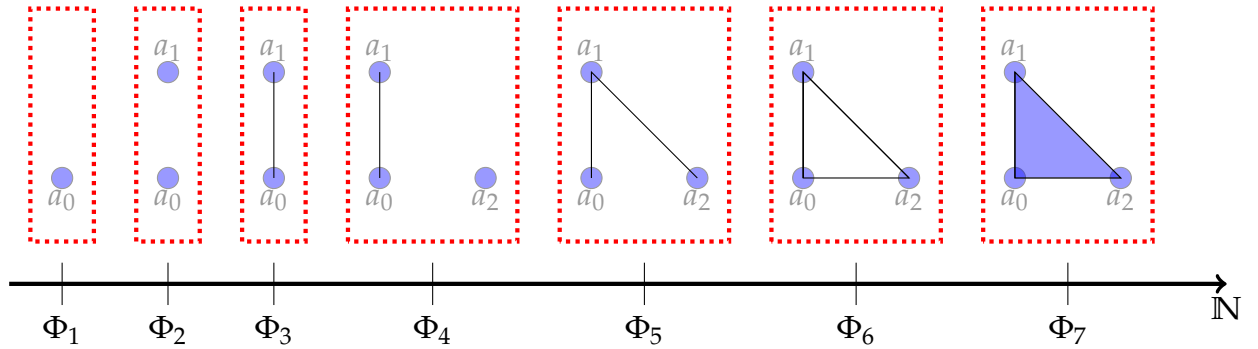


Figura 20: Exemplo de filtração simplicial $\mathcal{F} = (\Phi_i)_{i=0}^7$ que satisfaz a condição (A-1), do início desta seção. Note que as faces dos simplexos devem aparecer primeiro que o simplexo que as contém, se quisermos aplicar o algoritmo desta seção.

Antes de apresentarmos o algoritmo, precisaremos de um resultado importante sobre o número de Betti, decorrente de estarmos trabalhando com o corpo $\mathbb{Z}_2$ (propriedade A-2) e com uma filtração satisfazendo a propriedade (A-1).

Definição 2.5.1. Seja $\mathcal{F}$ uma filtração simplicial, construída a partir de um conjunto não nulo, e satisfazendo a propriedade (A-1). Então, dizemos que um simplexo $\sigma^i = \langle a_0, \dots, a_p \rangle$, que foi adicionado na filtração exatamente no instante de tempo i , é um *simplexo positivo* quando existir um ciclo w , de mesma dimensão que σ^i , tal que σ^i pertence a w . Caso contrário, dizemos que σ^i é um *simplexo negativo*.

Observação 2.5.1. Quando dizemos que um k -simplexo $\sigma = \langle a_0, \dots, a_k \rangle$, pertencente a um complexo simplicial Φ , está contido em um ciclo de dimensão k , queremos dizer que, referente ao módulo livre gerado por todos os simplexos de dimensão k , existe o ciclo

$$c = \sum_{i \in I} \sigma_i \in Z_k(\Phi),$$

com I finito, tal que σ é elemento da soma formal acima.

Aproveitaremos agora para apresentar um resultado essencial, que surge graças ao fato de estarmos trabalhando com o corpo $\mathbb{Z}_2$:

Proposição 2.5.1. *Seja $\mathcal{F} = (\Phi_i)_{i \in I \subseteq \mathbb{N}_0}$ uma filtração simplicial, tomada a partir de uma nuvem de pontos não nula, e satisfazendo a condição (A-1) (que garante que os complexos simpliciais, que formam a filtração, sejam inclusões sucessivas de novos simplexos, adicionados um por vez). Então, se considerarmos que o corpo agindo sob os grupos de homologia é o $\mathbb{Z}_2$, teremos como resultado que, para todo k -simplexo σ , que surge na filtração exatamente no instante de tempo i , σ terá o efeito de: ou aumentar o número de Betti $\beta_k(\Phi_i)$ por uma unidade, caso seja um simplexo positivo,*

$$\beta_k(\Phi_i) = \beta_k(\Phi_{i-1}) + 1,$$

ou, caso contrário, diminuir o número de Betti $\beta_{k-1}(\Phi_i)$ por uma unidade,

$$\beta_k(\Phi_i) = \beta_k(\Phi_{i-1}) - 1.$$

Em outras palavras, o simplexo σ , quando nasce na filtração, tem o efeito de criar um novo ciclo de mesma dimensão, se for um simplexo positivo, ou anular um ciclo de dimensão inferior, transformando-o em um bordo, caso contrário.

Demonstração. Vide a Seção 11.2 da referência [27]. □

Finalmente, abaixo apresentaremos a matriz com a qual o algoritmo calculará a homologia persistente. Seja

$$\mathcal{F} = (\Phi_i = (\mathcal{K}_i, \mathcal{S}_i))_{i \in I \subseteq \mathbb{N}_0}, \quad \text{com} \quad I = \{0, 1, 2, \dots, N\} \quad \text{e} \quad \mathcal{S}_N = \{\sigma^1, \sigma^2, \dots, \sigma^N\}$$

uma filtração simplicial respeitando a condição (A-1), denotaremos por

$$M^{\mathcal{F}} \in M_{N \times N}(\mathbb{Z}_2)$$

a matriz de dimensão $N \times N$ induzida pela filtração $\mathcal{F}$, e cujas entradas são elementos no corpo $\mathbb{Z}_2$, satisfazendo

$$\forall i, j \in \{1, 2, 3, \dots, N\} \quad \text{vale} \quad M_{i,j}^{\mathcal{F}} := \begin{cases} 1, & \text{se o simplexo } \sigma^i \text{ for uma face de } \sigma^j \\ 0, & \text{caso contrário.} \end{cases} \quad (4)$$

Observação 2.5.2. Note que por estarmos trabalhando com o corpo $\mathbb{Z}_2$, a matriz $M^{\mathcal{F}}$ nada mais é que a matriz de transformação dos operadores bordo. De fato, dado $\sigma^j = \langle a_0, \dots, a_m \rangle$, vale

$$\partial(\sigma^j) = \sum_l (-1)^l \langle a_0, \dots, \widehat{a_l}, \dots, a_m \rangle = \sum_l \langle a_0, \dots, \widehat{a_l}, \dots, a_m \rangle.$$

Finalmente, dada a matriz $M^{\mathcal{F}} \in M_{N \times N}(\mathbb{Z}_2)$, denotaremos por $M_l^{\mathcal{F}}$, com $l \in \{1, 2, \dots, N\}$, como sendo o vetor da l -ésima coluna da matriz $M^{\mathcal{F}}$:

$$M_l^{\mathcal{F}} = \begin{bmatrix} M_{1,l}^{\mathcal{F}} \\ M_{2,l}^{\mathcal{F}} \\ \vdots \\ M_{N,l}^{\mathcal{F}} \end{bmatrix}.$$

Agora, estamos prontos para apresentar o principal algoritmo deste capítulo. Seja $\mathcal{F} = (\Phi_i)_{i \in I \subseteq \mathbb{N}_0}$ uma filtração simplicial satisfazendo A-1, com I um conjunto de índices finito, e $H_*(\mathcal{F})$ a homologia persistente da filtração $\mathcal{F}$, restrita a condição de que seus grupos de homologia, para cada módulo de persistência, sejam $\mathbb{Z}_2$ -espaços vetoriais. Iniciemos apresentando o **Algoritmo 1**, que servirá de auxílio para o Algoritmo principal:

Algoritmo 1: Low (j, M)

Input: matriz $M \in M_{N \times N}(\mathbb{Z}_2)$ e um índice $j \in \{1, 2, \dots, N\}$

Output: retorna o maior índice $i \in \{1, 2, \dots, N\}$ tal que $M_{i,j} \neq 0$. Caso não exista tal índice, retorna 0

```

1  $l \leftarrow 0$ ;
2 if  $\{i \in \{1, 2, 3, \dots, N\}; M_{i,j} \neq 0\} \neq \emptyset$  then
3    $l \leftarrow \max\{i \in \{1, 2, 3, \dots, N\}; M_{i,j} \neq 0\}$ ;
4 return  $l$ ;
```

Este algoritmo auxiliar serve apenas para expressarmos os elementos pivôs do processo de escalonamento, utilizado no algoritmo principal:

Algoritmo 2: Calculando a homologia persistente de uma filtração $\mathcal{F}$.

Input: Matriz $M^{\mathcal{F}} \in M_{N \times N}(\mathbb{Z}_2)$ referente a uma filtração $\mathcal{F}$ satisfazendo a condição A-1, e cujas entradas são elementos de $\mathbb{Z}_2$ e definidas em 4.

Output: Uma matriz $M \in M_{N \times N}(\mathbb{Z}_2)$ cujas colunas representam os intervalos de persistência

```

1 for  $j \leftarrow 1$  to  $N$  do
2    $l_j \leftarrow \text{Low}(j, M^{\mathcal{F}})$ ;
3    $k \leftarrow 1$ ;
4   while  $(k < j) \wedge (l_j \neq 0)$  do
5      $l_k \leftarrow \text{Low}(k, M^{\mathcal{F}})$ ;
6     if  $l_k = l_j$  then
7        $M_j^{\mathcal{F}} \leftarrow M_j^{\mathcal{F}} + M_k^{\mathcal{F}} \pmod{2}$ ;  $\triangleright$  Atualiza a  $j$ -ésima coluna de  $M^{\mathcal{F}}$ 
8        $l_j \leftarrow \text{Low}(j, M^{\mathcal{F}})$ ;
9        $k \leftarrow 1$ ;
10    else
11       $k \leftarrow k + 1$ ;
12 return  $M^{\mathcal{F}}$ 

```

Notemos que

- i O **Algoritmo 2** serve para escalonarmos a matriz dos operadores bordo $M^{\mathcal{F}}$, na direção crescente dos índices de suas colunas. Consequentemente, colunas distintas não poderão ter elementos pivôs na mesma linha, i.e.,

$$j \neq j' \implies \text{Low}(j, M) \neq \text{Low}(j', M);$$

- ii Ao encerrarmos o escalonamento, a j -ésima coluna, da matriz resultante, será a combinação linear das colunas anteriores a j , que, por definição, são os operadores bordo dos simplexes σ^i , com $i < j$. De fato, denotando por C_j a j -ésima coluna, temos que

$$C_j = \partial(\sigma_j) + \sum_{i \in I_j} \partial(\sigma^i), \quad (5)$$

com $I_j \subseteq \{1, \dots, j-1\}$;

iii De maneira análoga, caso a j -ésima coluna, obtida pelo processo de escalonamento, for não nula, poderemos escrever C_j como combinação linear dos simplexes associados a entradas não nulas de C_j :

$$C_j = \sigma^{\text{Low}(j, M^{\mathcal{F}})} + \sum_{i \in I_{\text{Low}}(j, M^{\mathcal{F}})} \sigma^i, \quad (6)$$

com $I_{\text{Low}}(j, M^{\mathcal{F}}) \subseteq \{1, \dots, \text{Low}(j, M^{\mathcal{F}}) - 1\}$.

O resultado gerado pelo **Algoritmo 2** tem relevância, graças a seguinte proposição:

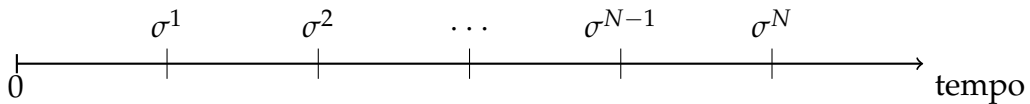
Proposição 2.5.2. *Seja*

$$\mathcal{F} = (\Phi_i = (\mathcal{K}_i, \mathcal{S}_i))_{i=0}^N, \quad \text{com} \quad \mathcal{S}_N = \{\sigma^1, \sigma^2, \dots, \sigma^N\}$$

uma filtração satisfazendo (A-1), com $M^{\mathcal{F}}$ a matriz induzida por $\mathcal{F}$ e cujas entradas são elementos de $\mathbb{Z}_2$ satisfazendo 4. Então, aplicando o **Algoritmo 2** a matriz $M^{\mathcal{F}}$, obtemos uma matriz M^{OUT} que satisfaz as seguintes propriedades:

- (i) toda coluna não nula de M^{OUT} estará associada a elementos do diagrama de persistência da forma $(\text{Low}(j, M^{\text{OUT}}), j)$, com j o índice da coluna, e vice-versa;
- (ii) toda coluna j de M^{OUT} , cujas entradas sejam zero, e que seja tal que não exista uma coluna não nula j' satisfazendo $\text{Low}(j', M^{\text{OUT}}) = j$, estará associada a um simplexo positivo j , responsável pela existência do ponto (j, ∞) no diagrama de persistência da filtração, e vice-versa.

Demonstração. Provemos a propriedade (i) primeiro. ($\implies$) Recordemos que a filtração $\mathcal{F}$ pode ser vista como a inclusão sucessiva de simplexes, tal qual o exposto na imagem abaixo, respeitando a restrição de que as faces dos simplexes devem, sempre, aparecer primeiro na filtração.



Para provarmos a primeira propriedade, recordemos das Equações 5 e 6, que se a j -ésima coluna, C_j , resultante do processo de escalonamento for não nula, então

$$\sigma^{\text{Low}(j, M^{\text{OUT}})} + \sum_{i \in I_{\text{Low}}(j, M^{\text{OUT}})} \sigma^i = C_j = \partial \left(\sigma^j + \sum_{i \in I_j} \sigma^i \right) \quad (7)$$

e, portanto,

$$\partial \left(\sigma^{\text{Low}(j, M^{\text{OUT}})} + \sum_{i \in I_{\text{Low}}(j, M^{\text{OUT}})} \sigma^i \right) = 0.$$

Concluí-se, desta igualdade, que $\sigma^{\text{Low}(j, M^{\text{OUT}})}$ é um simplexo positivo e, pela Proposição 2.5.1, temos garantido que este simplexo aumentará em uma unidade o número de Betti, referente ao grupo de homologia calculado a partir do complexo simplicial $\Phi_{\text{Low}(j, M^{\text{OUT}})}$. Observe, também, que a igualdade 7 fornece outra informação importante,

$$\sigma^{\text{Low}(j, M^{\text{OUT}})} + \sum_{i \in I_{\text{Low}}(j, M^{\text{OUT}})} \sigma^i = \partial \left(\sigma^j + \sum_{i \in I_j} \sigma^i \right), \quad (8)$$

de que no instante j o ciclo, onde o simplexo $\sigma^{\text{Low}(j, M^{\text{OUT}})}$ é face, torna-se um bordo. Além do mais este é o primeiro instante que isto acontece. De fato, como a matriz escalonada admite apenas um pivô por linha, então, se tivéssemos que o ciclo fosse um bordo, antes do instante j , teríamos a condição de unicidade de pivôs por linha quebrada. Logo, temos garantido o tempo de nascimento e morte de uma propriedade topológica pelo ponto $(\text{Low}(j, M^{\text{OUT}}), j)$, pertencente ao diagrama de persistência de mesma dimensão do simplexo positivo $\sigma^{\text{Low}(j, M^{\text{OUT}})}$. ($\Leftarrow$) Seja (j, j') um ponto do diagrama de persistência, então sabemos que

$$\partial(\sigma^{j'}) = \sigma^j + \sum_{i \in I_{j'} \setminus \{j\}} \sigma^i$$

e que a j' -ésima coluna da matriz M^{OUT} é não-nula, pois, se o fosse, $\sigma^{j'}$ seria um simplexo positivo, pois pela igualdade 7, vale que

$$0 = C_{j'} = \partial \left(\sigma^{j'} + \sum_{i \in I_{j'}} \sigma^i \right). \quad (9)$$

Suponhamos, por absurdo, que $\text{Low}(j', M^{\text{OUT}}) \neq j$. Portanto, pelo escalonamento temos que, para $I_{\text{Low}}(j, M^{\text{OUT}}) \subseteq \{1, \dots, N\} \setminus \{j\}$, vale a igualdade

$$\partial(\sigma^{j'}) = \sum_{i \in I_{j'}} \partial(\sigma^i) + \sum_{i \in I_{\text{Low}}(j', M^{\text{OUT}})} \sigma^i.$$

Juntando esta igualdade com a anterior, obtemos que

$$\sigma^j = \sum_{i \in I_{j'} \setminus \{j\}} \sigma^i + \sum_{i \in I_{j'}} \partial(\sigma^i) + \sum_{i \in I_{\text{Low}}(j', M^{\text{OUT}})} \sigma^i.$$

Todavia, a igualdade acima implica que

$$\sigma^j \in \sum_{i \in I_{j'}} \partial(\sigma^i) \implies \sigma^j \in \sum_{i \in I_{\text{Low}(j', M^{\text{OUT}})}} \sigma^i,$$

que é um absurdo. Desta forma, concluímos que $\text{Low}(j', M^{\text{OUT}}) = j$.

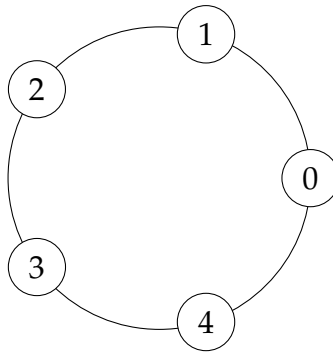
Encerremos provando a propriedade (ii), que é a propriedade mais fácil de se averiguar. ($\implies$) Sempre quando a coluna C_j for toda nula, então vale pela igualdade 7 que

$$0 = C_j = \partial \left(\sigma^j + \sum_{i \in I_j} \sigma^i \right) \quad (10)$$

e, portanto, σ^j é um simplexo positivo. Além do mais, o ciclo o qual este simplexo faz parte não se transformará em bordo, pois caso acontecesse, então existiria j' tal que $\text{Low}(j', M^{\text{OUT}}) = j$, como já provado em (i). Todavia, por hipótese sabemos que esta situação não pode acontecer. Com isso, provamos que a propriedade topológica que nasce no instante j jamais desaparece e, portanto, (j, ∞) é um elemento do diagrama de persistência. ($\Leftarrow$) Seja (j, ∞) um ponto do diagrama de persistência. Então, temos que a j -ésima coluna de M^{OUT} deve ter toda a sua entrada nula pois, caso contrário, teríamos pelo item (i) que o simplexo seria negativo, o que é um absurdo. $\square$

Observação 2.5.3. Com os intervalos de persistência obtidos pela Proposição 2.5.2 podemos escrever os módulos de persistência, referentes a homologia persistente da filtração em estudo, como isomorfa a soma direta dos módulos intervalares $\mathbb{I}_{[a,b)}$, com $[a, b)$ o intervalo obtido pela Proposição 2.5.2.

Exemplo 2.5.1. Consideremos o Exemplo 2.2.2, referente ao grafo de vértices $V =$



$\{0, 1, 2, 3, 4\}$. Lembremos, da filtração original deste exemplo, que o grafo admitia a triangularização $\Phi = (\mathcal{K}, \mathcal{S})$, com $\mathcal{K} = V$ e

$$\begin{aligned} \mathcal{S} = \{ & \emptyset, \{0\}, \{1\}, \{2\}, \{3\}, \{4\}, \{0, 1\}, \{1, 2\}, \{2, 3\}, \{3, 4\}, \{0, 4\}, \\ & \{0, 2\}, \{0, 3\}, \{1, 3\}, \{1, 4\}, \{2, 4\}, \\ & \{0, 1, 2\}, \{0, 1, 3\}, \{0, 1, 4\}, \{0, 2, 3\}, \{0, 2, 4\}, \{0, 3, 4\}, \{1, 2, 3\}, \\ & \{1, 2, 4\}, \{1, 3, 4\}, \{2, 3, 4\}, \\ & \{0, 1, 2, 3\}, \{0, 1, 2, 4\}, \{0, 1, 3, 4\}, \{0, 2, 3, 4\}, \{1, 2, 3, 4\}, \\ & \{0, 1, 2, 3, 4\} \}, \end{aligned}$$

Agora, estabelecendo a ordenação

$$\begin{array}{lllll} \sigma^1 = \{0\} & \sigma^2 = \{1\} & \sigma^3 = \{2\} & \sigma^4 = \{3\} & \sigma^5 = \{4\} \\ \sigma^6 = \{0, 1\} & \sigma^7 = \{1, 2\} & \sigma^8 = \{2, 3\} & \sigma^9 = \{3, 4\} & \sigma^{10} = \{0, 4\} \\ \sigma^{11} = \{0, 2\} & \sigma^{12} = \{0, 3\} & \sigma^{13} = \{1, 3\} & \sigma^{14} = \{1, 4\} & \sigma^{15} = \{2, 4\} \\ \sigma^{16} = \{0, 1, 2\} & \sigma^{17} = \{0, 1, 3\} & \sigma^{18} = \{0, 1, 4\} & \sigma^{19} = \{0, 2, 3\} & \sigma^{20} = \{0, 2, 4\} \\ \sigma^{21} = \{0, 3, 4\} & \sigma^{22} = \{1, 2, 3\} & \sigma^{23} = \{1, 2, 4\} & \sigma^{24} = \{1, 3, 4\} & \sigma^{25} = \{2, 3, 4\} \\ \sigma^{26} = \{0, 1, 2, 3\} & \sigma^{27} = \{0, 1, 2, 4\} & \sigma^{28} = \{0, 1, 3, 4\} & \sigma^{29} = \{0, 2, 3, 4\} & \sigma^{30} = \{1, 2, 3, 4\} \\ \sigma^{31} = \{0, 1, 2, 3, 4\} & & & & \end{array}$$

conseguimos definir uma filtração que satisfaz a condição (A-1). De fato, basta apenas considerarmos a filtração:

$$\mathcal{F} = (\Phi_i = (\mathcal{K}_i, \mathcal{S}_i))_{i=0}^{31}$$

com, $\mathcal{K}_i = V$, para todo $i \in \{0, 1, 2, \dots, 31\}$, tal que

$$\Phi_0 = (V, \{\emptyset\}) \quad \text{e, dado } i \in \{0, 1, \dots, 31\}, \quad \text{temos} \quad \mathcal{S}_{i+1} = \mathcal{S}_i \cup \sigma^i.$$

Desta filtração podemos induzir a matriz $M^{\mathcal{F}} \in M_{31 \times 31}(\mathbb{Z}_2)$, com entradas dadas por 4. De forma a manter a leitura da matriz $M^{\mathcal{F}}$ mais fácil, escreveremos apenas as entradas

o simplexo σ^1 será um simplex positivo. Como o inteiro não está relacionado nas outras colunas não nulas, temos que σ^1 será simplex positivo durante toda a filtração. O que de fato corresponde ao esperado, já que sabemos que em nossa filtração uma componente conexa se mantém o tempo todo. Da mesma forma, temos que a segunda coluna é, também vazia. Porém, diferentemente da primeira coluna, o algoritmo nos informa que o simplexo σ^2 aparece no instante 2 e desaparece no instante 6. Portanto, para o simplexo σ^2 não teremos seu tempo de vida dado por $[2, \infty)$, mas sim por $[2, 6)$. O mesmo vale para os outros.

Para visualizarmos melhor o resultado do Exemplo 2.5.1, exibimos abaixo, na Figura 21, seu diagrama de persistência. Note que pontos do diagrama de persistência afastados da diagonal representam características importantes. Como era esperado, temos uma característica que se destaca, o qual nos informa que uma componente conexa dura toda a filtração do grafo. Outro ponto que se destaca é o ponto $(10, 21)$, o qual nos informa que uma característica de dimensão 1, um buraco, aparece em nossa filtração e dura um tempo razoável nela. Mesmo o diagrama de persistência sendo de uma filtração distinta da original, do Exemplo 2.2.2, temos que os pontos mais distantes da diagonal nos trazem informação importante referente a filtração original.

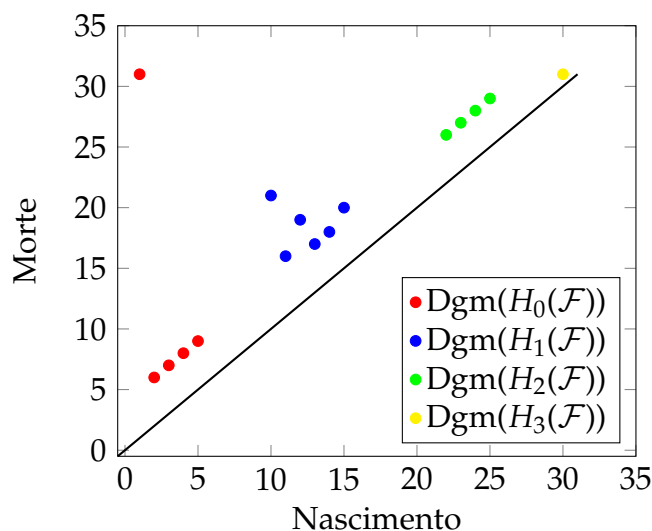


Figura 21: Diagrama de persistência da filtração do Exemplo 2.5.1

2.6 UTILIZANDO O AMBIENTE PYTHON

Nesta seção iremos ilustrar como o ambiente Python pode ser usado para o cálculo dos diagramas de persistência, das distâncias entre os diagramas, para o cálculo de algumas filtrações etc. As ferramentas computacionais necessárias para isto serão:

- Python⁴ versão mais nova (atualmente é a versão 3). Se estiver trabalhando em um ambiente Linux recente, como o Ubuntu, por exemplo, então o Python já estará instalado na versão mais recente;
- Biblioteca Numpy⁵, para cálculo numérico de alta-performance no ambiente Python. Para instalar esta biblioteca no Ubuntu basta digitar o seguinte comando no terminal: **sudo apt-get install python3-numpy**.
- Biblioteca Gudhi⁶, para a análise topológica de dados no ambiente Python. Para instalar esta biblioteca no Ubuntu basta digitar o seguinte comando no terminal: **sudo apt-get install python3-gudhi**.
- Biblioteca Matplotlib⁷, para gerarmos os gráficos no ambiente Python. Para instalar esta biblioteca no Ubuntu basta digitar o seguinte comando no terminal: **sudo apt-get install python3-matplotlib**.

A biblioteca importante para nós é a biblioteca Gudhi, o qual estaremos apenas interessados na parte de homologia persistente. As principais funções que iremos usar serão explicadas e, em seguida, iremos dar exemplos de códigos na prática. Para uma lista completa das funções recomenda-se a leitura do manual, referente a implementação para Python, que pode ser vista no sítio: <https://gudhi.inria.fr/python/latest/>.

Começemos listando as filtrações de Rips e $\check{c}$ ech. Um ponto importante é que a filtração de $\check{c}$ ech não é implementada na biblioteca, devido ao alto custo computacional. Entretanto, uma outra filtração é implementada: a α -filtração. Esta filtração tem duas vantagens: pode ser implementada computacionalmente de forma eficiente; e, segundo, esta filtração tem o mesmo tipo de homotopia da filtração de $\check{c}$ ech (vide [27, Capítulo

4 <https://www.python.org/>

5 <https://numpy.org/>

6 <http://gudhi.gforge.inria.fr/>

7 <https://matplotlib.org/>

6)). Portanto, se quiser implementar computacionalmente uma filtração de Čech, recomenda-se usar a α -filtração.

Note que esta biblioteca utiliza o conceito de orientação a objetos, por isto estaremos sempre fazendo referência a classes e seus métodos.

Dado um conjunto de dados armazenados em uma lista, chamemos este vetor de x . Para criarmos as filtrações de Čech ou Rips deveremos proceder da seguinte maneira:

FILTRAÇÃO DE ČECH : devemos criar uma instância de um objeto da classe **AlphaComplex**. Em seguida, devemos inicializar a estrutura de dados **simplex_tree**, o qual será responsável por lidar com os complexos simpliciais assim como realizar todos os cálculos.

FILTRAÇÃO DE RIPS : devemos criar uma instância de um objeto da classe **RipsComplex**. Em seguida, devemos inicializar a estrutura de dados **simplex_tree**.

Caso queira criar uma filtração manualmente é possível através do método **insert** da classe **SimplexTree**.

Na classe **SimplexTree** temos o método **persistence** para calcular o diagrama de persistência da filtração. Agora, caso seja do interesse calcular a distância Bottleneck entre diagramas de persistência, devemos utilizar a função **bottleneck**.

Abaixo exibiremos um exemplo de como implementar a biblioteca Gudhi no Python.

Exemplo 2.6.1. Seja X uma variável aleatória com distribuição uniforme no intervalo $[0, 2\pi)$ e Y uma variável aleatória com distribuição normal $\mathcal{N}(0, 0.2^2)$. Simulemos as seguintes variáveis aleatórias:

$$X_i \text{ i.i.d a variável } X \text{ e } Y_i \text{ i.i.d a variável } Y, \text{ com } i = 1, 2, \dots, 100.$$

Destas variáveis iremos estabelecer nossa nuvem de dados

$$\mathcal{X} = \{Z_1, Z_2, \dots, Z_{100}\}, \text{ onde } Z_i = ((1 + Y_i) \cos X_i, (1 + Y_i) \sin X_i), \forall i = 1, 2, \dots, 100.$$

Estabelecido a nuvem de pontos de dados, vamos calcular sua filtração de Čech e seu diagrama de persistência com o auxílio do Python. Abaixo, segue o script⁸

```
# Importando as bibliotecas
import numpy as np
import gudhi
```

⁸ Este script pode ser encontrado aqui: <https://github.com/rafaelpcarneiro>.

```
from matplotlib import pyplot as plt

# Constantes
SEMENTE          = 12345
TAM_AMOSTRA      = 1000
DESVIO_PADRAO    = 0.2

# Definindo a semente do algoritmo para gerarmos os numeros aleatorios.
np.random.seed( SEMENTE )

# Variaveis aleatorias iid a X e Y
X_i = np.random.uniform( 0, 2 * np.pi,      (TAM_AMOSTRA, 1))
Y_i = np.random.normal ( 0, DESVIO_PADRAO, (TAM_AMOSTRA, 1))

# Calculando a amostra referente ao eixo x e eixo y
amostra_x = (1 + Y_i) * np.cos( X_i )
amostra_y = (1 + Y_i) * np.sin( X_i )

# Agora vamos criar a nuvem de pontos de dados:
nuvem_pontos = np.concatenate( (amostra_x, amostra_y), axis=1)

# Caso queira gravar a nuvem de pontos em um arquivo csv:
np.savetxt( 'nuvem_pontos.csv', nuvem_pontos,
            delimiter = ',', fmt='%.3f')

# gerando o grafico de nossa nuvem de pontos de dados
plt.scatter( amostra_x, amostra_y )
plt.savefig( 'nuvem_pontos.pdf')

# para calcular a filtracao desejada devemos converter a
# nuvem de pontos de dados em uma lista
nuvem_pontos_lista = nuvem_pontos.tolist()

##### Agora calculamos a filtracao
filtracao = gudhi.AlphaComplex( nuvem_pontos_lista )

# E criamos a estrutura de dados principal: SimplexTree.
# Lembre que aqui filtracao eh uma instancia da classe AlphaComplex.
st = filtracao.create_simplex_tree()

# Agora podemos calcular o diagrama de persistencia
```

```

diagrama_persistencia = st.persistence( homology_coeff_field = 2 )

# E gerar seu grafico
gudhi.plot_persistence_diagram( diagrama_persistencia , legend = True)
plt.savefig( 'diagrama_persistencia.pdf' )

```

Criando um arquivo chamado "exemplo.py" com o conteúdo acima e rodando no terminal o comando

```
$ python3 exemplo.py
```

gerara um arquivo pdf contendo o gráfico da nuvem de pontos, Figura 22(a), e um arquivo pdf com o gráfico do diagrama de persistência, Figura 22(b). Percebe-se, pelo diagrama de persistência que duas características, de dimensão zero e um, se destacam, que são justamente as características referente a componente conexa e ao buraco, que condizem com a forma geométrica de nossa nuvem de pontos. Note, no diagrama de persistência, que as características próximas da diagonal não representam informação relevante para a análise, importando apenas as características mais afastadas da diagonal.

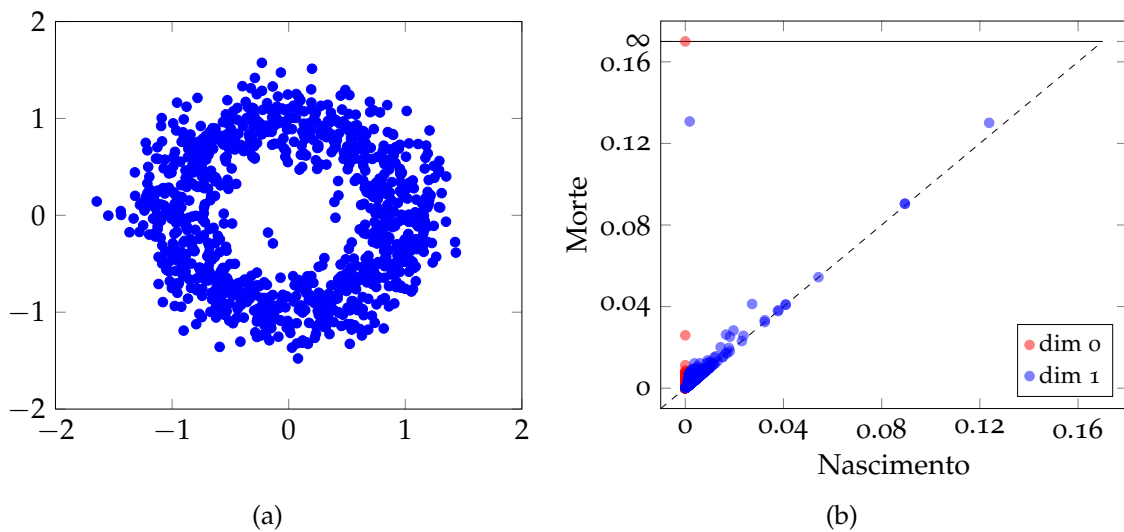


Figura 22: Gráfico da nuvem de pontos do Exemplo 2.6.1 e seu diagrama de persistência.

3 | HOMOLOGIA PERSISTENTE DE PROCESSOS PONTUAIS

Neste capítulo demonstraremos a Lei Forte dos Grandes Números e o Teorema do Limite Central para a homologia persistente¹ associada às classes de processos pontuais. Em outras palavras, estaremos interessados em estudar o limite da sequência

$$\frac{\beta_q^{r,s}(\mathbb{K}(\Phi \cap \Lambda_L^d))}{L^d},$$

com $0 \leq q \leq \dim(\mathbb{R}^d)$ e $d \in \mathbb{N}$ a dimensão do espaço vetorial, onde

(i) Φ é um processo pontual em $\mathbb{R}^d$;

(ii) Λ_L^d é a janela finita

$$\Lambda_L^d = [-L/2, L/2]^d \subseteq \mathbb{R}^d;$$

(iii) $\mathbb{K}(\Phi \cap \Lambda_L^d)$ é uma filtração simplicial do conjunto de pontos $\Phi \cap \Lambda_L^d$;

(iv) $\beta_q^{r,s}(\mathbb{K}(\Phi \cap \Lambda_L^d))$ é o número de persistência de Betti de dimensão q da filtração $\mathbb{K}(\Phi \cap \Lambda_L^d)$, levando em consideração o intervalo de tempo $[r, s] \subseteq \mathbb{R}_{\geq 0}$.

Sabemos dos artigos de Iraoka, Shirait e Trinh, vide [1], e do artigo de Yogeshwaran, Subag e Adler, vide [2], que para processos pontuais Φ em $\mathbb{R}^d$, estacionários, ergódicos e com todos os momentos finitos, então, para todo $q \in \mathbb{N}$ e para todo $r < s \in \mathbb{R}_{\geq 0}$, existirão constantes $\hat{\beta}_q^{r,s}$ tais que, com probabilidade 1, a convergência abaixo será verdadeira

$$\lim_{L \rightarrow \infty} \frac{\beta_q^{r,s}(\mathbb{K}(\zeta \cap \Lambda_L^d))}{L^d} = \hat{\beta}_q^{r,s}, \quad (11)$$

¹ Quando dizemos *propriedades de homologia persistente* (ou, *propriedades persistentes*) estamos, de forma grosseira, nos referindo ao número de persistência de Betti. Vale salientar que o número de Betti será sempre relativo a um processo pontual restrito à janela $\Lambda_L^d = [-L/2, L/2]^d \subseteq \mathbb{R}^d$. Restringimos o processo pontual a estas janelas pois, por processos pontuais serem medidas de Radon, temos garantido que a quantidade de pontos nestas janelas serão finitos e, portanto, vide [26, capítulo 1], teremos garantido a existência dos diagramas de persistência associados.

com $\Lambda_L^d = [-L/2, L/2]^d$ um intervalo e $\beta_q^{r,s}(\mathbb{K}(\phi \cap \Lambda_L^d))$ o número de persistência de Betti de dimensão q , referente a filtração $\mathbb{K}(\phi \cap \Lambda_L^d)$ e o intervalo de tempo $[r, s]$.

Neste capítulo provaremos que os processos pontuais dominados por um processo pontual de Poisson homogêneo, $\mathcal{P}_\tau$, também satisfazem a propriedade de convergência acima.

Finalmente, generalizamos o Teorema do Limite Central para o número de persistência de Betti associados à processos pontuais de Poisson homogêneos apresentado em [2] para uma classe de processos pontuais gibbsianos, que admitem esquema de simulação perfeita, considerando o número de persistência de Betti para características topológicas M -limitadas,

$$\frac{\beta_q^{M,r,s}(\mathbb{K}(\zeta \cap \Lambda_L^d)) - \mathbb{E} [\beta_q^{r,s}(\mathbb{K}(\zeta \cap \Lambda_L^d))]}{L^{d/2}} \xrightarrow{L \rightarrow \infty} \mathcal{N}(0, \hat{\sigma}_{r,s}^2), \quad (12)$$

onde a convergência acima é a convergência de distribuição e $\mathcal{N}(0, \hat{\sigma}^2)$ é a distribuição normal.

3.1 NOTAÇÃO BÁSICA

Seja $\mathcal{B}(\mathbb{R}^d)$ a σ -álgebra dos borelianos, reservaremos a letra ℓ para descrevermos a medida de Lebesgue neste espaço

$$\ell : (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \rightarrow [0, \infty].$$

Além disso, utilizaremos a notação $|A| = \ell(A)$, para representarmos o volume do conjunto $A \in \mathcal{B}(\mathbb{R}^d)$.

Dadas duas medidas, μ e ν , definidas no mesmo espaço de medida $(\mathbb{X}, \mathcal{A})$. Escreveremos $\mu \ll \nu$ sempre que a medida μ for absolutamente contínua com relação a ν , ou seja, sempre que existir uma função $\mathcal{A}$ -mensurável $f : (\mathbb{X}, \mathcal{A}) \rightarrow \overline{\mathbb{R}}$ tal que

$$\mu(A) = \int_A d\mu = \int_A f d\nu,$$

para todo conjunto $\mathcal{A}$ -mensurável A . Alternativamente, diremos que a função f , acima, é a derivada de Radon-Nykodim de μ relativamente à ν :

$$\frac{d\mu}{d\nu} = f \quad \text{ou} \quad d\mu = f d\nu.$$

Fixado um ponto qualquer x do espaço euclidiano $\mathbb{R}^d$, dotado da métrica usual, e um real positivo $r \geq 0$. Denotaremos por: (i) $B(x; r)$ a bola aberta de raio r e centrada em x ; (ii) $B[x; r]$ a bola fechada de raio r e centrada em x . Em outras palavras:

$$B(x; r) = \{a \in \mathbb{R}^d; \|a - x\| < r\} \quad \text{e} \quad B[x; r] = \{a \in \mathbb{R}^d; \|a - x\| \leq r\}.$$

Finalmente, seja $A \subseteq \mathbb{R}^d$ um subconjunto qualquer do espaço euclidiano, munido da métrica usual. Denotaremos a fronteira deste conjunto por:

$$\text{fr}(A) = \{x \in A; \forall r > 0 \ B(x; r) \cap A \neq \emptyset \text{ e } B(x; r) \cap A^c \neq \emptyset\}.$$

Além disso, dado um real $r \geq 0$ qualquer, representaremos por

$$\text{fr}(A, r) = \{x \in A; d(x, \text{fr}(A)) < r\}$$

o conjunto dos pontos de A , que distam de sua fronteira de um valor $\varepsilon < r$, com $d(x, \text{fr}(A)) = \inf_{y \in \text{fr}(A)} d(x, y)$ a distância do ponto x a fronteira $\text{fr}(A)$, como ilustrado na Figura 23.

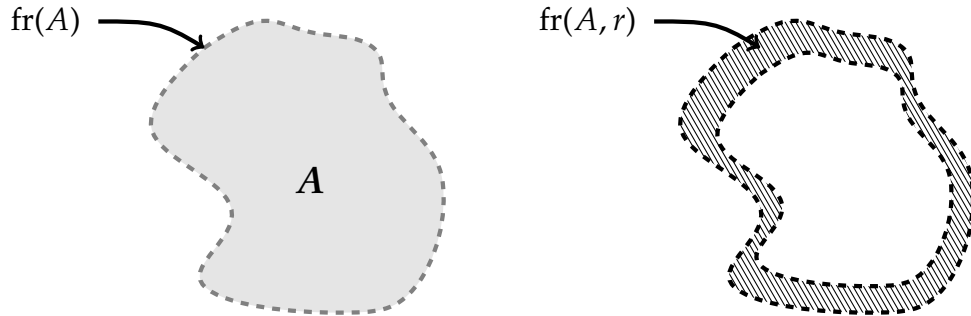


Figura 23: Fronteira de um conjunto A , na figura à esquerda, e o conjunto dos pontos de A suficientemente próximos da fronteira (distando de no máximo um valor r), figura hachurada à direita.

3.2 FILTRAÇÕES SIMPLICIAIS

Sejam $\mathcal{X} \subseteq \mathbb{R}^d$ um conjunto não nulo e localmente finito, e $\mathbb{K} = (K_r)_{r \geq 0}$ uma filtração simplicial cujos vértices são os pontos do conjunto $\mathcal{X}$. Nesta seção apresentaremos algumas condições à filtração $\mathbb{K}$ de modo que consigamos garantir, em um certo sentido, que a construção dos simplexes $\sigma \in K_r$ sejam: *invariante por translação*, *mensurável*, e

que, dado um ponto qualquer $x \in \mathcal{X}$, simplexes contendo esse ponto devem distar, no máximo, de um valor $\epsilon_r^{\rho, \kappa}$ que dependerá do índice da filtração e de duas funções auxiliares, associadas a filtração $\mathbb{K}$.

Definição 3.2.1 (Filtrações Simpliciais). Seja $\mathcal{X} \subseteq \mathbb{R}^d$ um conjunto não nulo e

$$\kappa : \{H \subseteq \mathbb{R}^d; H \text{ é localmente finito} \} \rightarrow [0, \infty] \quad (13)$$

uma função satisfazendo a condição, de que para todo par de conjuntos localmente finitos $s, t \subseteq \mathbb{R}^d$, vale $s \subseteq t \implies \kappa(s) \leq \kappa(t)$. Então, dizemos que $\mathbb{K} = (K_r)_{r \geq 0}$ é uma filtração simplicial de vértices em $\mathcal{X}$ e *indexada* por κ , se para todo simplexo $\sigma \in K_r$ valer a condição

$$\sigma \in K_r \iff \kappa(\sigma) \leq r, \quad (14)$$

para todo real $r \geq 0$.

Observação 3.2.1. Necessitamos impor a restrição $s \subseteq t \implies \kappa(s) \leq \kappa(t)$ para as funções κ , que indexam uma filtração $\mathbb{K}$ qualquer da Definição 3.2.1, devido ao fato de que quando trabalhamos com filtrações em homologia persistente, as faces de simplexes devem sempre nascer antes que os simplexes que as contem, i.e., devemos ter

$$\sigma \subseteq \sigma' \iff \min_{r \geq 0} \{r; \sigma \in K_r\} \leq \min_{r \geq 0} \{r; \sigma' \in K_r\}.$$

Destas filtrações $\mathbb{K}$, indexadas por uma função κ , temos o controle de como os simplexes são construídos. Resta-nos, agora, estabelecer algumas condições na função κ para que possamos, então, garantir as propriedades enunciadas no começo desta seção.

Definição 3.2.2 (Filtrações simpliciais - condições). Seja $\mathbb{K} = (K_r)_{r \geq 0}$ uma filtração simplicial indexada por uma função κ , (definida em 13), e cujos vértices sejam dados pelo conjunto não nulo $\mathcal{X} \subseteq \mathbb{R}^d$. Então, estabeleceremos que, no decorrer de todo este capítulo, para esta filtração $\mathbb{K}$ existirá uma função

$$\rho : [0, \infty] \rightarrow [0, \infty) \quad (15)$$

tal que, ambas κ e ρ , deverão satisfazer as seguintes restrições: (i) κ será mensurável, para algum espaço de medida a ser definido, dependendo do contexto em que estivermos trabalhando; (ii) κ será invariante por translação; (iii) ρ será uma função crescente tal que, dado o simplexo $\sigma = \{x, y\}$, valerá

$$|x - y| \leq \rho(\kappa(\sigma)). \quad (16)$$

Notemos que, dada uma filtração simplicial $\mathbb{K}$ de vértices não nulos $\mathcal{X} \subseteq \mathbb{R}^d$, indexada por uma função κ e satisfazendo as condições em 3.2.2, então teremos garantido pela desigualdade 16, pela condição 14 e pela função ρ ser crescente, que

$$\forall x, y \in \sigma, \sigma \in K_r \implies |x - y| \leq \rho(\kappa(\sigma)) \leq \rho(r),$$

pois $\sigma \in K_r$ implica $\kappa(\sigma) \leq r$.

Finalmente, encerramos esta seção observando que as *filtrações simpliciais de Čech* e de *Vietoris Rips*, que são as filtrações nas quais estaremos mais interessados, satisfazem as restrições impostas acima. (seção 1.3 do paper [1].)

3.3 NÚMERO DE PERSISTÊNCIA DE BETTI

Nesta seção definiremos o objeto central de estudo deste capítulo: *o número de persistência de Betti*. Como mostraremos a seguir os números de Betti estabelecem uma conexão direta entre diagramas de persistência, de uma dada filtração, e a medida de contagem de pontos contidos neste diagrama.

Sejam $\mathcal{X}$ um conjunto não nulo, $\mathbb{K}$ uma filtração simplicial, indexada em $\mathbb{R}_{\geq 0}$, do conjunto de pontos $\mathcal{X}$ e $H_*(\mathcal{F})$ os módulos de persistência desta filtração, tal qual apresentados no Capítulo 2. Sabemos que para todo par de índices $r, s \in \mathbb{R}_{\geq 0}$, vale o diagrama comutativo abaixo, Figura 24, sempre quando $r \leq s$.

$$\begin{array}{ccccccccc} (0) & \leftarrow & H_0(X_r) & \leftarrow & H_1(X_r) & \leftarrow & H_2(X_r) & \leftarrow & \cdots & \leftarrow & H_n(X_r) & \leftarrow & H_{n+1}(X_r) & \leftarrow & \cdots \\ & & \downarrow i_{*,0}^{r,s} & & \downarrow i_{*,1}^{r,s} & & \downarrow i_{*,2}^{r,s} & & & & \downarrow i_{*,n}^{r,s} & & \downarrow i_{*,n+1}^{r,s} & & \\ (0) & \leftarrow & H_0(X_s) & \leftarrow & H_1(X_s) & \leftarrow & H_2(X_s) & \leftarrow & \cdots & \leftarrow & H_n(X_s) & \leftarrow & H_{n+1}(X_s) & \leftarrow & \cdots \end{array}$$

Figura 24: Diagrama representando os módulos de persistência da filtração $\mathbb{K}$.

O *número de persistência de Betti* de dimensão q , no intervalo de tempo $[r, s]$, será a dimensão da imagem da transformação linear $i_{*,n}^{r,s}$:

$$\beta_q^{r,s}(\mathbb{K}(\mathcal{X})) := \dim \left(\text{Im}(i_{*,n}^{r,s}) \right),$$

com $i_{*,n}^{r,s}$ o operador induzido pelas inclusões da filtração. Antes de prosseguirmos, será importante estabelecermos uma notação menos carregada para o número de persistência de Betti, pois o mesmo será usado exaustivamente mais a frente.

Notação. Sempre quando formos calcular o número de persistência de Betti de uma filtração de um conjunto de pontos $\mathcal{X}$ e não houver risco de confusão, com a filtração sendo usada, então utilizaremos, por simplicidade, apenas a representação

$$\beta_q^{r,s}(\mathcal{X})$$

ao invés de sua forma completa

$$\beta_q^{r,s}(\mathbb{K}(\mathcal{X})).$$

Agora, apresentamos uma importante propriedade do número de persistência de Betti, que nos permitirá interpretá-lo como a medida de contagem de pontos em uma dada região do diagrama de persistência. Seja $\mathbb{K}$ uma filtração simplicial de um conjunto de pontos não nulo $\mathcal{X}$, e $H_*(\mathbb{K})$ sua respectiva homologia persistente, tal que, para todo $q \in \mathbb{N}_0$, tenhamos bem definido seus diagramas de persistência $\text{Dgm}(H_q(\mathbb{K}))$ que, por simplicidade, denotaremos apenas por Dgm_q . Então, considerando o subespaço

$$\Delta = \{(x, y) \in \mathbb{R}^2; x \leq y\},$$

e a medida

$$\text{Dgm}_q(B) = \sum_{(b,d) \in \text{Dgm}_q} m_{(b,d)} \delta_{(b,d)}(B), \quad (17)$$

onde $B \subseteq \Delta$ é um boreliano do espaço e $m_{(b,d)}$ é a quantidade de vezes em que o ponto (b, d) encontra-se no multiset $\text{Dgm}_q(B)$, com

$$\delta_{(b,d)}(B) = \begin{cases} 1 & \text{se } (b, d) \in B \\ 0 & \text{caso contrário,} \end{cases}$$

temos garantido (vide capítulo 1, Seção 3, de [26]) que as seguintes igualdades são verdadeiras:

$$(i) \text{Dgm}_p([-\infty, a] \times [b, \infty]) = \beta_p^{a,b}(\mathcal{X}) \quad (18)$$

$$(ii) \text{Dgm}_p([-\infty, a] \times [b, c]) = \beta_p^{a,b}(\mathcal{X}) - \beta_p^{a,c}(\mathcal{X}) \quad (19)$$

$$(iii) \text{Dgm}_p([a, b] \times [c, \infty]) = \beta_p^{b,c}(\mathcal{X}) - \beta_p^{a,c}(\mathcal{X}) \quad (20)$$

$$(iv) \text{Dgm}_p([a, b] \times [c, d]) = \beta_p^{b,c}(\mathcal{X}) - \beta_p^{a,c}(\mathcal{X}) + \beta_p^{a,d}(\mathcal{X}) - \beta_p^{b,d}(\mathcal{X}), \quad (21)$$

com $a \leq b \leq c \leq d$, e considerando que o número de persistência de Betti seja finito nos casos (ii)-(iv).

A Figura 25 ilustra como o número de persistência de Betti, $\beta_p^{a,b}(\mathcal{X})$, pode ser interpretado como a contagem de grupos de homologia que nasceram, até o instante de tempo $t = a$, e que sumiram, apenas após o instante de tempo $t = b$.

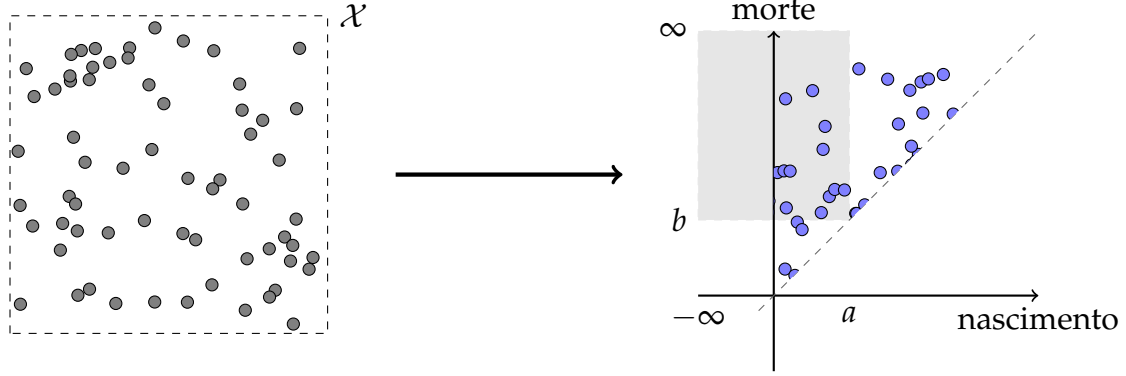


Figura 25: Diagrama de persistência de uma filtração de um conjunto de pontos $\mathcal{X}$ e seu respectivo número de persistência de Betti $\beta_q^{a,b}(\mathcal{X})$, região em cinza claro no diagrama.

3.4 CONTAGEM DE SIMPLEXOS E DESIGUALDADES

Boa parte das desigualdades utilizadas neste capítulo serão resultantes da contagem de simplexos com, ou sem, intersecção com borelianos limitados contendo a origem. Devido a importância destas medidas de contagem, estabeleceremos uma notação para estas quantidades, seguindo de perto a notação proposta em [1].

Dado um conjunto não vazio $\mathcal{X} \subseteq \mathbb{R}^d$, um boreliano $A \subseteq \mathbb{R}^d$ e a filtração simplicial de vértices em $\mathcal{X}$, $\mathbb{K}(\mathcal{X}) = (K_r(\mathcal{X}))_{r \in \mathbb{R}_{\geq 0}}$, denotemos por

$$S_j(\mathcal{X}, l; A) = \text{card}\{\sigma \in K_l(\mathcal{X}); \sigma \cap A \neq \emptyset \text{ e } \dim(\sigma) = j\}$$

e

$$S_j(\mathcal{X}, l) = \text{card}\{\sigma \in K_l(\mathcal{X}); \dim(\sigma) = j\}$$

a contagem dos simplexos de dimensão j , pertencentes ao complexo simplicial $K_l(\mathcal{X})$, e satisfazendo a restrição de ter, ou não, intersecção com um boreliano. Um exemplo de cálculo, das contagens de simplexos acima, pode ser vista na Figura 26. De maneira similar, definimos o conjunto auxiliar

$$S_j(\mathcal{X}, l) = \{\sigma \in K_l(\mathcal{X}); \dim(\sigma) = j\},$$

com $j, l \in \mathbb{N}$, satisfazendo a igualdade

$$S_j(\mathcal{X}, l; A) = \text{card}(S_j(\mathcal{X}, l)).$$

A partir destas estruturas somos capazes de enunciar uma majoração muito importante, que estabelecerá uma cota superior para a diferença, em valor absoluto, do

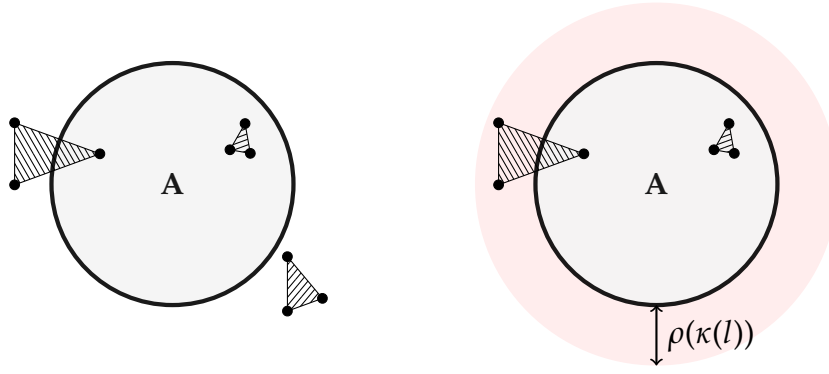


Figura 26: Na figura da esquerda temos uma ilustração em que a contagem de simplexes, de dimensão 2, e pertencentes ao complexo simplicial $K_l(\mathcal{X})$, vale 2, i.e., $S_2(\mathcal{X}, l; A) = 3$. Na figura da direita, a faixa em rosa representa a distância máxima em que os pontos podem estar, em relação ao conjunto A , tal que ainda possa existir um simplexo $\sigma \in K_l(\mathcal{X})$ tal que $\sigma \cap A \neq \emptyset$.

número de persistência de Betti entre duas filtrações. Do artigo [1], Lema 2.11, temos que dadas duas filtrações $\mathbb{K} = (K_r)_{r \geq 0} \subseteq \tilde{\mathbb{K}} = (\tilde{K}_r)_{r \geq 0}$, baseadas no mesmo conjunto de vértices $\mathcal{X}$, vale a seguinte desigualdade

$$|\beta_q^{r,s}(\tilde{\mathbb{K}}(\mathcal{X})) - \beta_q^{r,s}(\mathbb{K}(\mathcal{X}))| \leq \sum_{j=q, q+1} \text{card} \left(\tilde{\mathcal{S}}_j(\mathcal{X}, s) \setminus \mathcal{S}_j(\mathcal{X}, s) \right) + \sum_{j=q, q+1} \text{card} \left(\{ \sigma \in \mathcal{S}_j(\mathcal{X}, s) \setminus \mathcal{S}_j(\mathcal{X}, r); \tilde{t}_\sigma \leq r \} \right),$$

com

$$\mathcal{S}_j(\mathcal{X}, s) = \{ \sigma \in K_s; \dim(\sigma) = j \}, \quad \tilde{\mathcal{S}}_j(\mathcal{X}, s) = \{ \sigma \in \tilde{K}_s; \dim(\sigma) = j \}$$

e t_σ o instante em que o simplexo σ aparece pela primeira vez na filtração $\mathbb{K}$ (idem para $\tilde{t}_\sigma$).

Agora, consideremos o caso em que temos as constantes reais $L \leq L'$, um conjunto não nulo $\mathcal{X} \subseteq \mathbb{R}^d$, e o conjunto $\mathcal{X}_{\Lambda_L^d} = \mathcal{X} \cap \Lambda_L^d$, sendo

$$\Lambda_L^d = \left[\frac{-L}{2}, \frac{L}{2} \right)^d = \prod_{i=1}^d \left[\frac{-L}{2}, \frac{L}{2} \right).$$

Então, tomando as filtrações

$$\mathbb{K}(\mathcal{X}_{\Lambda_L^d}) = \{K_t\}_{t \geq 0} \subseteq \mathbb{K}(\mathcal{X}_{\Lambda_{L'}^d}) = \{K'_t\}_{t \geq 0},$$

cujas regras para construção dos simplexos sejam as mesmas, i.e., as funções κ e ρ são as mesmas (vide Seção 3.2), concluímos que

$$t'_\sigma \leq r \text{ e } \sigma \in \mathcal{S}_j(\mathcal{X}_{\Lambda_L^d}, s) \implies \dim(\sigma) = j \text{ e } \sigma \in K'_r \implies \dim(\sigma) = j \text{ e } \kappa(\sigma) \leq r,$$

ou seja

$$t'_\sigma \leq r \text{ e } \sigma \in \mathcal{S}_j(\mathcal{X}_{\Lambda_L^d}, s) \implies \sigma \in \mathcal{S}_j(\mathcal{X}_{\Lambda_L^d}, r),$$

que fornece a igualdade

$$\text{card} \left(\{ \sigma \in \mathcal{S}_j(\mathcal{X}, s) \setminus \mathcal{S}_j(\mathcal{X}, r); \tilde{t}_\sigma \leq r \} \right) = 0,$$

e, portanto, obtemos a **importante desigualdade**:

$$\left| \beta_q^{r,s}(\mathcal{X}_{\Lambda_{L'}}) - \beta_q^{r,s}(\mathcal{X}_{\Lambda_L}) \right| \leq \sum_{j=q, q+1} \text{card} \left(\mathcal{S}_j(\mathcal{X}_{\Lambda_{L'}^d}, s) \setminus \mathcal{S}_j(\mathcal{X}_{\Lambda_L^d}, s) \right) \quad (22)$$

para todo $0 < L \leq L'$.

3.5 CLASSES DE PROCESSOS PONTUAIS

Nesta seção começaremos apresentando a definição de processos pontuais, e em seguida, definiremos dois tipos de classes de processos pontuais que serão utilizadas no decorrer do capítulo:

- (i) a classe dos processos pontuais majorados por algum processo pontual de Poisson, homogêneo e de taxa unitária;
- (ii) a classe dos processos pontuais de Gibbs.

Esta seção seguirá, bem de perto, a abordagem adotada no livro [16] e nos artigos [3] e [7].

3.5.1 Processos Pontuais

Sejam $\mathbb{R}^d$ o espaço euclidiano munido da métrica usual e $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ o espaço de medida cuja σ -álgebra é a álgebra gerada pelos conjuntos abertos de $\mathbb{R}^d$, os borelianos. Deste espaço, consideremos o conjunto de todas as medidas de contagem própria

$$N = \left\{ \eta : \mathcal{B}(\mathbb{R}^d) \rightarrow \mathbb{N}_0; \eta = \sum_{i=1}^N \delta_{x_i}, \text{ com } x_i \in \mathbb{R}^d \text{ para todo } i \in \mathbb{N} \text{ e } N \in \mathbb{N} \cup \{\infty\} \right\}, \quad (23)$$

onde, para todo boreliano A e para todo $x \in \mathbb{R}^d$, vale

$$\delta_x(A) = \mathbb{1}_A(x) = \begin{cases} 1, & \text{se } x \in A; \\ 0, & \text{caso contrário.} \end{cases}$$

Para que o conjunto N admita estrutura de espaço de medida, definimos a seguinte σ -álgebra:

$$\mathcal{A} = \left\langle \left\{ \{ \eta \in N; \eta(B) = k \}; k \in \mathbb{N}_0, B \in \mathcal{B}(\mathbb{R}^d) \right\} \right\rangle, \quad (24)$$

com $\langle \cdot \rangle$ representando a σ -álgebra gerada pela família de conjuntos que tornam as funções

$$\begin{aligned} N_B : N &\rightarrow \mathbb{N}_0 \\ \eta &\mapsto \eta(B) \end{aligned} \quad (25)$$

$\mathcal{A}$ -mensuráveis, para todo $B \in \mathcal{B}(\mathbb{R}^d)$.

Definição 3.5.1 (Processos Pontuais em $\mathbb{R}^d$). Sejam $(\Omega, \mathcal{S}, \mu)$ um espaço de probabilidade e $(N, \mathcal{A})$ um espaço de medida, com N o conjunto definido em 23 e $\mathcal{A}$ a σ -álgebra definida em 24. Então, toda função mensurável

$$\Phi : (\Omega, \mathcal{S}, \mu) \rightarrow (N, \mathcal{A})$$

será chamada de *processo pontual*, enquanto que a medida

$$\nu(B) = \mathbb{E}[N_B \circ \Phi] = \int_{\Omega} N_B(\Phi(w)) \mu(dw),$$

para todo $B \in \mathcal{A}$, com N_B a função dada em 25, será chamada de *medida de intensidade* do processo Φ .

Observação 3.5.1. Todo processo pontual $\Phi : (\Omega, \mathcal{S}, \mu) \rightarrow (N, \mathcal{A})$ que iremos estudar no decorrer deste capítulo devem satisfazer as condições:

- (i) Para todo boreliano B em $\mathbb{R}^d$, compacto, temos quase certamente que a quantidade de pontos nesse conjunto será finita:

$$\Phi \circ \mu([N_B < \infty]) := \mu(\Phi^{-1}([N_B < \infty])) = 1;$$

- (ii) Para todo ponto $x \in \mathbb{R}^d$, não podemos ter mais de uma ocorrência neste mesmo ponto:

$$\Phi \circ \mu([N_{\{x\}} \leq 1]) = 1.$$

Uma propriedade importante, que alguns processos pontuais detêm, é a *invariância por translação* ou *estacionariedade*. Seja $x \in \mathbb{R}^d$ um ponto qualquer e Θ_x a translação das medidas próprias $\eta \in \mathcal{N}$ (medidas estas definidas em 23), dadas por

$$\theta_x \eta(B) = \eta(B + x),$$

com B um boreliano qualquer. Então, dizemos que um processo pontual Φ é invariante por translação, ou estacionário, se for válida a igualdade em distribuição

$$\theta_x \Phi \stackrel{d}{=} \Phi, \quad (26)$$

para todo ponto $x \in \mathbb{R}^d$.

Notemos, que da invariância por translação temos a seguinte propriedade: para todo boreliano $B \in \mathcal{B}(\mathbb{R}^d)$, todo ponto $x \in \mathbb{R}^d$ e todo natural $k \in \mathbb{N}_0$, a seguinte igualdade é verdadeira

$$\Phi \circ \mu([N_B = k]) = \theta_x \Phi \circ \mu([N_{B+x} = k]),$$

cuja intuição apresentamos na Figura 27.

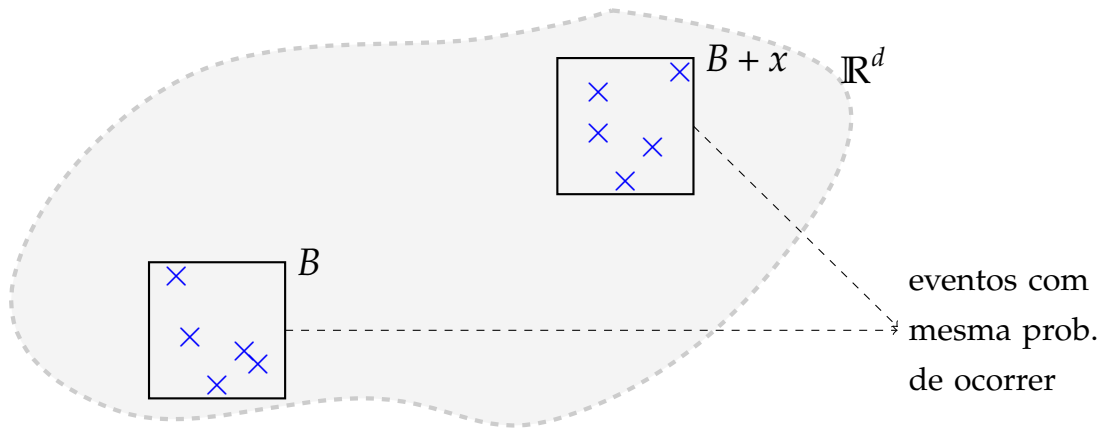


Figura 27: Representação de um processo pontual estacionário Φ , onde translação de janelas do $\mathbb{R}^d$ devem ter a mesma probabilidade de conter a mesma quantidade de pontos de Φ .

Finalmente, para simplificarmos a notação, adotaremos a seguinte convenção daqui em diante: sejam B um boreliano qualquer em $\mathbb{R}^d$ e Φ um processo pontual, então, por simplicidade, denotaremos $N_B \circ \Phi$ por simplesmente $\Phi(B)$.

3.5.2 Processos Pontuais majorados por um PPP homogêneo

Os processos pontuais de Poisson homogêneos terão a característica de controlarem a média do número de persistência de Betti quando olhamos para processos pontuais Φ , satisfazendo quase certamente a majoração $\Phi(B) < \mathcal{P}(B)$, com B um boreliano. Esta propriedade nos permitirá estender resultados de convergência de processos pontuais estacionários, para processos pontuais majorados por um PPP homogêneo.

Em vista do exposto acima, estabelecemos abaixo a definição dos PPP.

Definição 3.5.2 (Processos Pontuais de Poisson, homogêneos e de taxa λ). Sejam $(\Omega, \mathcal{S}, \mu)$ um espaço de probabilidade e $(N, \mathcal{A})$ um espaço de medida, com N o conjunto definido em 23 e $\mathcal{A}$ a σ -álgebra definida em 24, e $\mathcal{P}_\lambda$ um processo pontual, tal qual definido em 3.5.1. Então, $\mathcal{P}_\lambda$ será um processo pontual de Poisson, homogêneo e de taxa λ , se as seguintes condições forem satisfeitas:

- (i) Para todo boreliano $B \in \mathcal{A}$, a variável aleatória $\mathcal{P}_\lambda(B)$ tem como distribuição de probabilidade a distribuição de uma Poisson de taxa $\lambda|B|$;
- (ii) Sejam $B_1, B_2, \dots, B_N$ borelianos disjuntos do $\mathbb{R}^d$. Então, as variáveis de contagem $\mathcal{P}_\lambda(B_1), \mathcal{P}_\lambda(B_2), \dots, \mathcal{P}_\lambda(B_N)$ serão variáveis aleatórias independentes.

A partir dos PPP homogêneos construímos a família de processos pontuais $\mathcal{C}$ cujas propriedades topológicas serão exploradas mais a frente, na Seção 3.6, dada por

$$\mathcal{C} := \left\{ \Phi \text{ é um processo pontual; } \Phi(B) \stackrel{q.c.}{\leq} \mathcal{P}(B), \forall B \in \mathcal{B}(\mathbb{R}^d) \right\}. \quad (27)$$

Observação 3.5.2. A família de processos pontuais $\mathcal{C}$ contém todos os PPP não homogêneos de intensidade

$$d\nu = f d(x),$$

com $f : \mathbb{R}^d \rightarrow [0, 1]$. De fato, Consideremos o PPP homogêneo, de taxa unitária e marcado Φ , tal que, para todo $w \in \Omega$, vale

$$\Phi(w) = \sum_{i=1}^N \delta_{(x_i, y_i)},$$

com $N \in \mathbb{N} \cup \{\infty\}$ e $y_1, \dots, y_N$ variáveis aleatórias, independentes entre si e entre as x_i , com $y_i \in \{0, 1\}$. Assumindo que o kernel deste processo marcado seja a função

$$\begin{aligned} K : \mathbb{R}^d \times \{0, 1\} &\rightarrow [0, 1] \\ (x, y) &\mapsto f(x)\delta_1(y) + (1 - f(x))\delta_0(y) \end{aligned}$$

temos, vide Seção 5.3 do livro [16], que o processo pontual emagrecido Φ^1 , usando a restrição $y_i = 1$ (vide Figura 28, onde pontos vermelhos são marcas valendo zero e pontos azuis são marcas valendo um), será um PPP de intensidade

$$d\nu = f d(x).$$

Portanto, como processos pontuais de Poisson, não homogêneos, e com função de intensidade assumindo valores no compacto $[0, 1]$ podem ser interpretados como o emagrecimento de um PPP, homogêneo e de taxa 1, concluimos que estes processos estão contidos em $\mathcal{C}$.

Observação 3.5.3. A classe $\mathcal{C}$ admite processos pontuais que não são estacionários. De fato, decorre da Observação 3.5.2, tomando funções de intensidade que não sejam invariantes por translação.

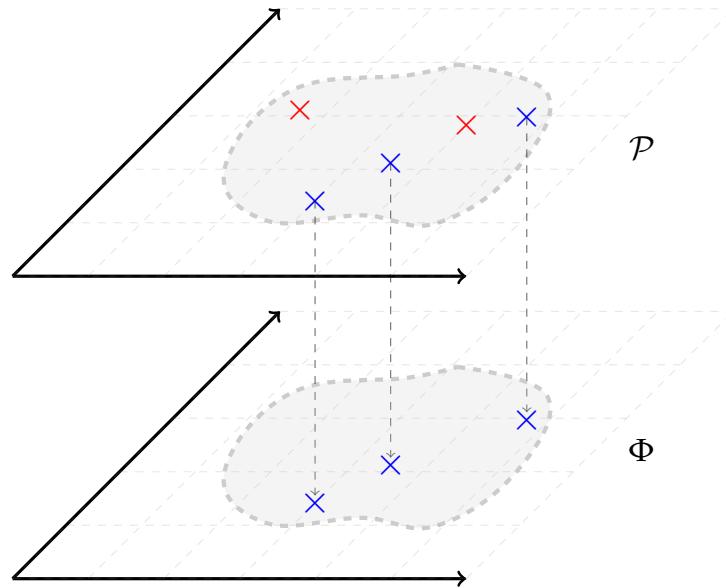


Figura 28: Processo pontual pertencente a classe $\mathcal{C}$, equação 27, o qual é resultante do emagrecimento de um processo pontual de Poisson.

3.5.3 Processos pontuais de Gibbs

A segunda classe de processos que citamos no início desta seção, e que abordaremos quando estivermos estudando convergência em distribuição, serão os *processos pontuais gibbsianos*. A seguir, mostraremos que tais processos, sob certas condições, terão

comportamento próximos ao de processos de Poisson, e garantindo, no limite, que possamos usar o Teorema do Limite Central provado em [3].

Para a construção destes processos, precisaremos da noção de *funções de energia* ou *potenciais* (vide os artigos [3] e [7]). Dizemos que uma função

$$\Psi : \{\mathbf{X} \subset \mathbb{R}^d; \mathbf{X} \text{ é finito}\} \rightarrow \mathbb{R} \cup \{\infty\}.$$

é uma função de energia (ou, potencial) se as seguintes condições forem satisfeitas

(i) Ψ é invariante por translação e por rotação;

(ii) Ψ é não degenerada:

$$\Psi(\emptyset) < +\infty;$$

(iii) Para todo subconjunto finito $\mathbf{X}$ de $\mathbb{R}^d$, e para todo ponto $x \in \mathbf{X}$, vale:

$$\Psi(\mathbf{X}) < +\infty \implies \Psi(\mathbf{X} \setminus \{x\}) < +\infty;$$

(iv) Existe uma constante real $A \in \mathbb{R}$ tal que, para todo subconjunto finito $\mathbf{X}$ de $\mathbb{R}^d$, temos

$$\Psi(\mathbf{X}) \geq A \text{ card}(\mathbf{X}).$$

Agora que definimos as funções de energia, conseguimos definir o processo de Gibbs, que será nada mais que um processo absolutamente contínuo a um processo de Poisson.

Definição 3.5.3. Sejam Ψ uma função de energia e $\beta > 0$ uma constante, que chamaremos de *inverso da temperatura*. Então, um processo pontual Φ^β é um processo pontual de Gibbs se, para todo boreliano limitado $\Lambda \subseteq \mathbb{R}^d$, sua distribuição de probabilidade, μ_Λ , for absolutamente contínua com relação a distribuição de probabilidade μ_Λ^0 de um processo pontual de Poisson, homogêneo e de taxa τ , satisfazendo a igualdade

$$\mu_\Lambda(d\mathbf{X}) = \frac{\exp(-\beta \Psi(\mathbf{X} \cap \Lambda))}{Z(\beta \Psi_\Lambda)} \mu_\Lambda^0(d\mathbf{X}),$$

com $\mathbf{X}$ subconjuntos finitos de $\mathbb{R}^d$ e $Z(\beta \Psi_\Lambda)$ uma constante normalizadora.

Decorre da definição acima, que para a construção de processos gibbsianos, precisamos apenas definir funções de energia Φ que satisfaçam as condições (i)-(iv) listadas acima.

Exemplo 3.5.1. Sejam $K \subseteq \mathbb{R}^d$ um compacto qualquer, $\Lambda \subseteq \mathbb{R}^d$ um conjunto limitado e μ_Λ^0 a distribuição de um PPP homogêneo de taxa τ . Então, o processo de interação de área $\Phi^{K,1}$, com distribuição de probabilidade

$$\mu_\Lambda(d\mathbf{X}) = \frac{\exp(-\ell(\mathbf{X} \cap \Lambda + K))}{Z(\beta\Psi_\Lambda)} \mu_\Lambda^0(d\mathbf{X})$$

é um processo pontual gibbsiano, com função de energia

$$\Psi(\mathbf{X}) = \ell(\mathbf{X} + K).$$

Como já observado inicialmente, processos gibbsianos, quando satisfazem certas hipóteses, tem a propriedade de se comportarem muito assemelhadamente a um processo de Poisson. De fato, uma condição suficiente para obtermos esta propriedade decorre da existência, ou não, de um esquema de simulação perfeita (vide o artigo [4]) para este processo. Uma forma de conseguirmos aplicar a simulação perfeita a um processo gibbsiano, é olhando para o alcance de sua função de energia:

Definição 3.5.4 (Alcance de interação finito). Seja Φ^β um processo pontual gibbsiano, com função de potencial Ψ . Então, dizemos que Φ^β tem *alcance de interação finito* se existir um real positivo $r^\Psi < \infty$ tal que, para todo conjunto finito $\mathcal{X} \subseteq \mathbb{R}^d$, valer

$$\Delta(0, \mathcal{X}) = \Delta(0, \mathcal{X} \cap B(0, r^\Psi)), \quad (28)$$

com

$$\Delta(0, \mathcal{X}) = \Psi(\mathcal{X} \cup \{0\}) - \Psi(\mathcal{X}). \quad (29)$$

Da propriedade de alcance finito, temos pelo Lema 3.5.1 (cuja demonstração pode ser vista em [3]) que para um certo par de parâmetros (τ, β) de um processo de Gibbs, satisfazendo uma certa condição, temos garantido a existência de um esquema de simulação perfeita para este processo:

Lema 3.5.1. *Seja Φ^β um processo gibbsiano, com distribuição de probabilidade absolutamente contínua com relação a um processo pontual de Poisson, homogêneo e de taxa τ . Então, se a função de energia de Φ^β tiver alcance finito, basta que os parâmetros τ e β satisfaçam a desigualdade*

$$\tau \ell(B(0; 1)) e^{-\beta m_0} (r_0 + 1)^d < 1, \quad (30)$$

com

$$m_0 = \inf_{\mathcal{X} \text{ finito}} \Delta(0, \mathcal{X})$$

para que o processo gibbsiano admita simulação perfeita.

A próxima definição formaliza a ideia de processos gibbsianos cujo comportamento se “aproxima” a um processo de Poisson, que chamaremos de *Processos Pontuais do tipo Poisson*.

Definição 3.5.5 (processos pontuais do tipo Poisson). Dizemos que Ξ é um Processo Pontual tipo Poisson quando:

- (i) Ξ for estocasticamente dominado por um processo pontual de Poisson de taxa τ , $\mathcal{P}_\tau$. Em outras palavras, Ξ é estocasticamente dominado por $\mathcal{P}_\tau$ quando para todo boreliano B e todo natural n valer

$$\mathbb{P}(\text{card}(\Xi \cap B) \geq n) \leq \mathbb{P}(\text{card}(\mathcal{P}_\tau \cap B) \geq n);$$

- (ii) Existir constantes $C_0, r_0 \in \mathbb{R}_{>0}$ tal que, para todo $x \in \mathbb{R}^d$, para todo raio $r \in [r_0, \infty)$ e para todo conjunto de pontos localmente finito $\mathcal{E}_r(x) \subseteq B(x; r)^c$, valer a desigualdade:

$$\mathbb{P}\left(\Xi \cap B(x; r) = \emptyset \mid \Xi \cap B(x; r)^c = \mathcal{E}_r(x)\right) \leq e^{-C_0 r^d}.$$

Finalmente, temos o Lema que nos garante que uma dada classe de processos gibbsianos é, também, um processo do tipo Poisson:

Lema 3.5.2. *Seja Φ^β um processo gibbsiano cuja função de potencial seja pertencente a classe dos processos pontuais de interação de área. Então, se Φ^β tiver alcance finito, teremos que Φ^β será, também, um processo pontual do tipo Poisson quando seus parâmetros satisfizerem a desigualdade*

$$\tau \ell(B(0; 1)) e^{-\beta m_0} (r_0 + 1)^d < 1,$$

onde τ é o parâmetro da Poisson tal que Φ^β é absolutamente contínua.

Demonstração. A demonstração deste fato sai imediatamente do fato de que, quando o processo gibbsiano tem alcance finito, a desigualdade acima é suficiente para termos garantido a simulação perfeita, de acordo com o Lema 3.5.1. Finalmente, usando o Lema 3.2 do artigo [3], temos o resultado. $\square$

3.6 LEI FORTE DO NÚMERO DE PERSISTÊNCIA DE BETTI

Nesta seção provaremos que, ao considerarmos qualquer processo pontual Φ pertencente a família de processos pontuais $\mathcal{C}$ da Subseção 3.5.2, em uma janela $\Lambda_L^d \subseteq \mathbb{R}^d$, e dotado de uma filtração satisfazendo as condições em 3.2, valerá que a taxa

$$\frac{\beta_q^{r,s}(\Phi_{\Lambda_L^d}) - \mathbb{E}[\beta_q^{r,s}(\Phi_{\Lambda_L^d})]}{|\Lambda_L^d|} \quad (31)$$

convergirá quase certamente para zero, conforme o volume da caixa Λ_L^d crescer para o infinito. Acima, os termos q, r, s denotam constantes do modelo, denotando, respectivamente, a dimensão da homologia e o intervalo de tempo $[r, s] \subseteq \mathbb{R}_{\geq 0}^d$.

Para provarmos a convergência quase certa do termo da equação 31 precisaremos provar alguns resultados preliminares. Começamos provando que, a razão do número de persistência de Betti pelo volume de sua caixa, é majorado por uma constante sempre que considerarmos qualquer processo pontual pertencente a família de processos pontuais $\mathcal{C}$.

Lema 3.6.1. *Seja Φ um processo pontual pertencente a família de processos pontuais $\mathcal{C}$ da Seção 3.5.2. Então, para toda filtração simplicial satisfazendo as restrições em 3.2, temos que, para todos reais $r \leq s$ positivos e para todo inteiro $0 \leq q < d$, existe uma constante real e positiva c_q tal que*

$$\frac{\mathbb{E}[\beta_q^{r,s}(\Phi_A)]}{|A|} \leq c_q,$$

com A um boreliano do espaço.

Demonstração. Notemos que o número de persistência de Betti admite a seguinte desigualdade

$$\beta_q^{r,s}(\Phi) = \dim \text{Im}(l_{*,q}^{r,s}) \leq \dim(H_q(K_s)) \leq \dim(Z_q(K_s)) \leq S_q(\Phi_A, s; A).$$

Além do mais, usando o fato de que existe um processo pontual de Poisson, homogêneo e de taxa unitária, $\mathcal{P}$, que majora o nosso processo pontual, obtemos que

$$\beta_q^{r,s}(\Phi) \leq S_q(\mathcal{P}, s; A),$$

onde, novamente, $S_q(\mathcal{P}, s; A)$ pode ser majorado superiormente por toda combinação possível de simplexos $\sigma = \langle a_0, \dots, a_q \rangle$ com pontos em Φ e com intersecção em A . Desta forma, concluímos que

$$\begin{aligned} S_q(\mathcal{P}, s; A) &\leq \sum_{x \in \Phi \cap A} \binom{\mathcal{P}(B_x^{\rho(s)})}{q} \\ &\leq \sum_{x \in \Phi \cap A} \left(\mathcal{P}(B_x^{\rho(s)}) \right)^{q+1}, \end{aligned}$$

com $B_x^{\rho(s)} := B(x; \rho(s))$. Finalmente, note que a medida

$$\mu(A) := \mathbb{E} \left[\sum_{x \in \Phi \cap A} \left(\mathcal{P}(B_x^{\rho(s)}) \right)^{q+1} \right]$$

é uma medida invariante por translação em $\mathbb{R}^d$. Logo, existe uma constante $c_q \in \mathbb{R}$ tal que $\mu(A) = c_q |A|$, garantindo, assim, a desigualdade desejada

$$\frac{\mathbb{E} [\beta_q^{r,s}(\Phi_A)]}{|A|} \leq c_q.$$

□

Agora, sejam L uma constante real e m um número inteiro qualquer, ambos positivos. Mostremos que a variável aleatória

$$S_m = \sum_{j \in \{1, \dots, m\}^d} \frac{\beta_q^{r,s}(\Phi_{\Lambda_L^j}) - \mathbb{E} [\beta_q^{r,s}(\Phi_{\Lambda_L^j})]}{m^d |\Lambda_L^j|}, \quad (32)$$

é um super-martingale, considerando as mesmas restrições ao processo pontual Φ e sua filtração, como o do Lema acima, e considerando Λ_L^j a translação da janela $[-mL/2, -mL/2 + L]^d$, dada por

$$\Lambda_L^j := \left[-\frac{mL}{2}, -\frac{mL}{2} + L \right)^d + \sum_{i=1}^d (j_i - 1) L e_i,$$

com $j = (j_1, j_2, \dots, j_d) \in \{1, \dots, m\}^d$ e e_i um vetor da base canônica em $\mathbb{R}^d$ (vide a Figura 29).

Lema 3.6.2. *Seja Φ um processo pontual pertencente a família de processos pontuais $\mathcal{C}$ da Seção 3.5.2. Então, para toda filtração simplicial satisfazendo as restrições em 3.2, temos que, para todos reais $r \leq s$ positivos e para todo inteiro $0 \leq q < d$, as variáveis aleatórias*

$$S_m^p := S_{2m} \quad e \quad S_m^i := S_{2m+1},$$

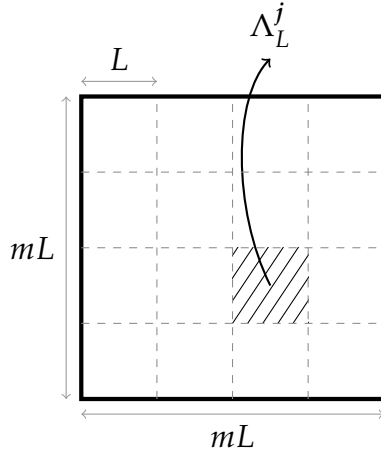


Figura 29: Particionamento da janela $[-\frac{mL}{2}, \frac{mL}{2}]^d$ em m^d janelas disjuntas de volume L^d . Em cada um das partições Λ_L^j , calculamos a razão do número de persistência de Betti pelo volume da janela, para, assim, compor os termos da somatória que define a variável aleatória S_m (equação 32).

com S_m a sequência definida em 32, $L \in \mathbb{R}_{\geq 0}$ uma constante fixa qualquer, e m um inteiro positivo, são super-martingales.

Demonstração. Provemos que a sequência S_m^p , que nada mais é que a sequência S_m indexada com números pares, é um super-martingale. Para a sequência S_m^i , a argumentação é equivalente e, portanto, iremos omitir. Seguindo a definição dada em [17], para provarmos que uma sequência é um super-martingale precisamos checar duas condições:

- (i) $\mathbb{E}[S_m] < \infty$;
- (ii) $\mathbb{E}[S_{m+1} | S_m, \dots, S_1] \leq S_m$, para todo m .

O item (i) é imediato, já que a esperança de S_m^p será sempre zero. Resta provarmos o item (ii). Notemos que a esperança de S_{m+1}^p , dado que temos informação das variáveis $S_1^p, S_2^p, \dots, S_m^p$, equivale a calcular a esperança das caixas com área hachurada da Figura 30, com o retângulo interior a área hachurada equivalendo a informação das variáveis $S_1^p, S_2^p, \dots, S_m^p$.

Denotando por $\bar{\Lambda}$ o conjunto das partições contidas na área hachurada, temos que

$$\mathbb{E}[S_{m+1}^p | S_m^p, \dots, S_1^p] = \sum_{A \in \bar{\Lambda}} \mathbb{E} \left[\frac{\beta_q^{r,s}(\Phi_A) - \mathbb{E}[\beta_q^{r,s}(\Phi_A)]}{(m+1)^d |A|} \right] + \frac{m^d}{(m+1)^d} S_m^p \leq S_m^p,$$

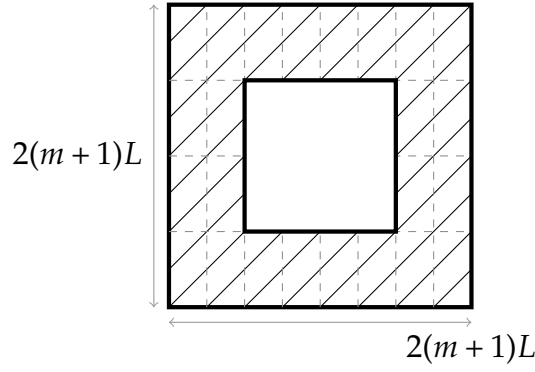


Figura 30: Diferença das partições, quando consideramos índices pares ou ímpares

pois a esperança das caixas na região hachurada é zero. Desta forma, concluímos que S_m^p é uma sequência de super-martingales. $\square$

Resultante deste Lema somos capazes de provar a seguinte proposição:

Proposição 3.6.1. *Seja Φ um processo pontual pertencente a família de processos pontuais $\mathcal{C}$ da Seção 3.5.2. Então, para toda filtração simplicial satisfazendo as restrições em 3.2, temos que, para todos reais $r \leq s$ positivos e para todo inteiro $0 \leq q < d$, a seguinte convergência é verdadeira*

$$\lim_{m \rightarrow \infty} |S_m| \stackrel{q.c.}{=} 0$$

com S_m a sequência definida em 3.2, $L \in \mathbb{R}_{\geq 0}$ uma constante fixa qualquer, e m um inteiro positivo.

Demonstração. Para provarmos esta proposição, avaliaremos a convergência das sub-sequências de índices pares e ímpares:

$$S_m^p := S_{2m} \quad \text{e} \quad S_m^i := S_{2m+1}.$$

Pelo Lema 3.6.2, sabemos que ambas sequências são super-martingales. Além do mais, usando o Lema 3.6.1, temos que

$$\begin{aligned} \mathbb{E}[|S_m|] &\leq \sum_{j \in \{1, \dots, m\}^d} \frac{1}{m^d} \frac{2\mathbb{E}\left[\beta_q^{r,s}(\Phi_{\Lambda_L^j})\right]}{|\Lambda_L^j|} \\ &\leq \sum_{j \in \{1, \dots, m\}^d} \frac{1}{m^d} 2c_q \\ &\leq c_q, \end{aligned}$$

para todo inteiro positivo m . Sendo assim, concluimos que

$$\sup_m \mathbb{E}[|S_m^p|] \leq c_q \quad \text{e} \quad \sup_m \mathbb{E}[|S_m^i|] \leq c_q.$$

Usando o Teorema de Doob para a convergência de super-martingales, vide [17], temos que existem variáveis aleatórias integráveis X^p e X^i tal que

$$\lim_{m \rightarrow \infty} S_m^p \stackrel{q.c.}{=} X^p \quad \text{e} \quad \lim_{m \rightarrow \infty} S_m^i \stackrel{q.c.}{=} X^i.$$

Para encerrarmos a demonstração, provemos que, com probabilidade 1, vale $X^p = X^i = 0$. Para isto, lembremos que decorre da desigualdade de Markov que, para todo real estritamente positivo a , valem as desigualdades

$$\mathbb{P}(|X^p| \geq a) \leq \frac{\mathbb{E}[|X^p|^2]}{a^2} \quad \text{e} \quad \mathbb{P}(|X^i| \geq a) \leq \frac{\mathbb{E}[|X^i|^2]}{a^2}.$$

Provemos, destas desigualdades, que

$$\mathbb{P}(|X^p| \geq a) = 0 \quad \text{e} \quad \mathbb{P}(|X^i| \geq a) = 0,$$

para todo $a > 0$. Mostremos apenas o caso da variável aleatória X^p , já que X^i é feito de forma análoga. Notemos que

$$\mathbb{E}[|X^p|^2] = \mathbb{E}[(X^p)^2] = \lim_{m \rightarrow \infty} \mathbb{E}[S_{2m}^2],$$

pois S_{2m} é majorada por uma constante e converge quase certamente a X^p , além do fato que a função $f(x) = x^2$ é contínua. Com isto, precisamos calcular a última esperança acima, o qual, denotando por $Y_j = \beta_q^{r,s}(\Phi_{\Lambda_L^j}) - \mathbb{E} \left[\beta_q^{r,s}(\Phi_{\Lambda_L^j}) \right]$, vale

$$\begin{aligned} \mathbb{E}[S_m^2] &= \frac{1}{m^{2d} L^{2d}} \mathbb{E} \left[\sum_i Y_i^2 + \sum_{i \neq j} Y_i Y_j \right] \quad (\text{caixas disjuntas são independentes}) \\ &= \frac{1}{m^{2d}} \mathbb{E} \left[\frac{\sum_i Y_i^2}{L^{2d}} \right] \\ &\leq \frac{2c_q^2}{m^d}, \end{aligned}$$

ou seja, $\lim_{m \rightarrow \infty} \mathbb{E}[S_{2m}^2] = 0$. Com isto, concluimos que para qualquer $a > 0$ vale $\mathbb{P}(|X^p| \geq a) = 0$. O mesmo mostra-se para X^i . Consequentemente, concluimos que

$$X^p = 0 \quad \text{e} \quad X^i = 0.$$

Portanto, decorre de $|S_m|$ admitir subsequências, nos índices pares e ímpares, que convergem quase certamente para zero, que concluímos o desejado

$$\lim_{m \rightarrow \infty} |S_m| \stackrel{q.c.}{=} 0$$

□

Finalmente, enunciaremos o último resultado auxiliar antes de provarmos nosso resultado desejado, que equivale ao Lema 3.2 do artigo [2], e que nos garante a convergência quase certa da contagem de simplexes S_q , quando consideramos filtrações aplicadas a processos pontuais ergódicos e estacionários.

Lema 3.6.3. *Seja ζ um processo pontual estacionário. Então, se ζ for um processo ergódico de intensidade unitária, teremos que existem constantes $\hat{S}_q$, para todo $q \in \mathbb{N}_0$, tal que*

$$\lim_{L \rightarrow \infty} \frac{S_q(\phi, r; \Lambda_L^d)}{L^d} \stackrel{q.c.}{=} \lim_{L \rightarrow \infty} \frac{S_q(\phi, r)}{L^d} \stackrel{q.c.}{=} \hat{S}_q.$$

Agora temos as ferramentas necessárias para provar a convergência quase certa da taxa do número de persistência de Betti pelo seu volume:

Teorema 3.1. *Seja Φ um processo pontual pertencente a família de processos pontuais $\mathcal{C}$ da Seção 3.5.2. Então, para toda filtração simplicial satisfazendo as restrições em 3.2, e para todos reais $r \leq s$ positivos, vale que*

$$\lim_{L \rightarrow \infty} \frac{\beta_q^{r,s}(\Phi_{\Lambda_L^d}) - \mathbb{E}[\beta_q^{r,s}(\Phi_{\Lambda_L^d})]}{|\Lambda_L^d|} \stackrel{q.c.}{=} 0, \quad (33)$$

onde $0 \leq q < d$ é a dimensão da homologia e

$$\Lambda_L^d = \left[-\frac{L}{2}, \frac{L}{2} \right)^d$$

a janela de centro na origem o qual calculamos o número de persistência de Betti.

Demonstração. Seja $m > 0$ um inteiro qualquer e $L > 0$ um número real, o procedimento para provarmos a convergência da sequência

$$X(\Lambda_L^d) := \frac{\beta_q^{r,s}(\Phi_{\Lambda_L^d}) - \mathbb{E}[\beta_q^{r,s}(\Phi_{\Lambda_L^d})]}{|\Lambda_L^d|}$$

consistirá em provarmos que $X(\Lambda_L^d)$ pode ser majorada por uma soma finita de termos, que consistem, basicamente, na diferença de propriedades avaliadas tanto na janela Λ_L^d

quanto em janelas menores, resultantes de um particionamento de Λ_L^d . Em seguida, mostraremos que esta majoração, conforme aumentamos o volume da janela Λ_L^d para o infinito, tenderá a zero, i.e., o particionamento da janela, normalizado pelo volume das partições, carregam, no limite, a mesma informação topológica da janela maior.

Consideremos a janela

$$\Lambda_{mL}^d = \left[-\frac{mL}{2}, \frac{mL}{2} \right)^d$$

e seu particionamento em m^d janelas, de volume L^d , que denotaremos por

$$\Lambda_L^j := \left[-\frac{mL}{2}, -\frac{mL}{2} + L \right)^d + \sum_{i=1}^d (j_i - 1)L e_i,$$

com $j = (j_1, j_2, \dots, j_d) \in \{1, \dots, m\}^d$ e e_i um vetor da base canônica em $\mathbb{R}^d$, da mesma forma como ilustrado na Figura 29, valendo

$$\Lambda_{mL}^d = \bigcup_{\tilde{j} \in \{1, 2, \dots, m\}^d} \Lambda_L^{\tilde{j}}.$$

Deste particionamento vale a seguinte desigualdade:

$$\begin{aligned} \left| \frac{\beta_q^{r,s}(\Phi_{\Lambda_{mL}^d}) - \mathbb{E}[\beta_q^{r,s}(\Phi_{\Lambda_{mL}^d})]}{(mL)^d} \right| &\leq \left| \frac{\beta_q^{r,s}(\Phi_{\Lambda_{mL}^d})}{(mL)^d} - \frac{\sum_{j \in \{1, \dots, m\}^d} \beta_q^{r,s}(\Phi_{\Lambda_L^j})}{(mL)^d} \right| \\ &\quad + \left| \frac{\mathbb{E}[\beta_q^{r,s}(\Phi_{\Lambda_{mL}^d})]}{(mL)^d} - \frac{\sum_{j \in \{1, \dots, m\}^d} \mathbb{E}[\beta_q^{r,s}(\Phi_{\Lambda_L^j})]}{(mL)^d} \right| \\ &\quad + \left| \frac{\sum_{j \in \{1, \dots, m\}^d} \mathbb{E}[\beta_q^{r,s}(\Phi_{\Lambda_L^j})]}{(mL)^d} - \frac{\sum_{j \in \{1, \dots, m\}^d} \beta_q^{r,s}(\Phi_{\Lambda_L^j})}{(mL)^d} \right| \end{aligned}$$

Agora devemos inspecionar os três termos da desigualdade acima:

$$\begin{aligned} t_1 &= \left| \frac{\beta_q^{r,s}(\Phi_{\Lambda_{mL}^d})}{(mL)^d} - \frac{\sum_{j \in \{1, \dots, m\}^d} \beta_q^{r,s}(\Phi_{\Lambda_L^j})}{(mL)^d} \right|, \\ t_2 &= \left| \frac{\mathbb{E}[\beta_q^{r,s}(\Phi_{\Lambda_{mL}^d})]}{(mL)^d} - \frac{\sum_{j \in \{1, \dots, m\}^d} \mathbb{E}[\beta_q^{r,s}(\Phi_{\Lambda_L^j})]}{(mL)^d} \right|, \\ t_3 &= \left| \frac{\sum_{j \in \{1, \dots, m\}^d} \mathbb{E}[\beta_q^{r,s}(\Phi_{\Lambda_L^j})]}{(mL)^d} - \frac{\sum_{j \in \{1, \dots, m\}^d} \beta_q^{r,s}(\Phi_{\Lambda_L^j})}{(mL)^d} \right| \end{aligned}$$

Recordemos, da desigualdade 22, que

$$\left| \beta_q^{r,s}(\Phi_{\Lambda_{mL}^d}) - \sum_{j \in \{1, \dots, m\}^d} \beta_q^{r,s}(\Phi_{\Lambda_L^j}) \right| \leq \sum_{j=q, q+1} \text{card} \left(\mathcal{S}_j(\Phi_{\Lambda_{mL}^d}, s) \setminus \bigcup_{\tilde{j} \in \{1, \dots, m\}^d} \mathcal{S}_j(\Phi_{\Lambda_L^{\tilde{j}}}, s) \right),$$

com

$$\mathcal{S}_j(\Phi, s) = \{\sigma \in K_s(\Phi); \dim(\sigma) = j\},$$

e $K_s(\Phi)$ um elemento da filtração. Além do mais, temos que a diferença de simplexos

$$\mathcal{S}_j(\Phi_{\Lambda_{mL}^d}, s) \setminus \bigcup_{\tilde{j} \in \{1, \dots, m\}^d} \mathcal{S}_j(\Phi_{\Lambda_L^{\tilde{j}}}, s)$$

está contida na união de todas as faixas, de comprimento fixo, da fronteira das partições Λ_L^j , i.e., esta diferença na contagem de simplexos será majorada por toda a faixa rosa da Figura 31.

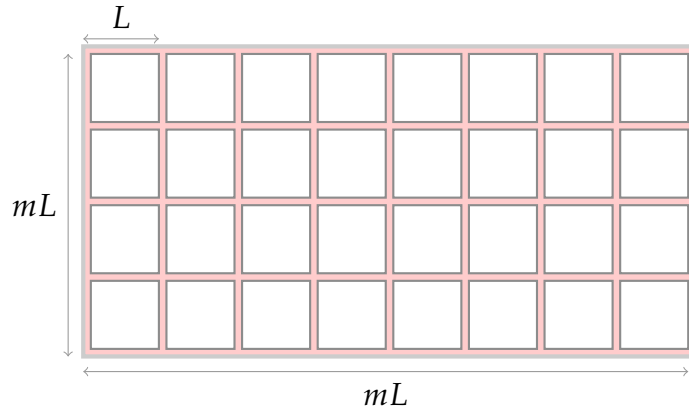


Figura 31: Diferença da contagem de simplexos entre a caixa Λ_{mL}^d e as caixas de seu particionamento, representada na faixa de cor rosa.

Com isto, temos que

$$\text{card} \left(\mathcal{S}_j(\Phi_{\Lambda_{mL}^d}, s) \setminus \bigcup_{\tilde{j} \in \{1, \dots, m\}^d} \mathcal{S}_j(\Phi_{\Lambda_L^{\tilde{j}}}, s) \right) \leq \sum_{\tilde{j} \in \{1, \dots, m\}^d} \mathcal{S}_j \left(\Phi_{\Lambda_{mL}^d}, s; \text{fr}(\Lambda_L^{\tilde{j}}, \rho(s)) \right),$$

com

$$|\text{fr}(\Lambda_L^{\tilde{j}}, \rho(s))| = 2dL^{d-1}\rho(s).$$

Usando que a contagem de simplexes é uma função crescente, e que nosso processo pontual Φ é majorado por um processo pontual de Poisson $\mathcal{P}$ homogêneo e de taxa unitária, obtemos a desigualdade

$$\left| \beta_q^{r,s}(\Phi_{\Lambda_{mL}^d}) - \sum_{j \in \{1, \dots, m\}^d} \beta_q^{r,s}(\Phi_{\Lambda_L^j}) \right| \leq \sum_{j=q, q+1} \sum_{\tilde{j} \in \{1, \dots, m\}^d} S_j \left(\mathcal{P} \cap \Lambda_{mL}^d, s; \text{fr}(\Lambda_L^{\tilde{j}}, \rho(s)) \right). \quad (34)$$

Desta desigualdade podemos tratar os termos t_1 e t_2 . Para o termo t_1 utilizaremos a ergodicidade do processo $\mathcal{P}$, que majora nossa processo pontual. Sendo, tanto a nossa filtração quanto a contagem de simplexes invariantes por translação, temos, pelo Teorema Multidimensional de Ergodicidade de Birkhoff, que

$$\lim_{m \rightarrow \infty} \frac{1}{m^d} \sum_{\tilde{j} \in \{1, \dots, m\}^d} S_j \left(\mathcal{P} \cap \Lambda_{mL}^d, s; \text{fr}(\Lambda_L^{\tilde{j}}, \rho(s)) \right) \stackrel{q.c.}{=} \mathbb{E} \left[S_j \left(\mathcal{P} \cap \Lambda_L^d, s; \text{fr}(\Lambda_L, \rho(s)) \right) \right],$$

para $j = q, q+1$. Utilizando o Lema 3.6.1, obtemos a desigualdade

$$\limsup_{m \rightarrow \infty} t_1 \leq \frac{1}{L^d} (c_q + c_{q+1}) 2dL^{d-1} \rho(s) = \frac{(c_q + c_{q+1}) 2d\rho(s)}{L}, \quad (35)$$

quase certamente. De forma equivalente, provamos uma desigualdade semelhante ao termo t_2 :

$$\limsup_{m \rightarrow \infty} t_2 \leq \frac{1}{L^d} (c_q + c_{q+1}) 2dL^{d-1} \rho(s) = \frac{(c_q + c_{q+1}) 2d\rho(s)}{L}. \quad (36)$$

Finalmente, para o termo t_3 temos garantido pela Proposição 3.6.1 que com, probabilidade 1, vale

$$\limsup_{m \rightarrow \infty} t_3 = 0. \quad (37)$$

Utilizando as desigualdades 35, 36, 37, obtemos, com probabilidade um, que

$$\limsup_{m \rightarrow \infty} \left| \frac{\beta_q^{r,s}(\Phi_{\Lambda_{mL}^d}) - \mathbb{E}[\beta_q^{r,s}(\Phi_{\Lambda_{mL}^d})]}{(mL)^d} \right| \leq \frac{(c_q + c_{q+1}) d(4\rho(s))}{L}. \quad (38)$$

Desta forma, dado $\varepsilon > 0$, fixemos um inteiro $L_0 > 0$ que garanta a desigualdade

$$\frac{(c_q + c_{q+1}) d(4\rho(s))}{L_0} < \varepsilon,$$

fornecendo, para a desigualdade 38, o seguinte

$$\limsup_{m \rightarrow \infty} \left| \frac{\beta_q^{r,s}(\Phi_{\Lambda_{mL_0}^d}) - \mathbb{E}[\beta_q^{r,s}(\Phi_{\Lambda_{mL_0}^d})]}{(mL_0)^d} \right| \leq \varepsilon, \quad (39)$$

com probabilidade 1.

Estamos na parte final para provar o Teorema. Dado $l \in \mathbb{N}$ suficientemente grande, consideremos o inteiro

$$m_l = L_0 \max\{m; mL_0 \leq l\}.$$

Disto, temos a desigualdade

$$m_l \leq l \leq m_l + L_0 \implies \frac{m_l}{l} \leq \frac{1}{l} + 1 \quad \text{e} \quad 1 - \frac{L_0}{l} \leq \frac{m_l}{l}.$$

Em outras palavras, temos

$$\frac{m_l}{l} \rightarrow 1, \text{ quando } l \rightarrow \infty.$$

Como já usado acima, sabemos que

$$\begin{aligned} \left| \beta_q^{r,s}(\Phi_{\Lambda_l^d}) - \beta_q^{r,s}(\Phi_{\Lambda_{m_l}^d}) \right| &\leq \sum_{j=q, q+1} \text{card} \left(\mathcal{S}_j(\Phi_{\Lambda_l^d}, s) \setminus \mathcal{S}_j(\Phi_{\Lambda_{m_l}^d}, s) \right) \\ &\leq \sum_{j=q, q+1} S_j(\mathcal{P} \cap \Lambda_l^d, s; W) \\ &\leq \sum_{j=q, q+1} S_j(\mathcal{P} \cap \Lambda_l^d, s) - S_j(\mathcal{P} \cap \Lambda_{m_l - 2\rho(s)}^d, s), \end{aligned}$$

onde

$$|W| = d2(l - m_l - \rho(s))l^{d-1}.$$

Dividindo tudo por l^d temos

$$\begin{aligned} \left| \frac{\beta_q^{r,s}(\Phi_{\Lambda_l^d})}{l^d} - \frac{\beta_q^{r,s}(\Phi_{\Lambda_{m_l}^d})}{l^d} \right| &\leq \sum_{j=q, q+1} \frac{S_j(\mathcal{P} \cap \Lambda_l^d, s)}{l^d} - \frac{S_j(\mathcal{P} \cap \Lambda_{m_l - 2\rho(s)}^d, s)}{l^d} \\ &\leq \sum_{j=q, q+1} \frac{S_j(\mathcal{P} \cap \Lambda_l^d, s)}{l^d} - \frac{S_j(\mathcal{P} \cap \Lambda_{m_l - 2\rho(s)}^d, s)}{(m_l - 2\rho(s))^d} \frac{(m_l - 2\rho(s))^d}{l^d}. \end{aligned}$$

Pelo Lema 3.6.3, existem as constantes $\hat{S}_{s,q}, \hat{S}_{s,q+1}$ tal que, com probabilidade um, os seguintes limites valem

$$\lim_{l \rightarrow \infty} \frac{S_q(\mathcal{P} \cap \Lambda_l^d, s)}{l^d} = \lim_{l \rightarrow \infty} \frac{S_q(\mathcal{P} \cap \Lambda_{m_l - 2\rho(s)}^d, s)}{(m_l - 2\rho(s))^d} = \hat{S}_{s,q}$$

e

$$\lim_{l \rightarrow \infty} \frac{S_{q+1}(\mathcal{P} \cap \Lambda_l^d, s)}{l^d} = \lim_{l \rightarrow \infty} \frac{S_{q+1}(\mathcal{P} \cap \Lambda_{m_l - 2\rho(s)}^d, s)}{(m_l - 2\rho(s))^d} = \hat{S}_{s,q+1}.$$

Logo, obtemos o importante resultado

$$\lim_{l \rightarrow \infty} \left| \frac{\beta_q^{r,s}(\Phi_{\Lambda_l^d})}{l^d} - \frac{\beta_q^{r,s}(\Phi_{\Lambda_{m_l}^d})}{l^d} \right| = 0, \quad (40)$$

com probabilidade um.

Finalmente, note que

$$\begin{aligned} \left| \frac{\beta_q^{r,s}(\Phi_{\Lambda_l^d}) - \mathbb{E}[\beta_q^{r,s}(\Phi_{\Lambda_l^d})]}{l^d} \right| &\leq \left| \frac{\beta_q^{r,s}(\Phi_{\Lambda_l^d})}{l^d} - \frac{\beta_q^{r,s}(\Phi_{\Lambda_{m_l}^d})}{l^d} \right| \\ &\quad + \left| \frac{\mathbb{E}[\beta_q^{r,s}(\Phi_{\Lambda_l^d})]}{l^d} - \frac{\mathbb{E}[\beta_q^{r,s}(\Phi_{\Lambda_{m_l}^d})]}{l^d} \right| \\ &\quad + \left| \frac{\beta_q^{r,s}(\Phi_{\Lambda_{m_l}^d}) - \mathbb{E}[\beta_q^{r,s}(\Phi_{\Lambda_{m_l}^d})]}{m_l^d} \frac{m_l^d}{l^d} \right| \end{aligned}$$

Utilizando 39 e 40, concluimos que com probabilidade um

$$\limsup_{l \rightarrow \infty} \left| \frac{\beta_q^{r,s}(\Phi_{\Lambda_l^d}) - \mathbb{E}[\beta_q^{r,s}(\Phi_{\Lambda_l^d})]}{l^d} \right| \leq \varepsilon.$$

Como $\varepsilon > 0$ é arbitrário, concluímos o desejado:

$$\lim_{l \rightarrow \infty} \left| \frac{\beta_q^{r,s}(\Phi_{\Lambda_l^d}) - \mathbb{E}[\beta_q^{r,s}(\Phi_{\Lambda_l^d})]}{l^d} \right| \stackrel{q.c.}{=} 0.$$

□

3.7 TEOREMA DO LIMITE CENTRAL

Sejam $(\Omega, \mathcal{A}, \mu)$ um espaço de probabilidade, $K \subseteq \mathbb{R}^d$ um compacto qualquer e $\Phi^{K,\beta}$ o processo pontual marcado

$$\Omega \ni \omega \mapsto \Phi^{K,\beta}(\omega) = \sum_{i \in I} \delta_{(x_i, K)}, \quad (41)$$

com I um conjunto de índices enumerável e $\sum_{i \in I} \delta_{(x_i, K)}$ uma medida própria, tal qual definida em 23. Mostraremos nesta seção que, quando estes processos forem dotados

de estrutura de um processo pontual de Gibbs, como mostrado em 3.5.3, e satisfazendo certas condições, induzirão diagramas de persistência cujos números de persistência de Betti, em uma janela de tempo $[r, s] \subseteq \mathbb{R}_{\geq 0}$, convergirão, em distribuição, à uma normal $\mathcal{N}(0, \bar{\sigma})$, como descrito abaixo:

$$\lim_{L \rightarrow \infty} \frac{\beta_q^{r,s}(\Phi_{\Lambda_L^d}^{K,\beta}) - \mathbb{E} \left[\beta_q^{r,s}(\Phi_{\Lambda_L^d}^{K,\beta}) \right]}{L^{d/2}} \stackrel{D}{=} \mathcal{N}(0, \bar{\sigma}),$$

onde $\Phi^{K,\beta}$ é um processo pontual de Gibbs, com medida de probabilidade absolutamente contínua a uma processo pontual, homogêneo e de taxa unitária, ou seja, satisfazendo

$$\mu_{\Lambda_L^d}(\mathrm{d}\mathbf{X}) = \frac{\exp(-\beta \Psi(\mathbf{X} \cap \Lambda_L^d))}{Z(\beta \Psi_\Lambda)} \mu_\Lambda^0(\mathrm{d}\mathbf{X}), \quad (42)$$

com $\mathbf{X}$ um subconjunto finito de $\mathbb{R}^d$, $Z(\beta \Psi_\Lambda)$ uma constante normalizadora e Ψ uma função de energia.

Notação. Denotemos por $\mathcal{C}_{\text{inter}}$ a família dos processos pontuais gibbsianos satisfazendo 41 e 42.

Para provarmos a convergência em distribuição acima, utilizaremos como ferramenta o Teorema 2.3 do paper [3], que exige, dentre suas hipóteses, que exista uma função escore

$$\xi : \mathbb{R}^d \times \mathcal{C} \rightarrow \mathbb{R}, \quad \text{com} \quad \mathcal{C} = \{A \subseteq \mathbb{R}^d; A \text{ é localmente finito} \},$$

tal que possamos escrever o número de persistência de Betti $\beta_q^{r,s}(\Phi_{\Lambda_L^d}^{K,\beta})$ em função deste escore

$$\beta_q^{r,s}(\Phi_{\Lambda_L^d}^{K,\beta}) = \sum_{x \in \Phi^{K,\beta} \cap \Lambda_L^d} \xi(x, \Phi^{K,\beta} \cap \Lambda_L^d). \quad (43)$$

Desta seção, em diante, todos nossos resultados serão restritos somente à filtrações simpliciais de Vietoris-Rips, visto que estas nos possibilitarão definir funções escores tais que a igualdade em 43 seja satisfeita, como veremos a seguir. Desta restrição à filtração $\mathcal{F} = (\mathbb{K}_i)_{i \in \mathbb{R}_{\geq 0}}$, recordemos que

$$x, y \in \sigma \in \mathbb{K}_r \iff B(x; r) \cap B(y; r) \neq \emptyset \iff d(x, y) \leq 2r.$$

Recordemos, também, que o número de persistência de Betti

$$\beta_q^{r,s}(\Phi_{\Lambda_L^d}^{K,\beta}), \quad \text{com} \quad \Lambda_L^d = \left[-\frac{L}{2}, \frac{L}{2} \right)^d, \quad L \in \mathbb{R}_{>0} \text{ e } q \in \mathbb{N}_0,$$

pode ser escrito como uma medida de contagem de pontos persistentes no boreliano $(-\infty, r] \times [s, \infty) \subseteq \mathbb{R}^2$, i.e., $\beta_q^{r,s}$ é, nada mais que, a soma de medidas pontuais de Dirac

$$\beta_q^{r,s}(\Phi_{\Lambda_L^d}^{K,\beta}) = \sum_{(b,d) \in \text{Dgm}_q(\zeta \cap \Lambda_L^d)} m_{(b,d)} \delta_{(b,d)}((-\infty, r] \times [s, \infty)),$$

com $m_{(b,d)}$ o número de vezes em que encontramos o ponto (b,d) , no diagrama de persistência $\text{Dgm}_q(\zeta \cap \Lambda_L^d)$, e $\delta_{(b,d)}$ a medida de Dirac em $\mathbb{R}^2$:

$$\delta_{(b,d)}(B) = \begin{cases} 1 & \text{se } (b,d) \in B \\ 0 & \text{c.c.} \end{cases}$$

Com isto, já podemos estabelecer as funções scores necessárias para obtermos 43, o qual separaremos em dois casos: funções escore para dimensão 0; e funções escore para dimensão $q \geq 1$.

Seja $A \subseteq \mathbb{R}^d$ um conjunto localmente finito. Sabemos que componentes conexas sempre nascem no instante de tempo zero e, conseqüentemente, a única propriedade persistente que temos, para caracterizar, é o instante de tempo em que componentes conexas somem do espaço simplicial. Dado um ponto $x \in A$, e definindo por

$$C_r(x, A) := \{x_0, x_1, \dots, x_R\} \subseteq A \text{ tal que } d(x_i, x_{i+1}) \leq 2r, \text{ para todo } i \in \{0, 1, \dots, R\} \text{ e } x \in \{x_0, x_1, \dots, x_R\}.$$

sua componente conexa, então, $C_r(x, A)$ sumirá da filtração simplicial, se ela se unir a outra componente conexa disjunta. Desta forma, quando o conjunto A for finito, temos que a componente $C_r(x, A)$ desaparecerá da filtração quando seu elemento maximal, no instante r , for menor que o observado no instante s , com $r < s$. Sendo assim, definindo

$$\text{death}(x, A) = \sup\{r \geq 0; \max C_r(x, A \cap B[x; r]) = x\},$$

onde consideramos $\mathbb{R}^d$ munido da ordem lexicográfica, como ilustrado na Figura 32 temos que o número de persistência de Betti, para dimensão zero, será igual a

$$\beta_0^{r,s} = \sum_{x \in \zeta \cap \Lambda_L^d} \delta_{(0, \text{death}(x, \zeta \cap \Lambda_L^d))}([0, r] \times [s, \infty))$$

e, portanto, concluímos que a função escore para dimensão 0 será dada por

$$\begin{aligned} \xi_0 : \mathbb{R} \times \{A \subseteq \mathbb{R}; A \text{ é localmente finito}\} &\rightarrow \overline{\mathbb{R}} \\ (x, A) &\mapsto \delta_{(0, \text{death}(x, A))}([0, r] \times [s, \infty)). \end{aligned} \tag{44}$$

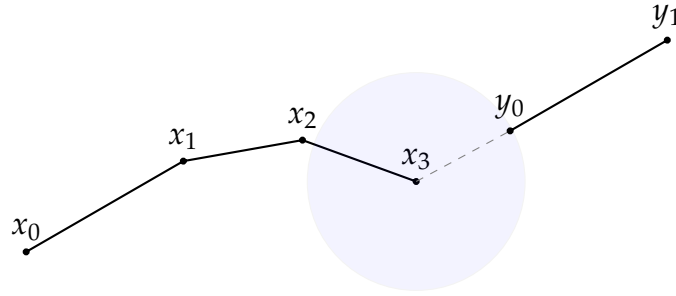


Figura 32: Ilustração da componente conexa $C_r(x_3, A)$ e do tempo, em que a mesma, some da filtração para se juntar a outra componente conexa $C_r(y_1, A)$.

Agora definiremos a função escore para dimensão $q \geq 1$, que, de fato, seguirá o mesmo caminho exposto acima, com apenas algumas adaptações. Sejam $x \in \Phi^{K,\beta}$ um ponto qualquer e $A \subseteq \mathbb{R}^d$ um conjunto localmente finito. Nossa função escore, para dimensão $q \geq 1$, será responsável em contar todos os buracos no espaço que surgiram, graças ao nascimento de algum simplexo de dimensão q , contendo algum dos ponto do processo pontual. Assim, como o número de persistência de Betti leva apenas em conta a quantidade de propriedades topológicas contidas no intervalo $[-\infty, r] \times [s, +\infty]$, com $0 < r \leq s$, temos a restrição natural, de que os simplexos responsáveis por criarem buracos devem aparecer, na filtração simplicial, em um tempo menor, ou igual, ao tempo r .

Notemos que o conjunto

$$A \cap B[x; r]$$

é finito e, portanto, a quantidade de q -simplexos neste conjunto é, também, finita. Agora, tomando o conjunto de índices

$$T_x = \{t \in \mathbb{R} \cap (0, r]; \text{ nasce no tempo } t \text{ um simplexo } \sigma, \text{ com } \dim(\sigma) = q, q-1 \text{ e } x \in \sigma\},$$

que é finito (pois, no pior caso, teremos indexado a quantidade máxima das q combinações dos elementos do conjunto acima) e considerando a filtração simplicial indexada por T_x ,

$$\mathbb{K}(A)_{T_x},$$

teremos que os diagrama de persistência, denotados simplesmente por $\text{Dgm}_q^{T_x}(A)$, estarão bem definidos. (Os diagramas de persistência estão bem definidos graças ao conjunto de índices ser finito, *Teorema 2.2 (Aplicação do Teorema de Gabriel, Parte (b))*.)

Em posse destas estruturas, agora somos capazes de estudar os buracos de dimensão q , influenciados pelo ponto x . Seja

$$(b, d) \in \text{Dgm}_q^{T_x}(A)$$

um ponto qualquer do diagrama de persistência e $H_q^{(b,d)}$ um buraco topológico de dimensão q , que nasceu no instante b e morreu no instante d . Dizemos que o ponto x é responsável pelo nascimento de $H_q^{(b,d)}$ se, no instante b , surgir um q -simplexo $\sigma = \{x_0, \dots, x_q\}$, com $x \in \sigma$, criando o buraco $H_q^{(b,d)}$. (Antes do instante b , o buraco $H_q^{(b,d)}$ não existia no espaço.) Como estes simplexos não serão obrigatoriamente únicos, definimos a função auxiliar

$$\underline{\text{birth}}\left(H_q^{(b,d)}\right) = \min \left\{ x; x \in \sigma \text{ e } \sigma \text{ é responsável pelo surgimento de } H_q^{(b,d)} \right\},$$

onde o mínimo é tomado considerando a ordem lexicográfica em $\mathbb{R}^d$ (vide a Figura 33). Para terminarmos, construiremos a função que calcula o tempo em que o buraco $H_q^{(b,d)}$ torna-se bordo. De fato, sabemos que este valor será d , entretanto, para deixarmos claro a dinâmica de comportamento da função escore, usaremos a função que calcula o tempo em que o buraco some da filtração

$$\underline{\text{death}}\left(H_q^{(b,d)}\right) = \inf \left\{ t \geq b; \mathbb{R}^d \setminus \bigcup_{\substack{z \in \sigma \\ \sigma \in H_q^{(b,d)}}} B[z; t] \text{ tem apenas uma comp. conexa,} \right\}$$

cujas intuições apresentamos, também, na Figura 33. Finalmente, teremos a seguinte função escore,

$$\xi_q : \mathbb{R} \times \{A \subseteq \mathbb{R}; A \text{ é localmente finito}\} \rightarrow \overline{\mathbb{R}},$$

valendo

$$\xi_q(x, A) = \sum_{(b,d) \in \text{Dgm}_q^{T_x}(A)} \mathbb{1}_{\{x\}} \left(\underline{\text{birth}}(H_q^{(b,d)}) \right) \mathbb{1}_{[s,\infty]} \left(\underline{\text{death}}(H_q^{(b,d)}) \right). \quad (45)$$

Agora, que já somos capazes de escrever o número de persistência de Betti em termos das funções escores 44 e 45, precisaremos estabelecer alguns critérios sobre estas funções, para que possamos usar o TLC apresentado em [3]. São eles:

Definição 3.7.1 (Raio de estabilização). Seja ξ uma função escore (tal qual a definida acima) e Ξ um Processo Pontual tipo Poisson. Então, dizemos que $R := R^\xi(x, \mathcal{X}^z) \in \mathbb{R}_{>0}$,

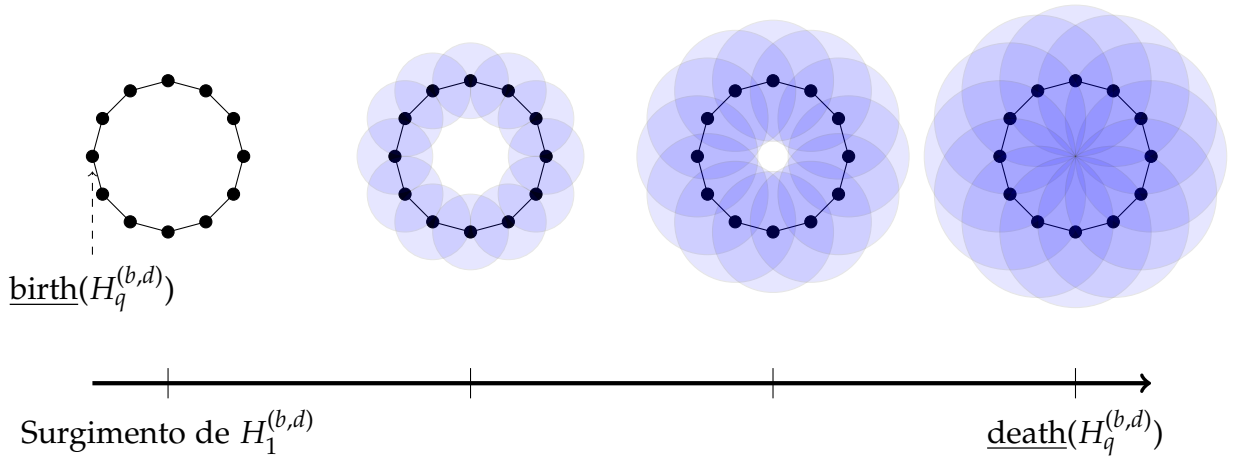


Figura 33: Nascimento e morte de buracos de dimensão $q \geq 1$.

com $\mathcal{X}^z := \mathcal{X} \cup \{z\}$, é um *raio de estabilização* para o ponto $x \in \mathbb{R}^d$ e para o ponto $z \in \mathbb{R}^d \cup \{\emptyset\}$ se, para quase certamente toda realização $\mathcal{X}$ de Ξ , a igualdade abaixo for verdadeira

$$\xi(x, \mathcal{X}^z \cap B(x, R)) = \xi(x, (\mathcal{X}^z \cap B(x, R)) \cup \mathcal{A}),$$

para todo $\mathcal{A} \subseteq \mathbb{R}^d \setminus B(x, R)$ conjunto localmente finito.

Definição 3.7.2 (Função escore exponencialmente estabilizadora). Seja ξ uma função escore e Ξ um Processo Pontual tipo Poisson. Então, dada o valor

$$t(w; \varepsilon) := \sup_{y \in \mathbb{R}^d} \sup_{z \in \mathbb{R}^d \cup \{\emptyset\}} \mathbb{P} \left(\sup_{x \in B(y, \varepsilon) \cap (\Xi \cup \{z\})} R^\xi(x, \Xi \cup \{z\}) > w \right),$$

com $s, \varepsilon > 0$, dizemos que ξ é exponencialmente estabilizadora se

$$\limsup_{\varepsilon \rightarrow 0} \limsup_{w \rightarrow \infty} \frac{\log t(w, \varepsilon)}{w} < 0.$$

Definição 3.7.3 (p-momento). Dado o potencial $\Psi \in \Psi^*$ e $p \in [0, \infty)$. Então, uma função score ξ satisfaz o *p-momento* se, para todo par $(\tau, \beta) \in \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$, tivermos

$$\sup_{L \in [1, \infty]} \sup_{u \in \Lambda_L^d} \sup_{\mathcal{Y} \in \mathcal{C}} \mathbb{E} |\xi(u, (\mathcal{P}^{\beta\Psi} \cap \Lambda_L^d) \cup \mathcal{Y})|^p < \infty,$$

onde $\mathcal{P}^{\beta\Psi}$ é um processo de Gibbs.

Finalmente, já estamos em condição de provar o Teorema do Limite Central, para o número de persistência de Betti dos processos gibbsianos $\Phi^{K, \beta} \in \mathcal{C}_{\text{inter}}$. Entretanto, antes

de prosseguirmos, há um ponto importante que devemos observar: as funções de escore definidas em 45, para dimensões maiores que 1, não admitem raio de estabilização.

Para ilustrarmos o fato sobre o raio de estabilização, olhemos para o caso exposto na Figura 34. Neste exemplo, suponhamos que os pontos em preto, de cada reta, estejam distantes entre si de 1 unidade, e a distância entre pontos de retas distintas sejam maiores que 2 unidades. Sendo assim, para o intervalo de tempo $[0, 1]$, concluímos que nossa função escore satisfaz

$$\xi_1(x_0, \mathcal{X}) = 0,$$

onde $\mathcal{X}$ é o conjunto dos pontos pretos. Portanto, qualquer que seja a bola aberta $B(x_0; r)$, sempre poderemos criar um caminho de pontos vermelhos (conjunto A), tal qual na Figura 34, com distância menor que um, onde somos capazes de induzir um buraco na filtração e, portanto,

$$\xi_1(x_0, (\mathcal{X} \cap B(x_0; r)) \cup A) = 1.$$

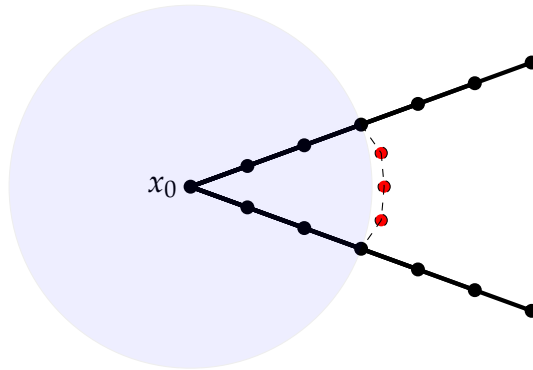


Figura 34: Motivação do porquê trabalharmos com propriedades topológicas limitadas, para dimensões maiores que zero.

Para lidarmos com o problema exposto acima, introduziremos o conceito de *propriedades topológicas de diâmetro limitado*. Seja $H_q^{(a,b)}$ um buraco de dimensão q , que nasce no instante a e morre no tempo b , em uma dada filtração. Dizemos que $H_q^{(a,b)}$ tem diâmetro finito, $\text{diam}(H_q^{(a,b)})$, se

$$\text{diam} \left(H_q^{(a,b)} \right) = \max \left\{ |x - y|; x, y \text{ pertencem a } q\text{-simplexos } \sigma_x, \sigma_y \in H_q^{(a,b)} \right\} < +\infty.$$

e, como consequência deste valor, estabelecemos o conceito de número de persistência de Betti de características topológicas M -limitadas:

Definição 3.7.4. Dado uma constante real positiva $M > 0$, um inteiro positivo $q \geq 0$ e $\mathcal{X} \subseteq \mathbb{R}^d$ um conjunto que admite diagrama de persistência

$$\text{Dgm}_q(\mathcal{X}),$$

para uma filtração $\mathbb{K}$ qualquer. Chamaremos de número de persistência de Betti, de características topológicas M -limitadas, e que denotaremos como $\beta_q^{M,r,s}(\mathbb{K}(\mathcal{X}))$, o valor

$$\beta_q^{M,r,s}(\mathbb{K}(\mathcal{X})) = \sum_{(b,d) \in \text{Dgm}_q(\mathcal{X}) \text{ e } \text{diam}(H_q^{(a,b)}) \leq M} \delta_{(b,d)}([0, r] \times [s, \infty)).$$

(Lembrando que, quando não houver risco de confusão com a filtração, denotaremos o número de persistência por apenas $\beta_q^{M,r,s}(\mathcal{X})$.)

Em posse destas estruturas, e desta última observação, somos capazes de apresentar o Teorema do Limite Central:

Teorema 3.2 (TLC). *Seja $\Phi^{K,\beta} \in \mathcal{C}_{\text{inter}}$ um processo gibbsiano, satisfazendo as condições em 41 e 42, e que admita simulação perfeita. Então, se existir uma constante H_0 , tal que seu potencial de energia Ψ satisfaça a desigualdade $\Psi(\mathcal{X}) > H_0$, teremos que para as funções escore definidas em 44 e 45, e para todo intervalo de tempo $[r, s] \subseteq \mathbb{R}_{\geq 0}$, existirão as constantes $\hat{\sigma}_q \geq 0$, com $q \in \mathbb{N}_0$, tal que as seguintes convergências em distribuição serão verdadeiras:*

$$\lim_{L \rightarrow \infty} \frac{\beta_0^{r,s}(\Phi_{\Lambda_L^d}^{K,\beta}) - \mathbb{E} \left[\beta_0^{r,s}(\Phi_{\Lambda_L^d}^{K,\beta}) \right]}{L^{d/2}} \stackrel{D}{=} \mathcal{N}(0, \hat{\sigma}_0), \quad (46)$$

e

$$\lim_{L \rightarrow \infty} \frac{\beta_q^{M,r,s}(\Phi_{\Lambda_L^d}^{K,\beta}) - \mathbb{E} \left[\beta_q^{M,r,s}(\Phi_{\Lambda_L^d}^{K,\beta}) \right]}{L^{d/2}} \stackrel{D}{=} \mathcal{N}(0, \hat{\sigma}_q), \quad (47)$$

para um M fixado a priori e $q > 0$.

Demonstração. Sendo $\Phi^{K,\beta} \in \mathcal{C}_{\text{inter}}$ um processo pontual gibbsiano, que admite simulação perfeita, então, pelo Lema 3.3 de [3], temos garantido que este processo é, também, do tipo quase de Poisson. Resta avaliarmos as seguintes condições: (i) ξ é exponencialmente estabilizadora; (ii) ξ satisfaz o p -momento, para $p > 2$. Começamos pelo caso de dimensão zero, já que para dimensão maiores será mais fácil, devido ao fato de estarmos trabalhando com propriedades topológicas limitadas quando $q > 0$. Sendo assim, para o caso de dimensão zero, começemos observando que para todo ponto $x \in \Phi^{K,\beta}$ existirá,

quase certamente, um ponto $x_0 \in x + (0, \infty)^d$, e, portanto, $R^{\tilde{\zeta}}(x, \Phi^{K, \beta} \cup \{z\})$ está bem definida. Além do mais, dado $w > s$, vale que

$$[R^{\tilde{\zeta}}(x, \Phi^{K, \beta} \cup \{z\}) > w] \iff \Phi^{K, \beta}(x + (0, w)^d) = 0.$$

Como, por hipótese, temos que a função de energia satisfaz $\Psi(\mathcal{X}) \geq H_0 > -\infty$, obtemos que

$$\mathbb{P} \left[\Phi^{K, \beta}(x + (0, w)^d) = 0 \right] = \frac{e^{-\beta H_0}}{e^{-\beta \Psi(\emptyset)}} e^{-\tau w^d}.$$

Como isto não depende de mais nenhum parâmetro, temos que

$$t(w; \varepsilon) = \frac{e^{-\beta H_0}}{e^{-\beta \Psi(\emptyset)}} e^{-\tau w^d},$$

onde $t(w; \varepsilon)$ está definida em 3.7.2. Logo, por $\Psi(\emptyset) < \infty$, pela construção dos processos gibbsianos, concluímos que

$$\limsup_{\varepsilon \rightarrow 0} \limsup_{w \rightarrow \infty} \frac{\log t(w, \varepsilon)}{w} = \limsup_{w \rightarrow \infty} \left(-\tau w^{d-1} + \frac{\log(e^{-\beta H_0}) - \log(e^{-\beta \Psi(\emptyset)})}{w} \right) = -\infty < 0.$$

Portanto, provamos que para dimensão 0 a função escore $\tilde{\zeta}_0$, para o nosso processo pontual, é exponencialmente estabilizadora. Para encerrarmos o caso de dimensão 0, mostremos que $\tilde{\zeta}_0$ satisfaz o p -momento (vide 3.7.3), para $p = 3$. Esta parte é imediata, visto que a $\tilde{\zeta}_0$ vale apenas zero ou um. Desta maneira, provamos todas as condições exigidas no Teorema 2.3 do paper [3] e, portanto, temos que existe uma constante $\hat{\sigma}_0$ tal que

$$\lim_{L \rightarrow \infty} \frac{\beta_0^{r,s}(\Phi_{\Lambda_L^d}^{K, \beta}) - \mathbb{E} \left[\beta_q^{r,s}(\Phi_{\Lambda_L^d}^{K, \beta}) \right]}{L^{d/2}} \stackrel{D}{=} \mathcal{N}(0, \hat{\sigma}_0).$$

Agora avaliemos o caso para $q > 0$. Começemos checando que $\tilde{\zeta}_q$ é exponencialmente estabilizadora. Esta propriedade decorre imediatamente do fato de estarmos trabalhando com propriedades topológicas limitadas e de diâmetro, de no máximo, M unidades. Com isto, para distâncias maiores que M , de um ponto qualquer x , não temos a criação de novos ciclos, criados por x . Portanto,

$$[R^{\tilde{\zeta}}(x, \Phi^{K, \beta} \cup \{z\}) > M] = \emptyset,$$

valendo

$$t(w; \varepsilon) = 0,$$

para todo $w > M$. Disto concluímos que

$$\limsup_{\varepsilon \rightarrow 0} \limsup_{w \rightarrow \infty} \frac{\log t(w, \varepsilon)}{w} = \log(0) = -\infty < 0,$$

adotando a convenção que $\log(0) = -\infty$. Agora que já somos capazes de provar que ξ_q é exponencialmente estabilizadora, para propriedades topológicas M -limitadas, provemos que vale, também, a propriedade do p -momento. Notemos que a propriedade de tomarmos o número de persistência de Betti, para características topológicas M -limitadas, garante que

$$\xi_q(u, \Phi^{K, \beta} \cap \Lambda_L) \leq \sum_{j=q-1, q} \binom{\Phi^{K, \beta}(B[u; M])}{j} \leq \sum_{j=q-1, q} \left(\Phi^{K, \beta}(B[u; M]) \right)^j$$

e, portanto, satisfaz a condição do p -momento, já que M é fixo e $\Phi^{K, \beta}$ admite todos os momentos finitos, na janela $B[u; M]$. Portanto, concluímos a última convergência em distribuição do teorema:

$$\lim_{L \rightarrow \infty} \frac{\beta_q^{M, r, s}(\Phi_{\Lambda_L^d}^{K, \beta}) - \mathbb{E} \left[\beta_q^{M, r, s}(\Phi_{\Lambda_L^d}^{K, \beta}) \right]}{L^{d/2}} \stackrel{D}{=} \mathcal{N}(0, \hat{\sigma}_q), \quad (48)$$

para um M fixado a priori, $q > 0$ e $\hat{\sigma}_q$ uma constante. $\square$

Como exemplo, vamos aplicar o resultado do TLC para um caso específico de processo de interação de área.

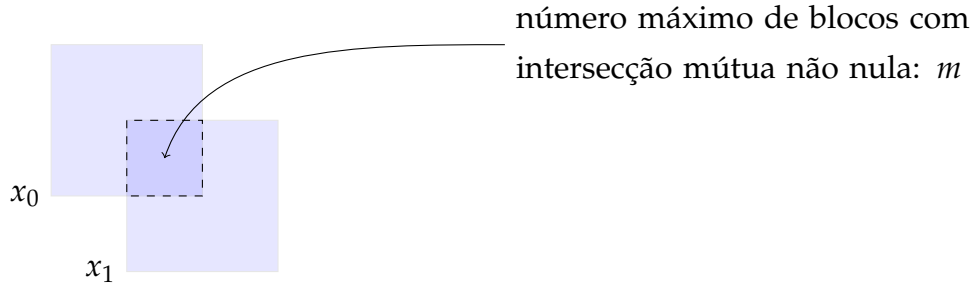


Figura 35: Processo de interação de área, com restrição de pontos em uma vizinhança compacta K .

Proposição 3.7.1. *Seja $K \subseteq \mathbb{R}^d$ um compacto qualquer, $m \in \mathbb{N}$ um número natural qualquer e $\Phi^{K, \beta} \mathcal{C}_{inter}$ um processo de interação de área, gibbsiano, com função de energia*

$$\Psi_{\Lambda}(N) = \begin{cases} \infty, & \text{se } \max_{\substack{H \subseteq N \cap \Lambda; \\ \text{card}(H) \geq C}} \left\{ \mathbb{1} \left(\bigcap_{x \in H} x + K \neq \emptyset \right) \right\} = 1; \\ 0, & \text{caso contrário,} \end{cases}$$

como ilustrado na Figura 35. Então, existe um parâmetro τ , do processo de Poisson que gera $\Phi^{K,\beta}$, tal que, dado o intervalo não degenerado $[r, s] \subseteq \mathbb{R}_{\geq 0}$, temos que a razão do número de persistência de Betti $\beta_0^{r,s}(\Phi^{K,\beta})/|\Lambda_L|$, de dimensão 0, converge em distribuição para uma normal. Além do mais, para propriedades topológicas M -limitadas, temos que $\beta_0^{M,r,s}(\Phi^{K,\beta})/|\Lambda_L|$ converge, também, para uma normal.

Demonstração. Para provarmos as convergências em distribuição, devemos checar as seguintes hipóteses: (i) $\beta_0^{r,s}(\Phi^{K,\beta})$ deve ter alcance de interação finito; (ii) a função de potencial deve ser limitada inferiormente; (iii) deve existir uma constante m_0 tal que

$$\tau \ell(B(0; 1)) e^{-\beta m_0} (r_0 + 1)^d < 1,$$

com τ a taxa do processo pontual de Poisson $\mathcal{P}_\tau$, a ser escolhida, tal que $\Phi^{K,\beta}$ admita esquema de simulação perfeita. Para provarmos a condição (i), devemos checar todos os casos possíveis para a função 29. Suponhamos que $\mathcal{X}$ seja tal que $\text{card}(\mathcal{X}) = m$ e

$$\bigcap_{x \in \mathcal{X}} (x + K) \neq \emptyset,$$

então temos dois casos possíveis

$$\bigcap_{x \in \mathcal{X} \cup \{0\}} (x + K) \neq \emptyset \quad \text{ou} \quad \bigcap_{x \in \mathcal{X} \cup \{0\}} (x + K) = \emptyset.$$

Nestes casos, teremos que

$$\bigcap_{x \in \mathcal{X} \cup \{0\}} (x + K) \neq \emptyset \implies \Delta(0, \mathcal{X}) = \infty$$

e

$$\bigcap_{x \in \mathcal{X} \cup \{0\}} (x + K) = \emptyset \implies \Delta(0, \mathcal{X}) = 0.$$

Portanto, só precisamos nos preocupar com o raio de estabilização para o caso em que $\Delta(0, \mathcal{X}) = \infty$. Todavia, achar o raio de estabilização para este caso é fácil, basta tomarmos

$$r_0 = 2 \text{diam}(K).$$

Com isto, usando o raio acima, temos que a igualdade 28 é satisfeita e, portanto, a função de potencial tem alcance de interação finito. A condição (ii) é imediata. Resta, finalmente, provarmos hipótese (iii). Este item resulta do Lema 3.5.1, pois como nosso processo pontual tem alcance finito, se tomarmos

$$m_0 = \inf_{\mathcal{X} \text{ localmente finito}} \Delta(0, \mathcal{X}) = 0$$

temos garantido, para todo parâmetro τ satisfazendo

$$\tau \ell(B(0;1))e^{-\beta m_0}(r_0+1)^d < 1,$$

que o processo pontual gibbsiano $\Phi^{K,\beta}$ admite esquema de simulação perfeita. Portanto, concluímos que as hipóteses do Teorema 3.2 são satisfeitas, garantido as convergências em distribuição para uma normal, dos números $\beta_0^{r,s}(\Phi^{K,\beta})/|\Lambda_L|$ e $\beta_0^{M,r,s}(\Phi^{K,\beta})/|\Lambda_L|$. $\square$

4

DETECTANDO NOTÍCIAS FALSAS NO TWITTER

Ao longo desses anos vivenciamos um fenômeno que não para de ganhar força em nossa sociedade, destacando-se principalmente nos meios virtuais, que é a propagação desenfreada de *mensagens falsas* (que muitas vezes nos referiremos pelo seu termo em inglês: *fake news*). O que no começo era visto como algo inofensivo, mostrou-se, com o decorrer do tempo, ser extremamente nocivo, tendo impacto nos mais diversos setores de nossa sociedade: como na saúde, nas eleições, no meio ambiente, etc. Motivados por estes impactos, que as fake news podem exercer em nossa sociedade, decidimos elaborar uma metodologia de análise que, usando as ferramentas de *Homologia Persistente*, seja capaz de avaliar os grafos de interação dos indivíduos que podem, ou não, estar sujeitos a propagação de mensagens falsas em uma rede social, o *Twitter* ¹.

Quando concentramos nossas atenções às redes sociais, os efeitos das fake news tornam-se ainda mais evidentes, seguindo um padrão já conhecido onde usuários mal intencionados, querendo espalhar estas mensagens falsas, contratam empresas especializadas que, por meio de *robôs*, forçam interações em larga escala, (e cujo tópico central da conversa seja a mentira do usuário mal intencionado), induzindo os algoritmos das redes sociais a concluírem que estes tópicos discutidos em massa são, de fato, relevantes ao seus usuários e os indicando sua leitura.

Uma forma para mitigar os efeitos das fake news, comumente aplicada no mercado, é a aplicação de modelos de *Processamento de Linguagem Natural* para a classificação destas mensagens, as classificando como falsas ou não, que, posteriormente, servirão como base para as devidas medidas preventivas a serem tomadas. Vale ainda ressaltar que todas estas técnicas utilizam-se de classificadores muito parecidos, que, de forma simplificada, nada mais são do que aproximações de modelos probabilísticos às palavras usadas, avaliando as *chances* destas palavras, dentro de certos contextos (que também são modelados e estimados suas chances), estarem ligadas a um determinado

¹ A escolha pela rede social Twitter dá-se por uma série de fatores, o qual podemos destacar o fato dela ter seu código aberto e propiciar um ambiente favorável para pesquisadores coletarem dados nela.

tópico. Atualmente, diversos modelos de *Processamento de Linguagem Natural* vem se destacando, muito devido ao recente sucesso da ferramenta *ChatGPT*².

Note que, por mais consolidado que estejam as técnicas de processamento de linguagem, todos seus modelos possuem a mesma estrutura em comum: um espaço amostral formado único, e exclusivamente, por uma vasta e abundante coleção de *documentos*³. Todavia, como neste texto trataremos apenas de uma rede social em específico, o *Twitter*, o espaço amostral não será composto por apenas documentos em isolado, mas sim por três componentes importantes:

- (i) As mensagens (*tweets* que formarão a coleção de documentos);
- (ii) Os usuários que geram estas mensagens;
- (iii) A interação dos usuários entre si, por meio das mensagens que eles produzem.

Como já ressaltado no parágrafo acima, destes três itens, apenas os dois primeiros são tratados extensivamente, com o primeiro item sendo notoriamente abordado pelas técnicas de Processamento de Linguagem Natural. Entretanto, vale ressaltar que o item (iii) é raramente tratado, muito devido a sua complexidade de estrutura, e a seu pouco apelo intuitivo, quando comparado aos itens (i) e (ii).

Com base nos pontos levantados acima, nossa contribuição com este trabalho será, além de tratar um ponto de alta relevância a nossa sociedade, será de propor o uso de novas técnicas de análise, que venham a preencher a lacuna decorrente do item (iii) acima, pouco estudada pelo público em geral.

De fato, neste texto são propostas duas metodologias de detecção de fake news: uma usando a *Homologia Persistente* e a outra usando a ferramenta de análise *Botômetro*⁴ (o termo Botômetro será um aportuguesamento do termo em inglês, *botometer*). Ambas metodologias vem a levar em conta os três itens citados acima, com a Homologia Persistente classificando os grafos de interação dos usuários em dois clusters distintos, levando em consideração apenas as estruturas topológicas dos grafos, e o Botômetro classificando os grafos em dois clusters também, porém com o Botômetro os classificando em função dos perfis dos usuários que os compoem, o que se encaixa nos itens (i) e (ii).

² <https://openai.com/blog/chatgpt>

³ Os documentos podem ser qualquer coisa, como o texto de um blog, uma página html, uma entrada de uma página da wikipédia, etc.

⁴ <https://botometer.osome.iu.edu/>

Destes estudos, somos capazes de inferir que ambas metodologias tem uma similaridade de algo em torno de 70%, ou seja, avaliar a estrutura topológica nos fornece resultados equiparáveis a técnicas já consolidadas, que avaliam o perfil dos usuários, em função do histórico de *tweets* publicado por eles. Além do mais, veremos que os métodos de clusterização, resultantes dos dados gerados pela Homologia Persistente, são mais assertivos em suas classificações do que a clusterização obtida por meio dos dados do Botômetro.

Abaixo listamos nossa metodologia de estudo, onde cada item apresenta uma descrição resumida de cada passo do estudo. Cada uma dessas etapas podem ser vistas no esquema visual da Figura 36.

ETAPA 1 : Coleta de *tweets* para criação do banco de dados. Convém notar que este processo de coleta de dados será baseado, fortemente, em um conjunto de palavras chaves, escolhidas convenientemente, tal que possamos usá-las no processo de busca de tweets, no serviço do Twitter;

ETAPA 2 - HOMOLOGIA : Transformação dos dados brutos, coletados anteriormente, em uma coleção de grafos ordenados. O objetivo desta etapa consiste em capturar as interações dos usuários entre si, sendo que, por interações entre usuários, entenderemos como o ato de um usuário u responder, ou retuitar, ou citar outro usuário;

ETAPA 3 - HOMOLOGIA : A partir destes grafos orientados, e com peso, aplicaremos a Homologia Persistente de Caminhos, gerando informação por meio dos diagramas de persistência;

ETAPA 4 - HOMOLOGIA : Devido ao alto volume de pontos nos diagramas de persistência, teremos que condensar toda esta informação em estruturas mais simples para trabalharmos. Com isso, introduziremos duas ferramentas: os *funcionais de homologia persistente* e a *entropia persistente*.

ETAPA 5 - HOMOLOGIA : Realizaremos o agrupamento dos dados da Etapa 4, os dividindo em dois conjuntos disjuntos. Deste agrupamento, resultante do algoritmo *K-means*, construiremos o classificador C_{Hom} ;

ETAPA 6 - BOTÔMETRO : Nesta etapa, iniciaremos aplicando a ferramenta *botometer* aos dados coletados na Etapa 1. Em seguida, tal qual fizemos para a Etapa 5 da

Homologia, realizaremos o agrupamento em dois clusters dos histogramas dos perfis de usuários obtidos a partir do Botômetro. Finalmente, deste agrupamento resultante do algoritmo *K-means*, construiremos o classificador $\mathcal{C}_{\text{Bot}}$;

ETAPA 7 - FINAL : Concluiremos nossa análise comparando os dois classificadores: $\mathcal{C}_{\text{Hom}}$ e $\mathcal{C}_{\text{Bot}}$.

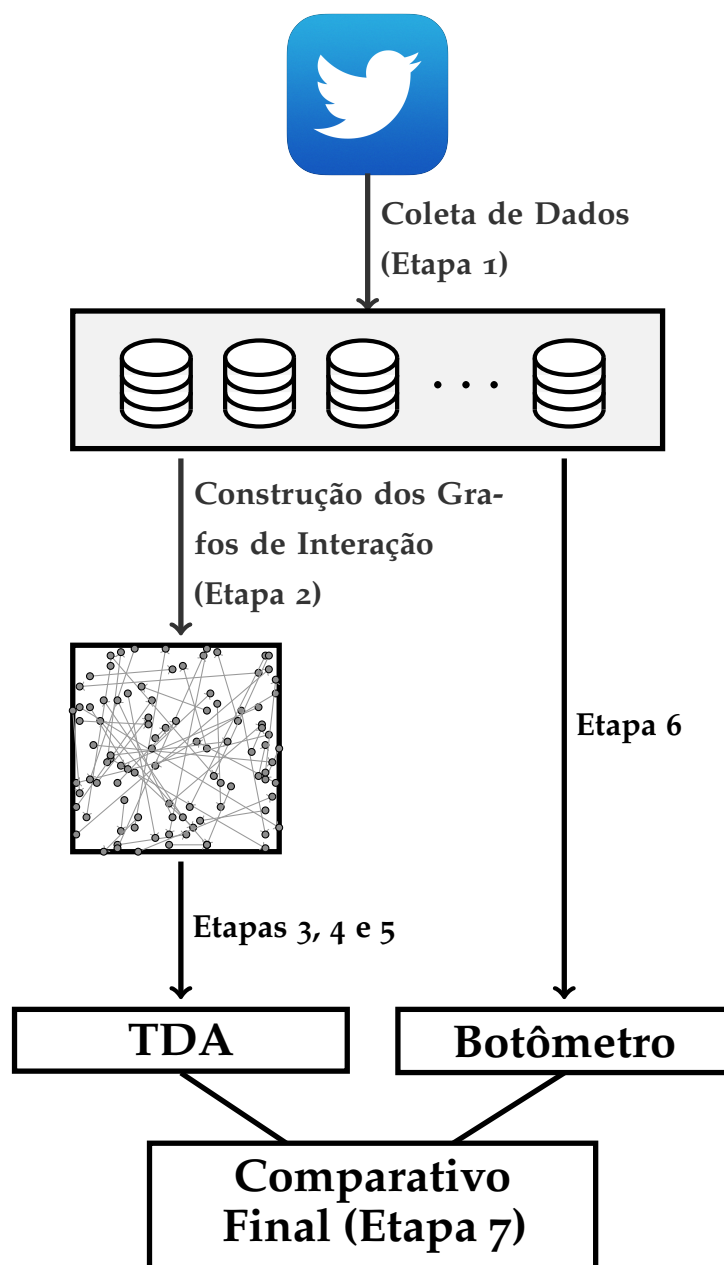


Figura 36: Passos da metodologia de Estudo.

4.1 COLETA DE DADOS

Nossa amostra será composta pela seguinte sequência de conjuntos: ⁵

$$T(t_0^1, t_f^1, K^1), T(t_0^2, t_f^2, K^2), \dots, T(t_0^{20}, t_f^{20}, K^{20}),$$

onde cada uma das tuplas (t_0^i, t_f^i, K^i) que indexam os conjuntos $T(t_0^i, t_f^i, K^i)$, com $i \in \{1, 2, 3, \dots, 20\}$, representam as propriedades

- (1) t_0^i é o instante inicial de tempo para coleta de tweets. Este tempo leva em conta o dia, o mês, o ano, a hora e o minuto para iniciarmos a coleta;
- (2) t_f^i é o instante final de tempo para coleta de tweets. Este tempo leva em conta o dia, o mês, o ano, a hora e o minuto para encerrarmos a coleta;
- (3) K^i é o conjunto de palavras chaves o qual os tweets devem conter pelo menos um elemento.

Já o elemento central da amostragem, os conjuntos $T(t_0^i, t_f^i, K^i)$, podem ser interpretados como uma coleção de árvores finitas e disjuntas T_j^i :

$$T(t_0^i, t_f^i, K^i) = \bigcup_{j=1}^{N_i} T_j^i,$$

onde todo vértice destas arvorés $w_j^i = (m_u, u) \in T_j^i$ é um par, formado por um tweet m_u o qual foi escrito pelo usuário u , e que satisfaz as restrições

- (i) de tempo de criação: $m_u \in [t_0^i, t_f^i]$;
- (ii) e de conteúdo: m_u deve conter ao menos uma palavra, ou expressão, contida no conjunto de palavras chaves K^i .

As arestas das árvores T_j^i são definidas como

$$(m_u, u) \rightarrow (m_w, w) \iff m_w \text{ é uma resposta, ou um retweet, ou uma citação ao tweet } m_u.$$

⁵ **Nota técnica:** Para a coleta dos dados, utilizando os scripts disponíveis em minha conta do GitHub <https://github.com/rafaelpcarneiro/fakeNewsAnalysis>, é necessário, antes de mais nada, ter uma conta de desenvolvedor do Twitter (otida em <https://developer.twitter.com/en>).

Na Figura 37 temos esquematizado uma sequência de conversas no Twitter, em sua forma de árvore. Nesta Figura temos uma representação visual do que nossa amostra será: uma coleção de árvores, construídas seguindo o critério de indexação descrito anteriormente.

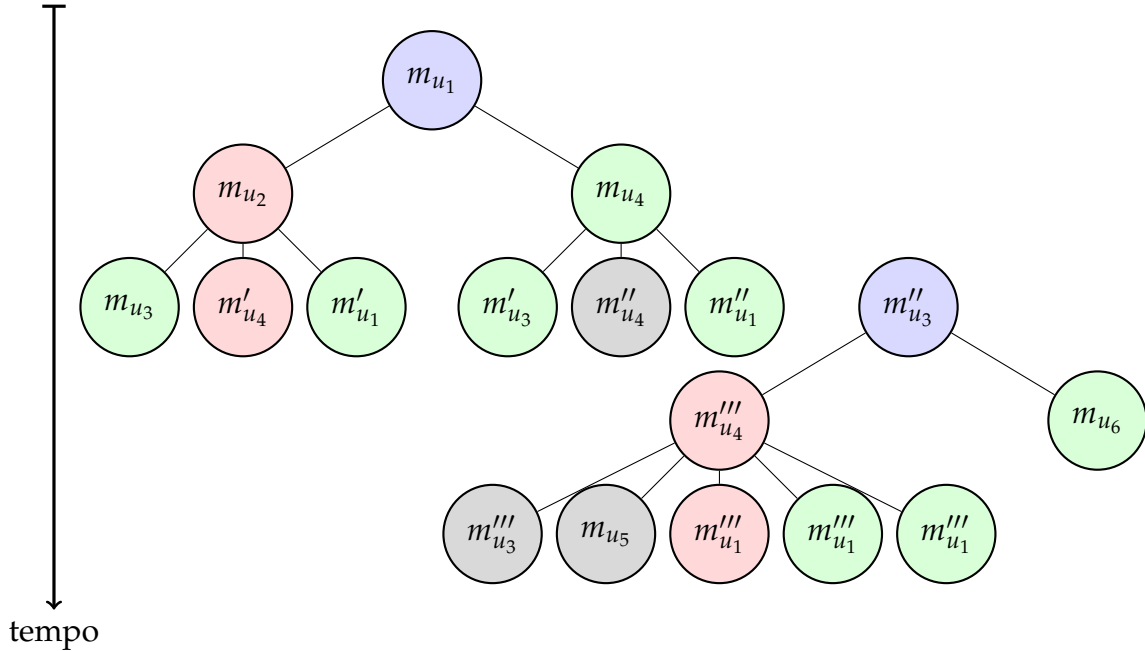


Figura 37: Tweets coletados seguindo uma restrição de palavras chaves e de tempo de publicação. Aqui, neste exemplo, nós de cor ■ representam mensagens simples, que não interagem com ninguém. Enquanto isso, nós de cor ■ são retweets, nós da cor ■ são respostas e, finalmente, nós da cor ■ são respostas com citações.

Finalmente, como já expostos na **Etapa 1** da metodologia de estudo, os conjuntos

$$T(t_0^1, t_f^1, K^1), T(t_0^2, t_f^2, K^2), \dots, T(t_0^{20}, t_f^{20}, K^{20}),$$

“serão” separados em dois grupos

Grupo dos tweets orgânicos G_1 :

$$T(t_0^1, t_f^1, K^1), T(t_0^2, t_f^2, K^2), T(t_0^{10}, t_f^{10}, K^{10})$$

Grupo dos tweets influenciados por fake news G_2 :

$$T(t_0^{11}, t_f^{11}, K^{11}), T(t_0^{12}, t_f^{12}, K^{12}), T(t_0^{20}, t_f^{20}, K^{20})$$

Esta separação entre grupos orgânicos e não orgânicos dá-se, única e exclusivamente, pela escolha das palavras chaves nos conjuntos $K_1, K_2, \dots, K_{20}$. Para K_1, K_2, K_{10} selecionamos palavras chaves relacionadas com tópicos de futebol, esportes e outros. Já

para os grupos fortemente influenciados por fake news, escolhemos, como esperado, tópicos fortemente correlacionados com mensagens falsas, como, por exemplo, tópicos de política, economia, etc.

Para encerrarmos, notemos que o tempo de coletas basearam-se em períodos de 12h, do meio-dia à meia-noite, e em dias tal que os os tópicos de interesse fossem encontrados com maior facilidade.

4.2 DESCRIÇÃO DOS GRAFOS AMOSTRADOS

Abaixo listaremos os três itens, que caracterizam completamente as árvores de tweets, de cada que um dos conjuntos $T_{(t_0^1, t_f^1, K^1)}, T_{(t_0^2, t_f^2, K^2)}, \dots, T_{(t_0^{20}, t_f^{20}, K^{20})}$. Para o item referente as palavras chaves, a cada grafo amostrado descreveremos o conteúdo de seus tweets, os quais estão diretamente associados as palavras chaves usadas nas coletas de dados. Já para a questão do dia de publicação dos tweets, ressaltamos que todos as publicações analisadas foram escritas em um dia específico, começando a coleta ao meio dia, e finalizando às 3h da manhã do dia seguinte.

Grafo	Publicação	Conteúdo dos tweets	Nome do grafo
$T_{(t_0^1, t_f^1, K^1)}$	14/09/2022	tweets sobre o programa de televisão <i>A Fazenda</i> , em um dia que o programa supostamente atrairia maior atenção de seus telespectadores;	A Fazenda
$T_{(t_0^2, t_f^2, K^2)}$	12/05/2022	tweets sobre o evento de um <i>Buraco Negro</i> , no dia em que fotos foram publicadas por astrônomos;	Buraco Negro
$T_{(t_0^3, t_f^3, K^3)}$	01/11/2022	tweets sobre a Deputada Federal que, dias antes da eleição, perseguiu armada um cidadão por, “supostamente”, ter sido agredida por ele;	Deputada Federal armada
$T_{(t_0^4, t_f^4, K^4)}$	20/11/2022;	tweets falando sobre a abertura da Copa do Mundo de futebol;	Copa do Mundo

$T(t_0^5, t_f^5, K^5)$	07/11/2022;	tweets cujo conteúdo fazem menção a uma piada que tornou-se viral, gerando muita repercussão por enganar diversas pessoas com seu trocadilho e por escancarar, de forma bem humorada, o efeito de mensagens falsas no Twitter;	Coronel Benjamin
$T(t_0^6, t_f^6, K^6)$	16/02/2021;	tweets sobre o Deputado Federal que ameaçou integrantes do Supremo Tribunal Federal, resultando em sua prisão;	Deputado Federal preso
$T(t_0^7, t_f^7, K^7)$	12/10/2022;	tweets sobre o dia das crianças;	Dia das Crianças
$T(t_0^8, t_f^8, K^8)$	02/11/2022	tweets sobre os resultados das eleições de domingo, dia 30/11/2022;	Eleições 2022
$T(t_0^9, t_f^9, K^9)$	04/03/2022	tweets sobre o Deputado Estadual que teve áudios seus vazados. Destes áudios, escancarou-se sua total falta de empatia com as mulheres ucranianas em situação de vulnerabilidade, decorrente da guerra em seu país;	Deputado Estadual e ucranianas
$T(t_0^{10}, t_f^{10}, K^{10})$	10/10/2020	tweets sobre as falas do Presidente da República do Brasil sobre a postura dos EUA com relação às queimadas na floresta Amazônica;	Presidente Brasileiro e Biden
$T(t_0^{11}, t_f^{11}, K^{11})$	08/02/2022	tweets sobre o posicionamento do youtuber Monark com relação ao nazismo;	Monark
$T(t_0^{12}, t_f^{12}, K^{12})$	12/10/2022	tweets sobre o feriado de Nossa Senhora Aparecida;	Nossa Senhora Aparecida

$T(t_0^{13}, t_f^{13}, K^{13})$	07/10/2022	tweets sobre o último episódio da novela <i>Pantanal</i> ;	Novela Pantanal
$T(t_0^{14}, t_f^{14}, K^{14})$	28/03/2022;	tweets sobre a agressão do ator Will Smith ao humorista Chris Rock, na premiação do Oscar;	Tapa no Oscar
$T(t_0^{15}, t_f^{15}, K^{15})$	14/11/2021	tweets sobre a corrida de Fórmula 1 que ocorreu no Domingo, em Interlagos, e que teve a vitória do piloto Lewis Hamilton;	Corrida F1
$T(t_0^{16}, t_f^{16}, K^{16})$	02/12/2021;	tweets sobre a rodada final do Campeonato Brasileiro, que consagrou o Atlético Mineiro como campeão;	Futebol
$T(t_0^{17}, t_f^{17}, K^{17})$	05/11/2021;	tweets sobre o acidente fatal envolvendo a cantora de sertanejo, Marília Mendonça;	Marília Mendonça.
$T(t_0^{18}, t_f^{18}, K^{18})$	21/09/2021;	tweets sobre a ida do Presidente Brasileiro à ONU;	ONU
$T(t_0^{19}, t_f^{19}, K^{19})$	13/10/2021	tweets de cunho político, e em espanhol, envolvendo o debate entre candidatos que ocorreu na Argentina;	Política Argentina
$T(t_0^{20}, t_f^{20}, K^{20})$	16/02/2022	tweets sobre o encontro do Presidente Brasileiro com o Vladimir Putin, na Rússia, antes da Rússia anunciar guerra contra a Ucrânia;	Visita a Putin.

Tabela 1: Descrição da amostragem dos grafos em estudo. Aqui também estabelecemos termos que usaremos daqui a diante para nos referirmos a eles.

4.3 TRANSFORMAÇÃO DOS DADOS BRUTOS PARA OS GRAFOS DE INTERAÇÃO ENTRE USUÁRIOS

Esta seção, referente à Etapa 2 da metodologia de estudo, servirá para estabelecermos a estrutura de real interesse de nossa pesquisa: o *Grafo de interação de usuários*.

Diferentemente das árvores da Figura 37, que são simples componentes conexas, estudaremos os Grafos orientados $G^i = (V^i, A^i)$ e com função peso $\mathcal{W}^i : A^i \rightarrow \mathbb{R}_{\geq 0}$ onde, para cada índice $i \in \{1, 2, \dots, 20\}$, vale

$$V^i = \{u \text{ um usuário} ; \exists w_u \text{ um tweet tal que } (w_u, u) \text{ é um vértice de } T_{t_0^i, t_f^i, K^i}\}$$

e

$A^i = \{(u, u') ; u, u' \in V^i \text{ e } u' \text{ interage com } u, \text{ seja respondendo, dando RT, ou citando}\}$, com a aresta orientada (u, u') sendo representada, também, por $u \rightarrow u'$. Já a matriz de peso é definida por

$$\begin{aligned} \mathcal{W}^i : A^i &\rightarrow \mathbb{R}_{\geq 0} \\ (u, u') &\mapsto \frac{\sum_{w \in N(u, u')} \tau_{(u, u')}^w}{|N(u, u')|} \end{aligned}$$

com $\tau_{(u, u')}^w$ o tempo demorado por u' para interagir com u , por meio de um tweet $w \in N(u, u')$ ($N(u, u')$ o conjunto finito de tweets que u' interagiu com u). Em outras palavras, a matriz peso expressa o tempo médio com que o usuário u' interage com o usuário u .

Para cada conjunto de nosso banco dados, com $i \in \{1, 2, \dots, 20\}$, faremos esta transformação, resultando no esquema

Grupo dos tweets orgânicos G_1

$$\begin{aligned} T_{(t_0^1, t_f^1, K^1)} &\rightarrow G^1 = (V^1, A^1) \text{ e } \mathcal{W}^1 : A^1 \rightarrow \mathbb{R}_{\geq 0} \\ T_{(t_0^2, t_f^2, K^2)} &\rightarrow G^2 = (V^2, A^2) \text{ e } \mathcal{W}^2 : A^2 \rightarrow \mathbb{R}_{\geq 0} \\ T_{(t_0^{10}, t_f^{10}, K^{10})} &\rightarrow G^{10} = (V^{10}, A^{10}) \text{ e } \mathcal{W}^{10} : A^{10} \rightarrow \mathbb{R}_{\geq 0} \end{aligned}$$

Grupo dos tweets influenciados por fake news (ou, simplesmente, tweets não orgânicos) G_2

$$\begin{aligned} T_{(t_0^{11}, t_f^{11}, K^{11})} &\rightarrow G^{11} = (V^{11}, A^{11}) \text{ e } \mathcal{W}^{11} : A^{11} \rightarrow \mathbb{R}_{\geq 0} \\ T_{(t_0^{12}, t_f^{12}, K^{12})} &\rightarrow G^{12} = (V^{12}, A^{12}) \text{ e } \mathcal{W}^{12} : A^{12} \rightarrow \mathbb{R}_{\geq 0} \\ T_{(t_0^{20}, t_f^{20}, K^{20})} &\rightarrow G^{20} = (V^{20}, A^{20}) \text{ e } \mathcal{W}^{20} : A^{20} \rightarrow \mathbb{R}_{\geq 0}. \end{aligned}$$

A Figura 38 ilustra como um banco de dados, formado por tweets com estrutura semelhante ao da Figura 37, gera um grafo de interação de usuários.

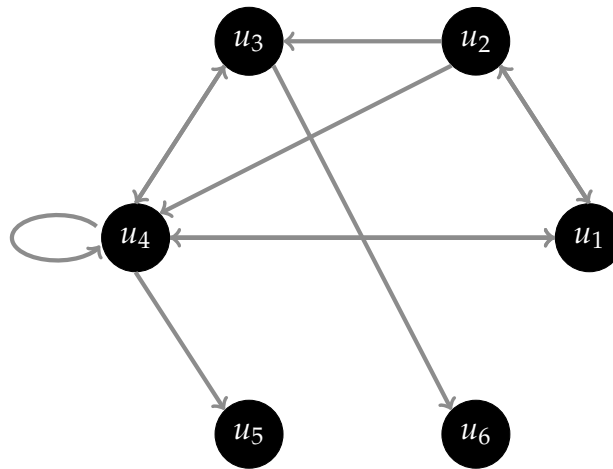


Figura 38

4.4 ANÁLISE EXPLORATÓRIA DOS GRAFOS DE INTERAÇÃO DE USUÁRIOS

Nesta seção apresentaremos algumas das propriedades básicas dos grafos, orientados e com pesos nas arestas, obtidos pelo nosso processo de coleta e tratamento dos tweets. Iniciamos comparando a quantidade de vértices e arestas de cada grafo, que pode ser visto na Tabela 2.

Conforme podemos ver na Tabela 2, existem grafos que explodem em tamanho, fugindo do padrão médio. O grafo G^7 , com mais de 1 milhão de arestas, por exemplo. Este grafo, referente ao acidente da cantora de sertanejo, Marília Mendonça, admite grande quantidade de usuários interagindo muito devido ao prestígio da cantora e, também, pela tragédia que gera uma comoção geral nas pessoas.

Abaixo, aprofundaremos nossa visão dos grafos de interação de usuários, apresentando:

- (i) Uma visual espacial dos grafos, focando nos usuários mais influentes dos grafos, e suas capacidades de influir os pontos vizinhos; Esta representação espacial segue a proposta em [31], utilizando o programa *free-software* graphviz⁶.

⁶ <https://graphviz.org>

Grafo amostrado	Quantidade de Vértices	Quantidade de Arestas
G^1 (A Fazenda)	10.819	15.387
G^2 (Buraco Negro)	17.543	19.790
G^3 (Copa do Mundo)	150.102	217.867
G^4 (Dia das Crianças)	269.592	438.142
G^5 (Corrida F1)	61.394	109.972
G^6 (Futebol)	119.441	219.700
G^7 (Marília Mendonça)	536.887	1.069.079
G^8 (Nossa Senhora Aparecida)	125.392	398.053
G^9 (Tapa no Oscar)	326.797	563.556
G^{10} (Novela Pantanal)	28.875	75.618
G^{11} (Deputada Federal armada)	170.948	399.042
G^{12} (Coronel Benjamin)	25.302	21.532
G^{13} (Deputado Federal preso)	29.444	53.802
G^{14} (Eleições 2022)	245.027	613.520
G^{15} (Deputado Estadual e ucranianas)	34.321	66.418
G^{16} (Presidente Brasileiro e Biden)	24.825	39.802
G^{17} (Visita a Putin)	166.814	412.967
G^{18} (Monark)	73.680	109.650
G^{19} (ONU)	111.541	262.818
G^{20} (Política Argentina)	31.131	61.033

Tabela 2: Quantidade de vértices e arestas de todos os grafos de interação obtidos.

- (ii) Distribuição dos horários de publicação dos tweets;
- (iii) A distribuição empírica da matriz de pesos de cada grafo, representando as interações entre os usuários;
- (iv) A distribuição empírica do grau de entrada e saída dos vértices dos grafos.

Na Figura 40 e 41 apresentamos os grafos de interação dos usuários, onde vértices com grau de saída maior, são representados por círculos de raios maiores. Para as arestas dos grafos, cores de maior intensidade representam a interação de usuários quaisquer com um usuário u_0 , que exerce uma alta influência na rede. A escala de cores usadas, para representar o peso das arestas, segue a ordem da Figura 39. Quanto

maior a intensidade da cor, maior a influência de um dos usuários que compõem a aresta.



Figura 39: Relação mapeando as cores das arestas. Quanto maior o valor mais influente é o vértice compondo a aresta (u_0, u_1) .

Das representações dos grafos das Figuras 40 e 41 surgem alguns padrões: a quantidade de pontos isolados nos grafos G^1 (A Fazenda) e G^2 (Buraco Negro) se distoam quando comparados aos restantes dos grafos, grafos G^3 até o G^{10} , Figura 40; já para os grafos da Figura 41, o grafo G^{12} (Coronel Benjamin) é o grafo com maior quantidade de pontos isolados; em geral, aparentemente temos a impressão de que a conectividade dos grafos da Figura 40 ocorre em menor intensidade do que os grafos da Figura 41.

Com relação aos histogramas dos itens (ii), (iii) e (iv), descritos acima, apresentaremos, nesta seção, apenas os histogramas do grafo G^7 (Marília Mendonça) e G^{14} (Eleições 2022), que são os grafos com maior quantidades de arestas, e cujas temáticas dos tweets são bem divergentes (tanto em assunto como comportamento dos usuários que publicaram os tweets).

Dos gráficos abaixo, Figuras 42 e 43, já observamos um comportamento curioso: a quantidade de tweets publicados do grafo G^7 (Marília Mendonça) apresenta um pico bem definido, decrescendo exponencialmente fora desse intervalo de pico. Enquanto isso, para o grafo G^{14} (Eleições 2022) temos o efeito contrário: os tweets são publicados com alta intensidade e por um largo período de tempo. Este padrão se estende, também, para o tempo médio de interação entre usuários de ambos grafos. Olhando a mediana dos pesos das arestas, de ambos grafos, percebemos que 50% das interações entres usuários, do Grafo G^7 , acontecem em tempo menor que 1.776s. Já no Grafo G^{14} , 50% dos usuários interagem em tempo menor que 10.139s. Ou seja, usuários do Grafo G^{14} dão continuidade as conversas passadas, não deixando que elas sumam dos tópicos mais discutidos no Twitter.

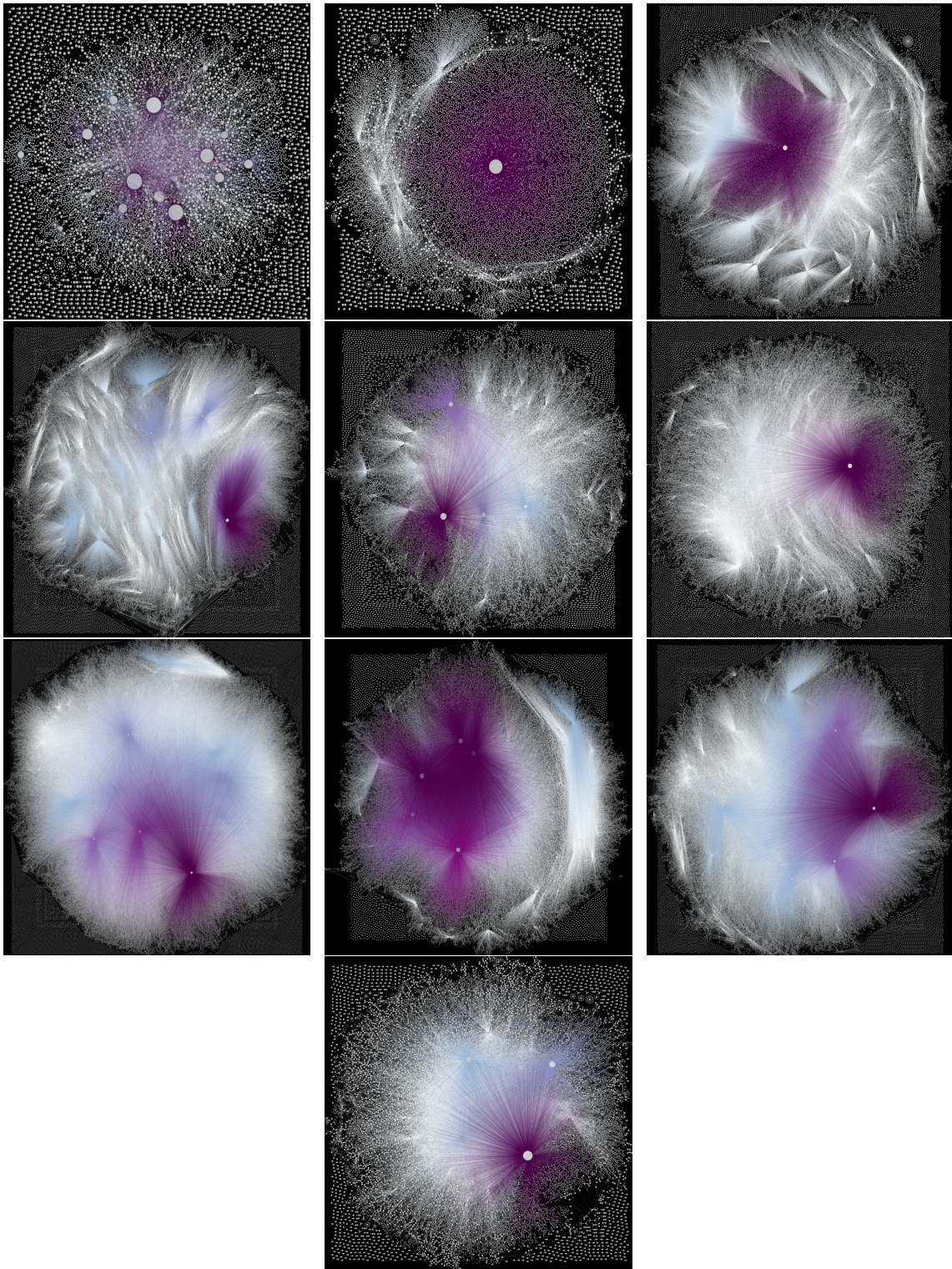


Figura 40: Grafos da esquerda para a direita: G^1 (A Fazenda), G^2 (Buraco Negro), G^3 (Copa do Mundo) (1ª linha); G^4 (Dia das Crianças), G^5 (Corrida F1), G^6 (Futebol) (2ª linha); G^7 (Marília Mendonça), G^8 (Nossa Senhora Aparecida), G^9 (Tapa no Oscar) (3ª linha); G^{10} (Novela Pantanal) (última linha). Quanto mais influente um usuário for, maior o raio do círculo que o representa. Arestas representadas por cores mais fortes, estão associadas a usuários com maior influência no grafo. (Em nosso contexto, influência é medida pelo o grau de saída do vértice.)

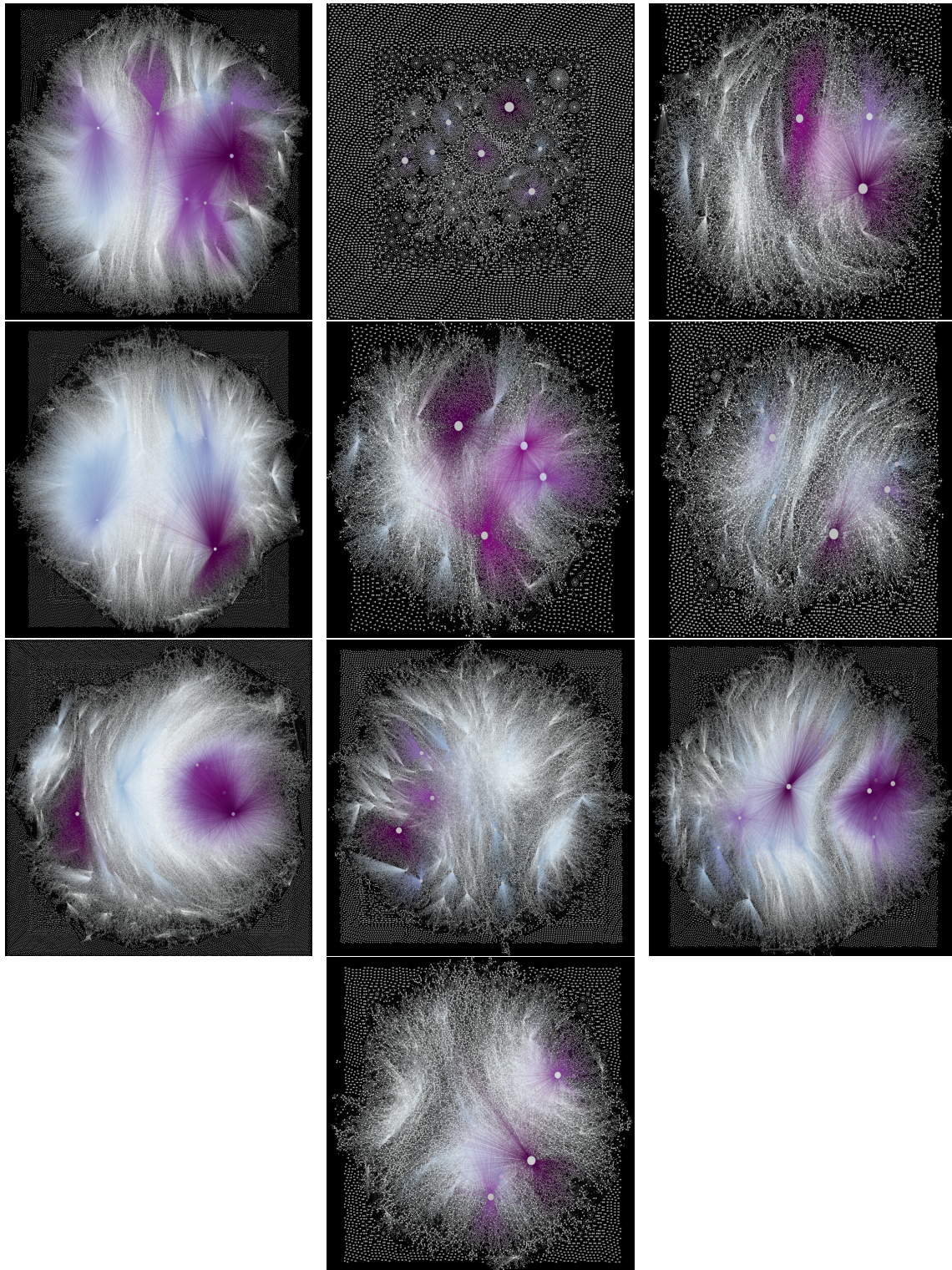


Figura 41: Grafos da esquerda para a direita: G^{11} (Deputada Federal armada), G^{12} (Coronel Benjamin), G^{13} (Deputado Federal preso) (1ª linha); G^{14} (Eleições 2022), G^{15} (Deputado Estadual e ucranianas), G^{16} (Presidente Brasileiro e Biden) (2ª linha); G^{17} (Visita a Putin), G^{18} (Monark), G^{19} (ONU) (3ª linha); G^{20} (Política Argentina) (última linha). Quanto mais influente um usuário for, maior o raio do círculo que o representa. Arestas representadas por cores mais fortes, estão associadas a usuários com maior influência no grafo. (Em nosso contexto, influência é medida pelo o grau de saída do vértice.)

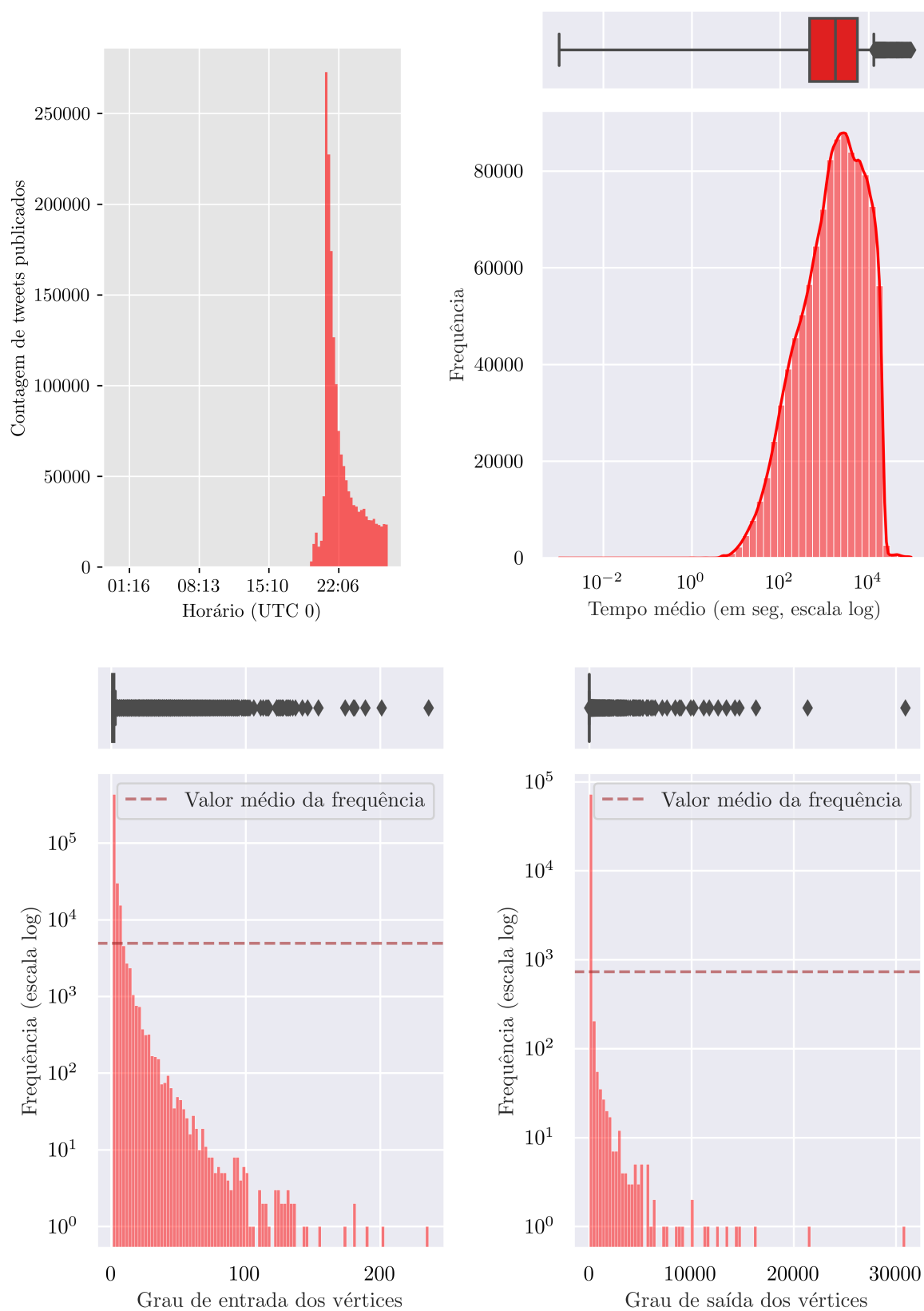


Figura 42: Histogramas dos horários de publicação dos tweets e dos pesos das arestas do **Grafo**

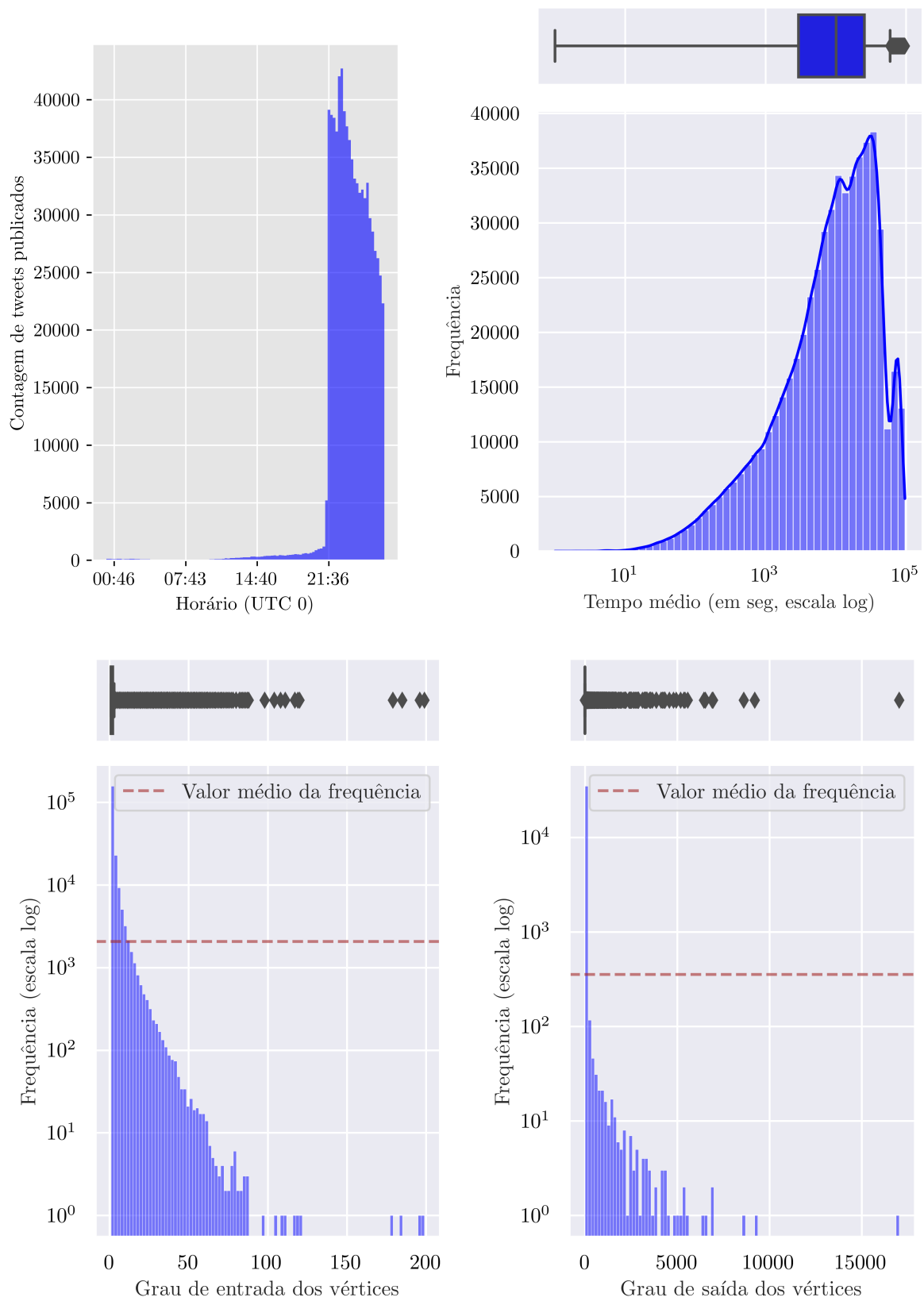


Figura 43: Histogramas dos horários de publicação dos tweets e dos pesos das arestas do **Grafo**

Já para os histogramas sobre os graus de entrada e saída, ainda observando as Figuras 42 e 43, não percebemos diferenças drásticas para os graus de entrada dos grafos. Contudo, para os graus de saída, aparentemente o Grafo G^7 (Marília Mendonça) possui vértices com maior influência que o Grafo G^{14} (Eleições 2022), olhando isoladamente. Para encerrarmos nossa análise inicial, apresentamos os boxplots dos tempos médios de interação dos usuários, de cada grafo amostrado, vide a Figura 44 abaixo. A priori, as medianas dos grafos não sugerem diferenças drásticas entre os grafos coletados.

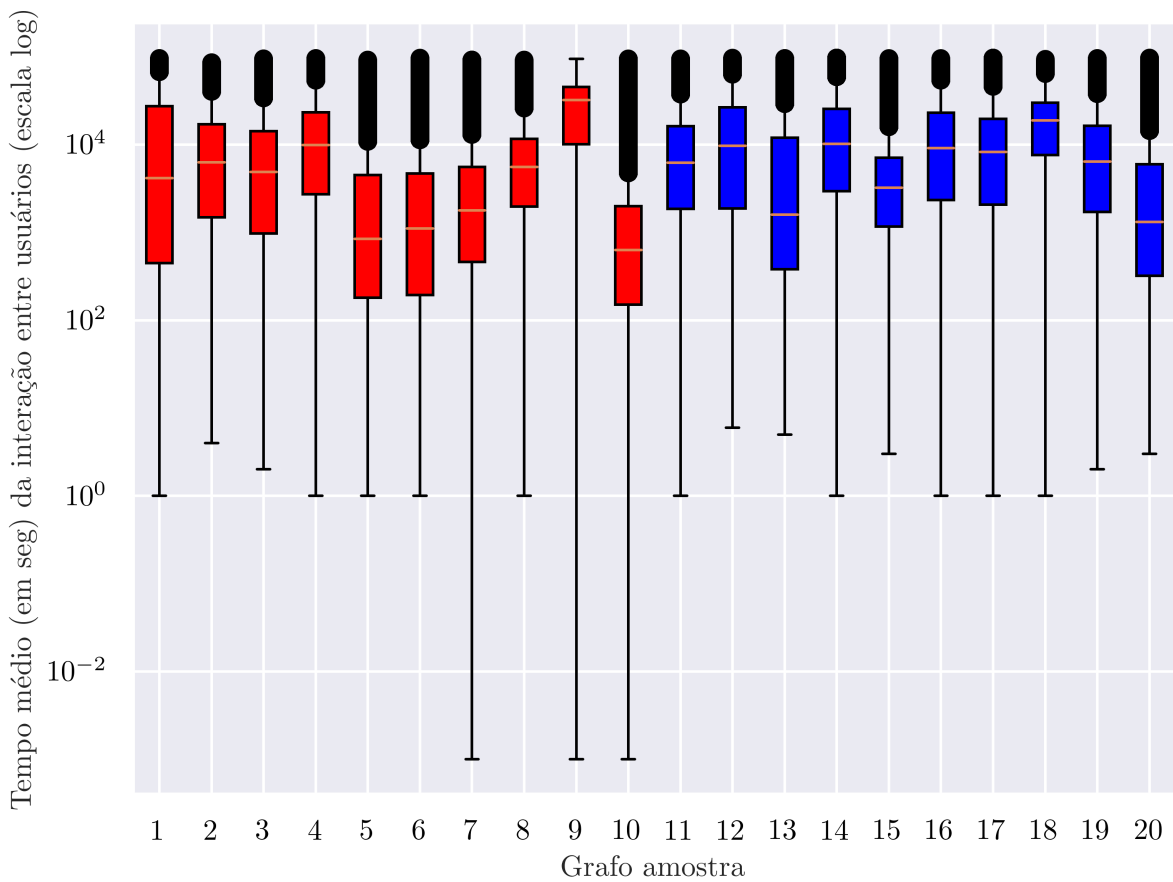


Figura 44: Boxplot dos pesos das arestas de todos os grafos amostrados.

A seguir, aplicaremos técnicas de homologia persistente para estudarmos os padrões dos grafos de interação $G^1, G^2, \dots, G^{20}$. Porém, a homologia persistente utilizada para detecção de padrões não será a usual, utilizada nos capítulos anteriores, mas sim uma versão modificada, para detectar grafos orientados, e chamada de *Homologia Persistente de Caminhos*. Tal proposta de homologia persistente pode ser encontrada

no artigo [8] ou no artigo [9] (o qual inclui a implementação do algoritmo para o cálculo da homologia persistente)

A escolha pela homologia persistente de caminhos é motivada pela assimetria do grafo estudado. Esta assimetria surge pois um usuário do Twitter pode se comunicar com outro usuário que, não necessariamente, continuará com a conversa (fato muito usual na prática). Desta maneira, o grafo de interação de usuários admitirá arestas ordenadas, cujos pesos serão distintos dependendo da orientação.

4.5 A HOMOLOGIA PERSISTENTE DE CAMINHOS

Esta seção será apenas uma introdução das ferramentas necessárias para a construção da homologia persistente de caminhos. De fato, sua construção é muito semelhante a homologia simplicial e a homologia persistente (exposto no Capítulo 1 e 2). Portanto, o interessado pode pular para a próxima Seção sem perda de generalidade. Todo o instrumental exposto aqui pode ser encontrado em [8] ou [9].

A escolha de utilizar a homologia persistente de caminhos, ao invés da homologia persistente usual, decorre do fato dela ser sensível a assimetrias. Isto é, grafos orientados terão sua orientação preservada quando criarmos os espaços vetoriais gerados pelos seus caminhos, ao invés do espaço vetorial gerado pelos simplexes abstratos (que são espaços quocientes de simplexes que preservam a mesma orientação).

Esta vantagem de detectar assimetrias tem um preço alto: complexidade computacional. Dependendo do grafo $\mathcal{G}$, pode-se ser necessário calcular a quantidade de todos os caminhos possíveis do grafo que, no pior cenário possível, pode ser algo da ordem de $|\mathcal{G}|^{n+2}$, com

$$|\mathcal{G}| = \text{quantidade de vértices do grafo}$$

e n a dimensão da homologia.

Por exemplo, um grafo de 10^5 vértices (100.000 vértices), para calcularmos sua homologia persistente de caminhos de dimensão 1 será necessário, no pior caso, uma quantidade de 10^{15} caminhos. Se considerarmos que cada caminho ocupe 1 byte de memória, será necessário o equivalente a 909 terabytes de memória para o processamento do algoritmo. Ou seja, cuidado deve ser tomado ao utilizar-se desta metodologia.

4.5.1 Caminhos em um grafo

Seja $\mathcal{G}$ um grafo finito. Dizemos que toda sequência de pontos $a_0, a_1, \dots, a_p \in \mathcal{G}$ é um caminho em $\mathcal{G}$ de tamanho p , sendo representado por

$$w = [a_0, a_1, a_2, \dots, a_p].$$

Além do mais, fixado o espaço vetorial gerado por todos os caminhos de tamanho p , definimos o operador bordo de dimensão p

$$\partial_p([a_0, a_1, a_2, \dots, a_p]) = \sum_{i=0}^p (-1)^i [a_0, a_1, a_2, \dots, \hat{a}_i, \dots, a_p].$$

Até aí nada de novo. Todavia, se compararmos com a teoria usual, estaríamos trabalhando em um espaço vetorial resultante de classes de equivalência, onde a relação de equivalência seria dada por

$$[a_0, a_1, a_2, \dots, a_p] \cong \text{sgn}(f)[a_{f(0)}, a_{f(1)}, a_{f(2)}, \dots, a_{f(p)}], \quad \text{com } f \text{ uma permutação.}$$

A partir desta relação de equivalência temos o efeito de perdermos as orientações destes caminhos.

Tendo em vista preservar as orientações dos caminhos de um grafo, o artigo [8] propõe estruturas auxiliares que nos possibilitem, ainda, obter homologias persistentes. Definiremos estas estruturas, preservando a notação do artigo original.

Tomando por Λ_p o espaço vetorial gerado por todos os caminhos de tamanho p de um grafo $\mathcal{G}$, dizemos que todo caminho $w = [a_0, a_1, a_2, \dots, a_p] \in \Lambda_p$ é regular se valer

$$a_i \neq a_{i+1}, \forall i \in \{0, 1, \dots, p-1\}.$$

Respectivamente, diremos que o caminho $w \in \Lambda_p$ é irregular se não for regular. A notação para estes caminhos será

$$\mathcal{R}_p = \{w \in \Lambda_p; w \text{ é regular}\} \quad \text{e} \quad \mathcal{I}_p = \Lambda_p \setminus \mathcal{R}_p.$$

A intenção agora é construir um complexo de cadeia a partir destes espaços, utilizando o operador bordo acima. Decorre que, ao se tomar o operador bordo de caminhos regulares pode acontecer de sua imagem ser um caminho que não seja possível de ocorrer no grafo.

Dizemos que um caminho $w = [a_0, a_1, a_2, \dots, a_p]$ de um grafo orientado $\mathcal{G}$ é um caminho possível se

$$(a_0, a_1), (a_1, a_2), \dots, (a_{p-1}, a_p)$$

forem arestas do grafo $\mathcal{G}$. Notação

$$\mathcal{A}_p = \{w \in \mathcal{R}_p; w \text{ for um caminho possível}\}.$$

Finalmente, dizemos que todo caminho w de tamanho $p > 0$ é ∂ -invariante se

$$\partial(w) \in \mathcal{A}_{p-1}.$$

Agora, para fazer o tratamento de homologia persistente será necessário estabelecer uma filtração do grafo em estudo.

A seguir será dada a filtração utilizada para se calcular a homologia persistente de caminhos. Dado um grafo orientado finito $\mathcal{G} = (V, A)$, o qual não admite loops (i.e., dado $x \in V$ não serão permitida arestas do tipo $(x, x) \in A$), e munido de pesos nas arestas:

$$\phi_{\mathcal{G}} : A \rightarrow \mathbb{R} \geq 0,$$

satisfazendo

$$\forall a, b \in V, \text{ vale } \phi_{\mathcal{G}}((a, b)) \geq 0 \text{ e } \phi_{\mathcal{G}}((a, b)) = 0 \iff a = b.$$

A filtração usada para se obter a homologia persistente de caminhos será $\mathcal{F} = (\mathcal{G}_i)_{i \in \mathbb{R}_{\geq 0}}$, onde $\mathcal{G}_i$ é um grafo orientado indexado pelos reais positivos tal que $\mathcal{G}_i = (V_i, A_i)$, satisfazendo

$$V_i = V \quad \text{e} \quad A_i = \{(a, b) \in A; \phi_{\mathcal{G}}((a, b)) \leq i\}.$$

Por meio desta filtração de grafos orientados, a Proposição 10 e o Teorema 15 do artigo [8] apresentam a Homologia Persistente de Caminhos, o qual satisfaz a condição de seu diagrama de persistência de caminhos ser estável a perturbações, i.e., pequenas translações dos vértices do grafo influenciarão pouco no diagrama de persistência de caminhos.

Vale notar, como era de se esperar, que o algoritmo para o cálculo da homologia persistente de caminhos continuará sendo o mesmo que o algoritmo original, visto no Capítulo 2. A única diferença é o número total de caminhos necessários para calcular a eliminação gaussiana na matriz do operador bordo. O algoritmo para cálculo da homologia persistente, tal qual alguns exemplos, pode ser visto na Seção 2.5 do Capítulo 2, ou na página final do artigo [9]

4.6 CLASSIFICAÇÃO DOS TREND TOPICS USANDO A HOMOLOGIA PERSISTENTE

Após coletarmos uma amostra significativa de mensagens publicados no Twitter, nos mais diversos assuntos, daremos continuidade ao nosso objetivo central: estudar os grafos de interação dos usuários do Twitter. Destes grafos, avaliaremos como a ferramenta da *Homologia Persistente de Caminhos* é capaz de detectar, e diferenciar, o comportamento destes grafos, influenciados pelos tópicos que os usuários discutiam.

Assim que classificarmos nossos grafos pelas suas propriedades topológicas, compararemos nossas medidas e classificações com uma outra ferramenta, que classifica usuários de acordo com seus históricos de tweets publicados. Desta forma, teremos meios de compararmos como a homologia persistente sai quando comparada com técnicas usuais e já consolidadas no mercado.

As próximas seções se dividirão da seguinte maneira: iniciaremos o estudo dos grafos de interação dos usuários do Twitter, usando a Homologia Persistente de Caminhos; em seguida, trataremos os mesmos grafos, porém utilizando, desta vez, o projeto Botometer⁷, que fornecerá sua classificação pelo histórico das mensagens (avaliando modelos que levam em consideração o comportamento do usuário apenas, sem levar dinâmicas de interação com outros usuários); finalmente, compararemos ambas classificações, para assim sermos capazes de inferir a compatibilidade em ambas técnicas.

4.6.1 Resultados

Seguindo o caminho natural de nosso trabalho, trataremos nesta seção como a Homologia de Persistência de Caminhos detecta padrões, dos grafos de interação dos usuários da Seção 4.3.

Ao aplicarmos as Homologias de Persistência de Caminhos aos vinte grafos de interação coletados, atribuiremos a cada diagrama de persistência medidas resumo, que nos possibilitarão transformar estes diagramas em pontos do espaço euclidiano $\mathbb{R}^n$. Em posse destes dados, encerraremos esta seção aplicando a técnica de clusterização *K-means*, para classificarmos os grafos em dois cluster distintos.

⁷ <https://botometer.osome.iu.edu/>

Para simplificarmos a exposição dos resultados das próximas subseções, começaremos nosso trabalho apresentando os resultados obtidos pelo K-means:

Cluster 0	Corrida F1, Novela Pantanal, Futebol, Marília Mendonça, Copa do Mundo, Tapa no Oscar Visita a Putin, Deputado Estadual e ucranianas
Cluster 1	A Fazenda, Buraco Negro, Dia Das Crianças, Nossa Senhora Aparecida, Política na Argentina Coronel Benjamin, Eleições 2022, Monark ONU, Deputada Federal armada, Deputado Federal preso Presidente Brasileiro e Biden

Tabela 3: Grafos classificados segundo suas propriedades topológicas pelo K-means.

4.6.2 Diagramas de Persistência

Devido as dificuldades computacionais, resultantes da quantidade massiva de dados que coletamos, estabeleceremos o seguinte critério: as filtrações dos simplexes de dimensão 1, e dimensão 2, deverão ter uma quantidade de simplexes variando entre 1 milhão até 2 milhões de simplexes. Desta maneira, seremos capazes de detectar propriedades, de dimensão zero e um, em tempo razoável.

Consequentemente, ao invés de calcularmos a homologia de persistência de caminhos dos grafos de interação G^i , que em muitos casos não é possível, devido a alta demanda de tempo, ou de equipamento (dados na ordem de terabytes), consideraremos os seguintes subgrafos (dos grafos das mensagens)

$$\hat{T}_{(t_0^i, t_f^i, K^i)} \subseteq T_{(t_0^i, t_f^i, K^i)},$$

definidos em 4.1, tal que, ao criarmos os grafos de interação de usuários, induzidos dos grafos $\hat{T}_{(t_0^i, t_f^i, K^i)}$, teremos os grafos $\hat{G}^i$ satisfazendo a restrição que as filtrações, de caminhos simpliciais de dimensão 1 e 2, terão entre 1 milhão até 2 milhões de simplexes.

Para alcançarmos este objetivo, escolheremos subgrafos $\hat{T}_{(t_0^i, t_f^i, K^i)}$ tal que os tweets que os compõem, tenham sido publicados dentro de um intervalo de tempo que seja simétrico, sempre que possível, e com centro no instante de tempo de maior publicação

de tweets, do grafo original $T_{(t_0^i, t_f^i, K^i)}$. Na Figura 45, temos os intervalos escolhidos para construirmos cada subgrafo $\hat{T}_{(t_0^i, t_f^i, K^i)}$.

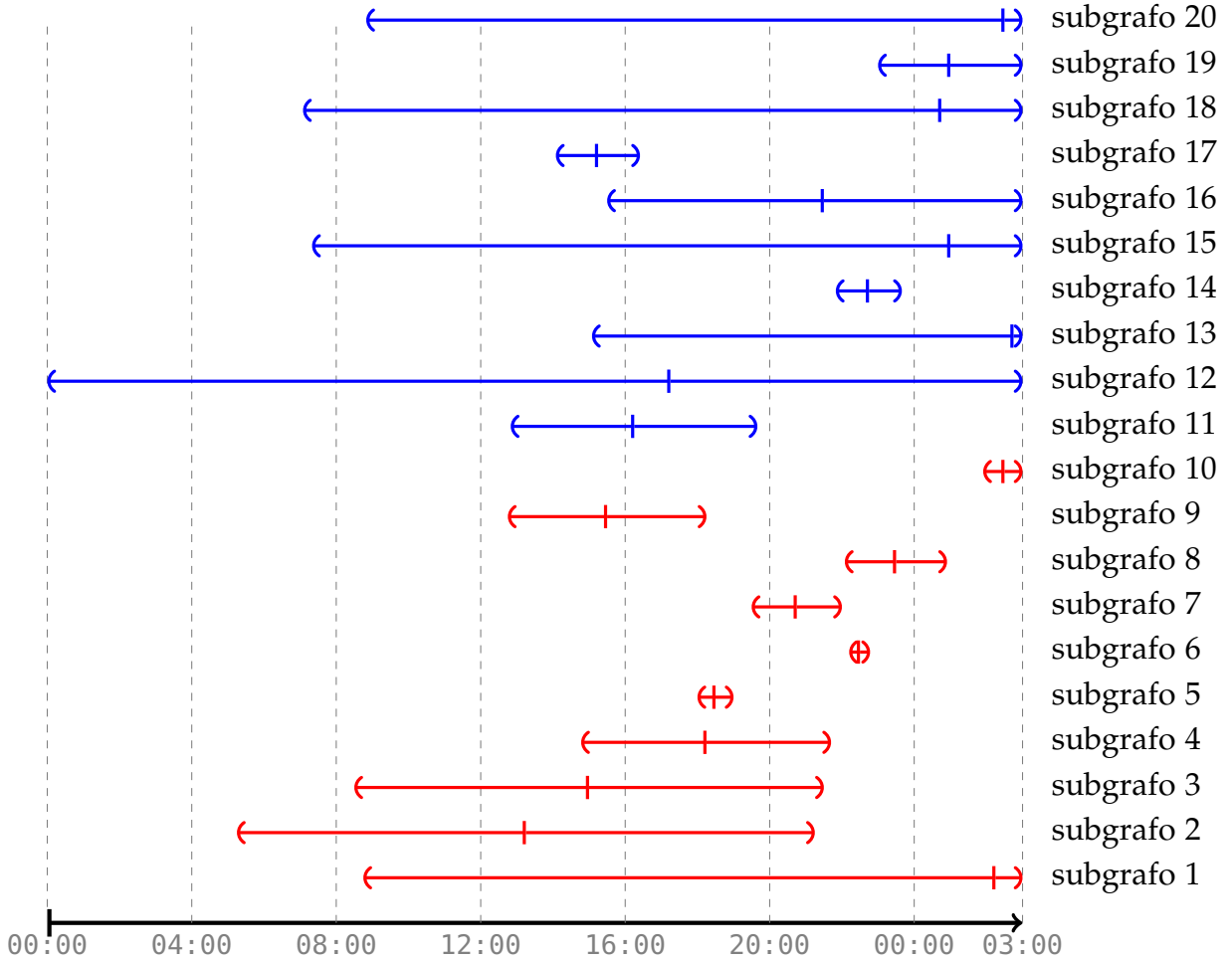


Figura 45: Intervalos de tempo utilizados para construirmos os subgrafos $\hat{T}_{(t_0^i, t_f^i, K^i)}$. Com estes intervalos somos capazes de calcular, em tempo razoável, os diagramas de persistência dos grafos de interação dos usuários.

Com as faixas de tempo, propostas na Figura 45 acima, somos capazes, agora, de criarmos os subgrafos $\hat{G}^i$ dos grafos de interação de usuários, com $i \in \{1, 2, \dots, 20\}$, tal que os diagramas de persistência de caminhos possam ser calculados em tempo razoável. Abaixo, nas Figuras 46 e 47, apresentamos os diagramas de persistência de cada subgrafo $\hat{G}^i$. Nos gráficos destas figuras, restringimos nossa atenção a apenas os simplexes que nascem e morrem em tempo finito, i.e., plotamos apenas os pontos (t_0, t_1) , do respectivo diagrama de persistência, tal que $t_0 \leq t_1 < \infty$, com t_0 o tempo de nascimento do simplexo e t_1 o tempo de morte do simplexo.

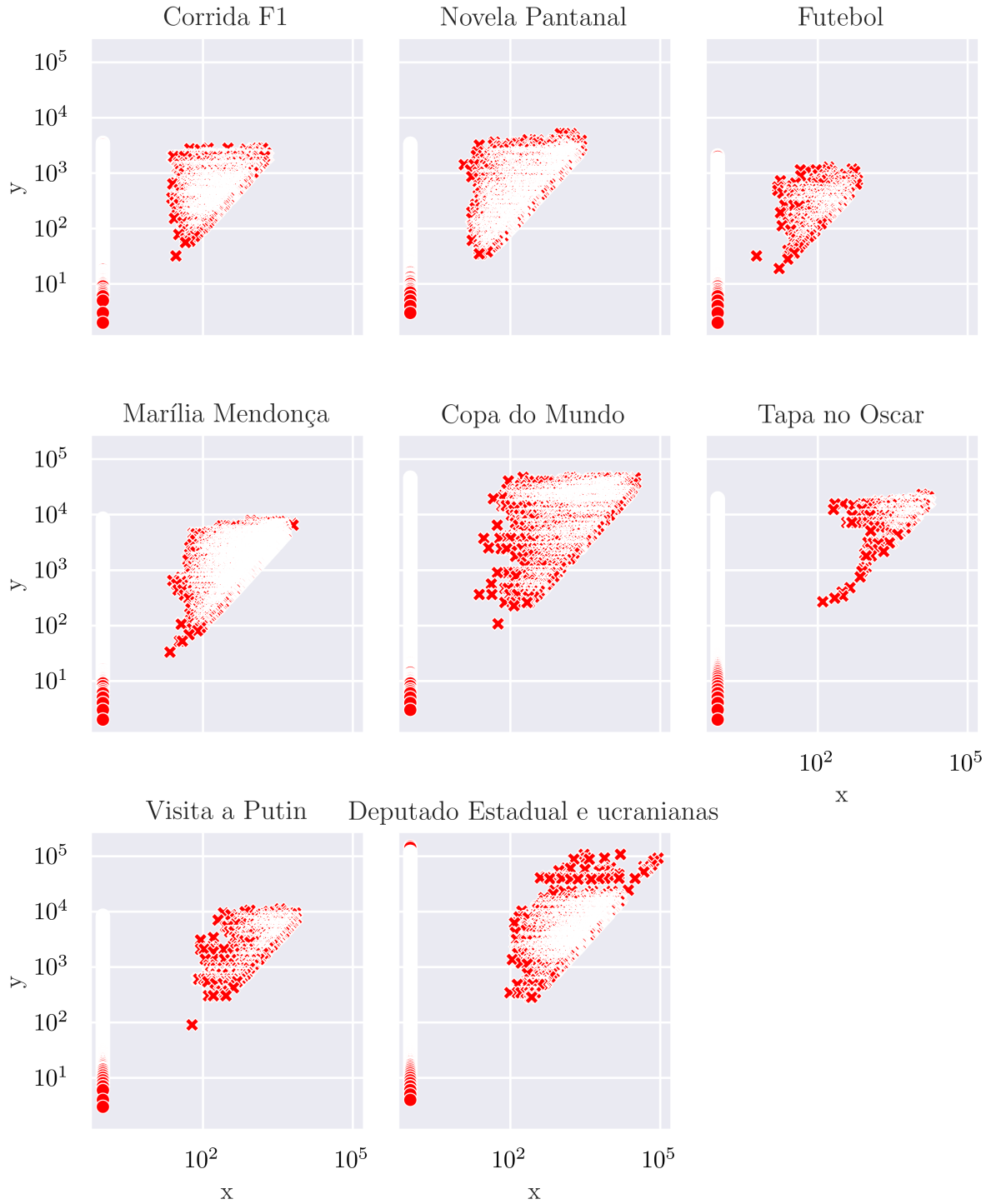


Figura 46: Diagramas de persistência de todos os subgrafos $\hat{G}^i$ do cluster 0, Tabela 3. Todos os gráficos estão em escala logarítmica e, tanto x , quanto y , estão em segundos. Pontos de dimensão zero são representados por $\bullet$, e pontos de dimensão um são representados por $\times$. (Obs: os grafos gerados aqui são obtidos pelo refinamento proposto na Figura 45. Além disso, apenas os ciclos que nascem e morrem em tempo finito são representados aqui.)

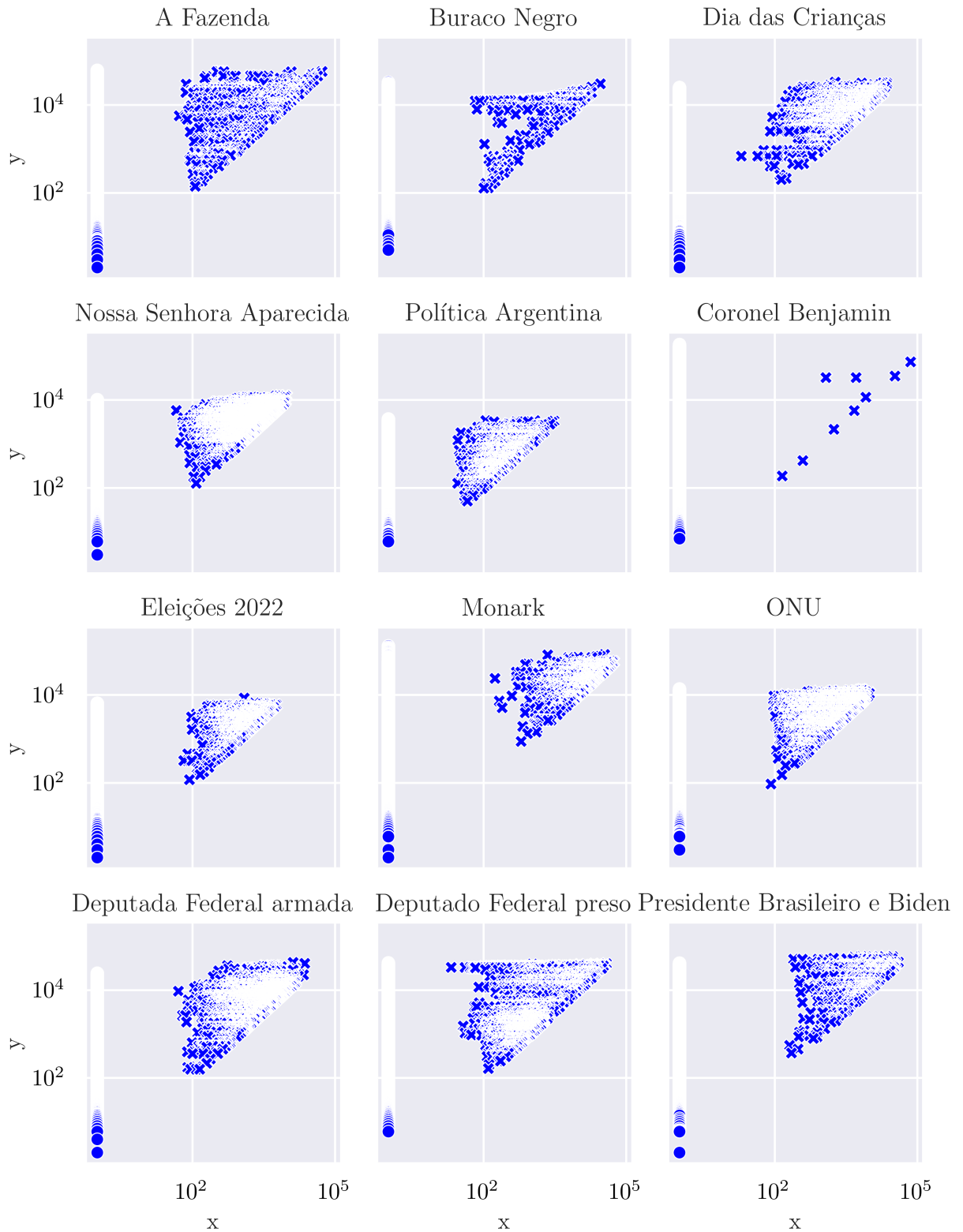


Figura 47: Diagramas de persistência de todos os subgrafos $\widehat{G}^i$ do cluster 1, Tabela 3. Todos os gráficos estão em escala logarítmica e, tanto x , quanto y , estão em segundos. Pontos de dimensão zero são representados por $\bullet$, e pontos de dimensão um são representados por $\times$. (Obs: os grafos gerados aqui são obtidos pelo refinamento proposto na Figura 45. Além disso, apenas os ciclos que nascem e morrem em tempo finito são representados aqui.)

Como é possível observar, ainda é muito difícil avaliar diferenças entre os grafos usando apenas as informações diretamente dos diagramas de persistência. Com esta perspectiva, de extrair informações dos diagramas de persistência, com o fim de avaliarmos suas diferenças, assim como estabelecermos comparações, iremos utilizar as ferramentas: *Silhuetas de Persistência* e *Entropia de Persistência*. Estas ferramentas serão abordadas na próxima seção.

Observação 4.6.1. Note que as propriedades topológicas que não morrem, das filtrações dos subgrafos $\widehat{G}^i$, dependem fortemente dos intervalos que selecionamos na Figura 45. Ou seja, se aumentarmos estes intervalos, então muito provavelmente estes ciclos desapareceram. Logo, devido a esta condição, induzida por não conseguirmos analisar os grafos G^i por completo, consideraremos apenas os ciclos que tornam-se bordo em tempo finito.

4.6.3 Silhuetas de Persistência e Entropia de Persistência

Nesta seção apresentaremos a definição de duas técnicas para extração de informação de diagramas de persistência: *silhuetas de persistência* e *entropia persistente*. A partir destas ferramentas, conseguimos extrair informação dos diagramas das Figuras 46 e 47, que junto com técnicas de aprendizado de máquina, mais explicitamente, com técnicas de clusterização, conseguimos agrupar os grafos em dois clusters distintos, com diferenças estatísticas significantes. Para uma exposição mais aprofundada destas duas técnicas, recomendamos os papers [10], [11], sobre silhuetas de persistência, e [12], sobre a entropia persistente.

Começemos discutindo as silhuetas de persistência. Seja $\mathcal{D}$ um diagrama de persistência finito (conjunto não nulo), então dizemos que as curvas das silhuetas de persistência de $\mathcal{D}$ são as funções $s_{\mathcal{D}} : \mathbb{R} \rightarrow \mathbb{R}$ definidas por

$$s_{\mathcal{D}}(x) = \frac{\sum_{(t_0, t_1) \in \mathcal{D}; t_1 < \infty} w_{(t_0, t_1)} \Lambda_{(t_0, t_1)}(x)}{\sum_{(t_0, t_1) \in \mathcal{D}; t_1 < \infty} w_{(t_0, t_1)}}, \quad (49)$$

com $w_{(t_0, t_1)} = 1$ e $\Lambda_{(t_0, t_1)}$ a função real

$$\Lambda_{(t_0, t_1)}(x) = \begin{cases} 0 & \text{se } x \leq t_0\sqrt{2} \text{ ou } x \geq t_1\sqrt{2}; \\ t - t_0\sqrt{2} & \text{se } x > t_0\sqrt{2} \text{ e } x \leq (t_0 + t_1)\sqrt{2}/2; \\ t_1\sqrt{2} - t & \text{se } x > (t_0 + t_1)\sqrt{2}/2 \text{ e } x \leq t_1\sqrt{2}, \end{cases}$$

que nada mais é que o triângulo isósceles da Figura 48 calculado com relação a diagonal do $\mathbb{R}^2$.

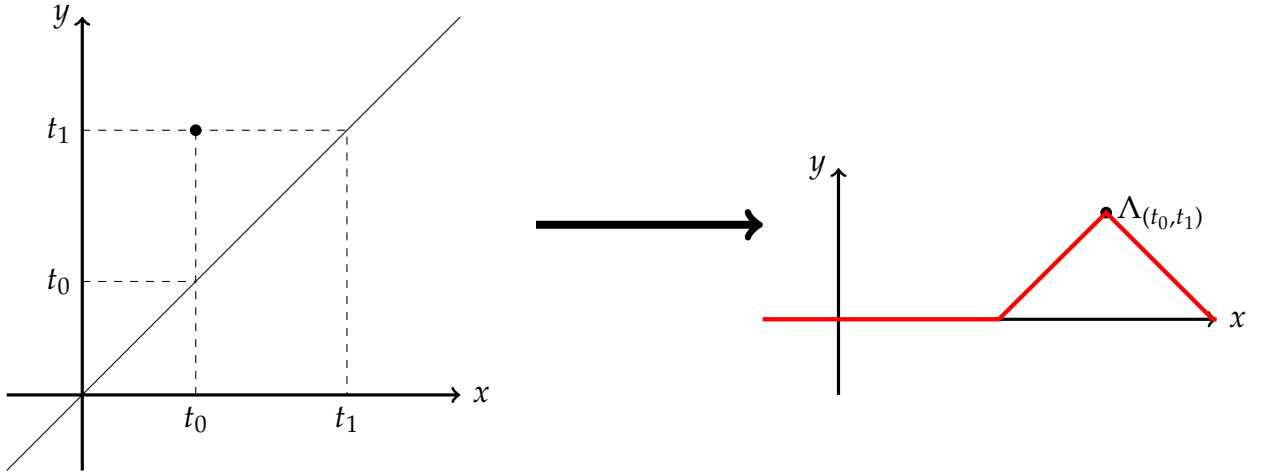


Figura 48: Representação das funções $\Lambda_{(t_0, t_1)}$, com (t_0, t_1) um ponto do diagrama de persistência $\mathcal{D}$, com $t_1 < \infty$. Essas curvas são os triângulos isósceles descritos na base formada pela diagonal do $\mathbb{R}^2$ e o vetor ortogonal a diagonal.

Note que as curvas de silhueta podem ser reescritas da seguinte maneira:

$$s_{\mathcal{D}}(x) = \frac{\sum_{(t_0, t_1) \in \mathcal{D}; t_1 < \infty} \Lambda_{(t_0, t_1)}(x)}{\text{card}\{(t_0, t_1) \in \mathcal{D}; t_1 < \infty\}}, \quad (50)$$

representando uma média das curvas induzidas pelos triângulos isósceles formados por cada ponto do diagrama de persistência.

Observação 4.6.2. Recordemos que estamos considerando apenas as propriedades topológicas que nascem e desaparecem em tempo finito. Casos em que ciclos permanecem até o infinito são ignorados em nossa análise.

Aplicando a silhueta de persistência aos diagramas de persistência das Figuras 46 e 47 obtemos os gráficos da Figura 49, para para propriedades de dimensão 0, e os gráficos da Figura 50, para para propriedades de dimensão 1. Destas curvas já conseguimos resumir melhor o comportamento dos diagramas de persistência e, a seguir, selecionaremos as características mais marcantes destas curvas, para realizarmos nossos testes.

Agora iremos introduzir a última ferramenta que iremos utilizar: a entropia de persistência. Novamente, consideremos um diagrama de persistência $\mathcal{D}$ finito e formados

por pontos (t_0, t_1) , com $t_1 < \infty$. Então a curva da entropia de persistência do diagrama $\mathcal{D}$ será a função real

$$E_{\mathcal{D}}(t) = - \sum_{(t_0, t_1) \in \mathcal{D}; t_1 \leq t} p_{(t_0, t_1)} \log(p_{(t_0, t_1)}), \quad (51)$$

com

$$p_{(t_0, t_1)} = \frac{t_1 - t_0}{S_t} \quad \text{e} \quad S_t = \sum_{(t_0, t_1) \in \mathcal{D}; t_1 \leq t} (t_1 - t_0).$$

Aplicando a definição de entropia persistente aos nossos diagramas obtemos os gráficos da Figura 51 e 52, separados pelas dimensões das homologias.

4.6.4 Extraíndo as informações das silhuetas e da entropia

A fim de agruparmos os nossos grafos de interação, nos clusters da Tabela 3, precisamos realizar mais um passo: extrair as características marcantes das silhuetas e entropias, tal que sejamos capazes de condensar essas informações no espaço euclidiano $\mathbb{R}^n$. Já no espaço euclidiano final, aproveitaremos, de suas boas propriedades métricas, para aplicarmos algoritmos de agrupamento nestes pontos, comparando, depois, com a classificação oriunda do Botometer (próximas seções).

Fixada uma dimensão, 0 ou 1, iremos associar a cada silhueta de persistência os seguintes pontos:

- (i) **pMax**: o ponto máximo da silhueta de persistência de dimensão d ;
- (ii) **support**: o suporte (valor aproximado) da silhueta de persistência de dimensão d ;
- (iii) **area**: a área da silhueta de persistência de dimensão d .

Todos os pontos acima serão tomando em consideração as silhuetas de persistência na escala logarítmica.

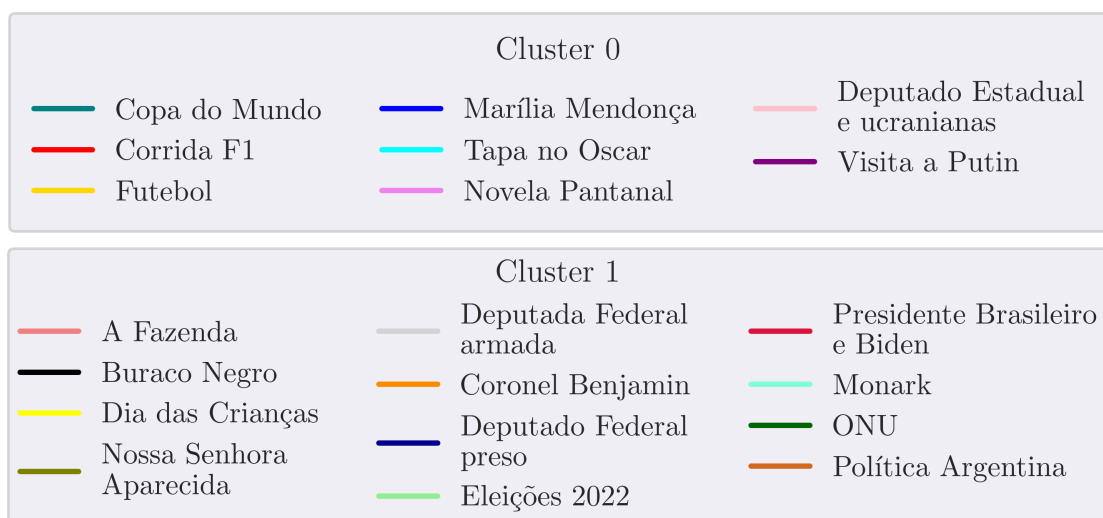
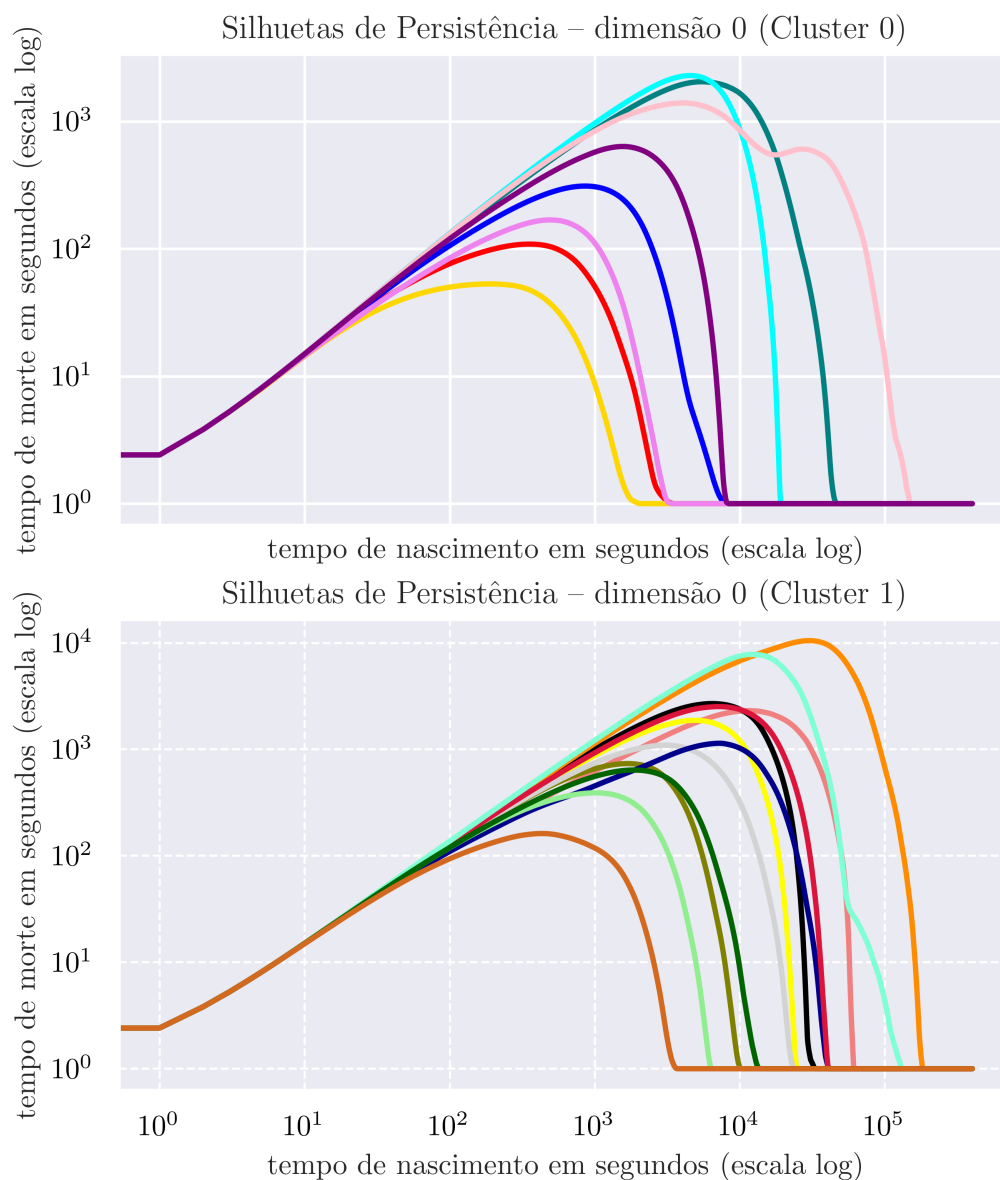


Figura 49: Silhuetas de persistência de dimensão 0.

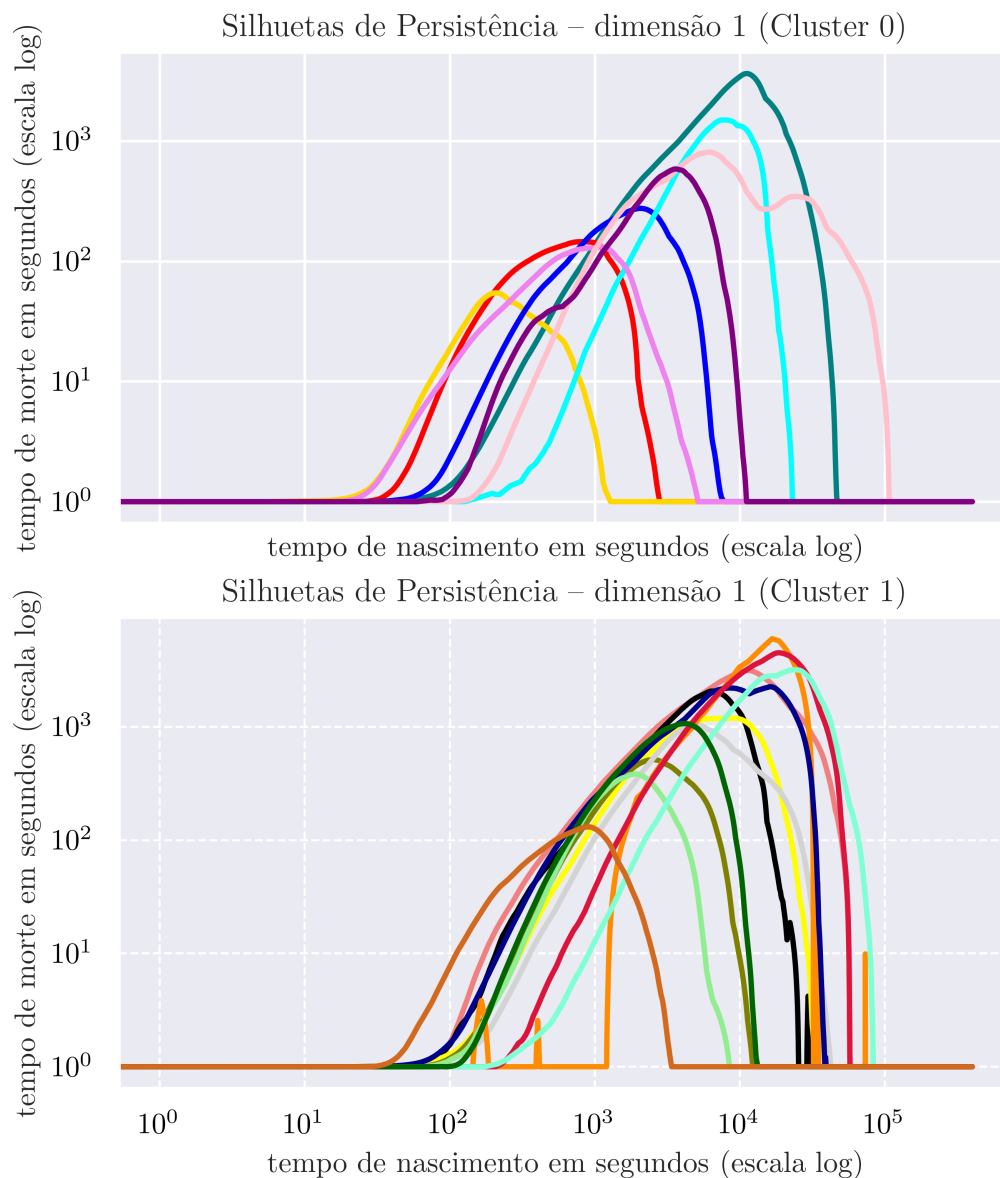


Figura 50: Silhuetas de persistência de dimensão 1.

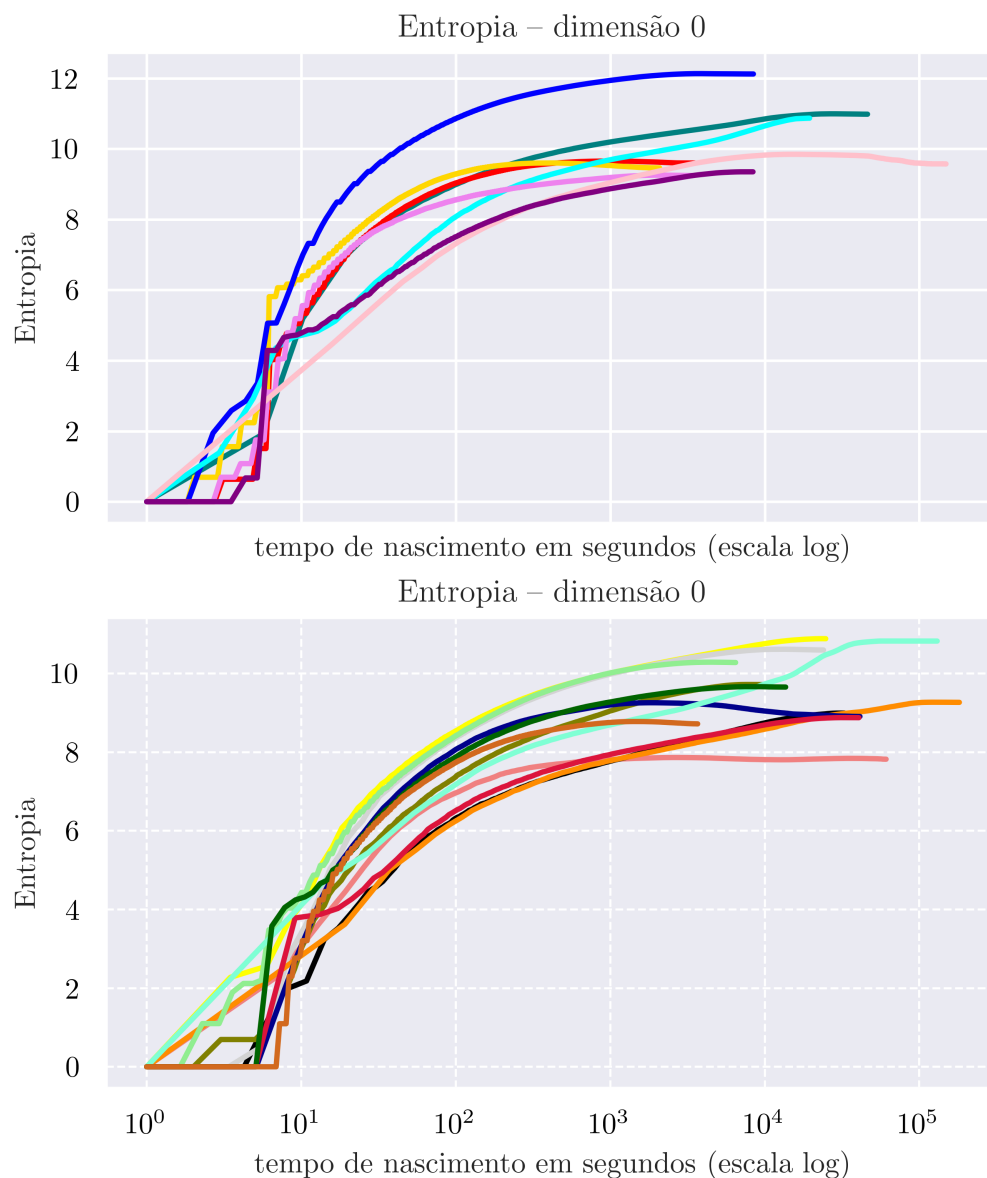


Figura 51: Entropia de persistência de dimensão 0.

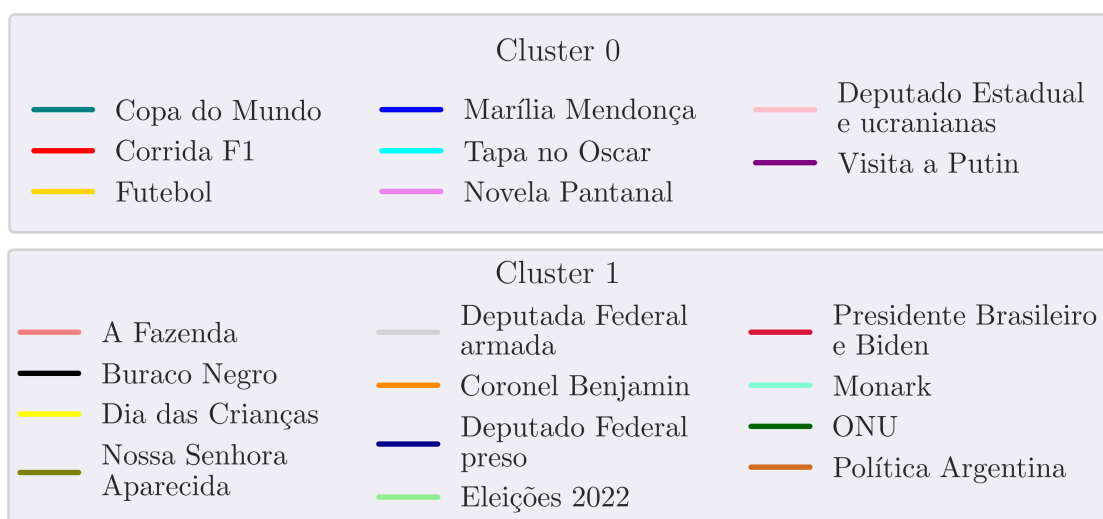
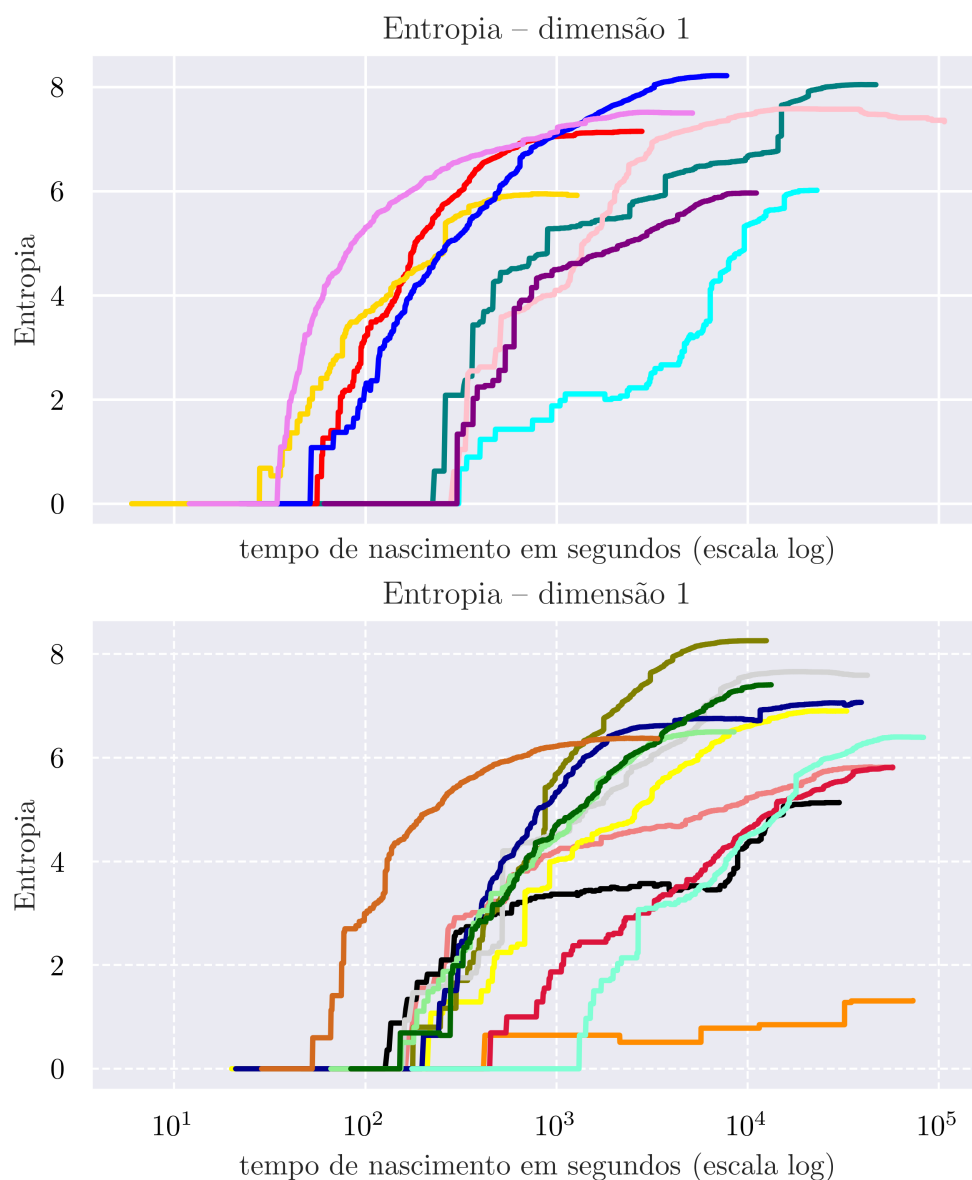


Figura 52: Entropia de persistência de dimensão 1.

Já para as entropias de persistência, consideraremos os valores

$$\text{entropia} = - \sum_{(t_0, t_1) \in \mathcal{D}; t_1 < \infty} p_{(t_0, t_1)} \log(p_{(t_0, t_1)}),$$

com

$$p_{(t_0, t_1)} = \frac{t_1 - t_0}{S_t} \quad \text{e} \quad S_t = \sum_{(t_0, t_1) \in \mathcal{D}; t_1 < t} (t_1 - t_0),$$

e $\mathcal{D}$ o diagrama de persistência do grafo $\hat{G}^i$. Resultante destas transformações, teremos que para cada índice $i = \{1, 2, \dots, 20\}$, nós basearemos na seguinte transformação:

$$\hat{G}^i \rightarrow (\text{pMax}_0, \text{support}_0, \text{area}_0, \text{pMax}_1, \text{support}_1, \text{area}_1, \text{entropia}_0, \text{entropia}_1) \in \mathbb{R}^8.$$

Na Tabela 4 apresentamos os valores extraídos das silhuetas e das curvas de entropia. Notemos, da Figura 53, que os pontos de máximo, o suporte, e a área das curvas, das silhuetas de persistência, são fortemente correlacionados, sendo um forte indicativo de que podemos, se quisermos, reduzir a dimensão da transformação acima.

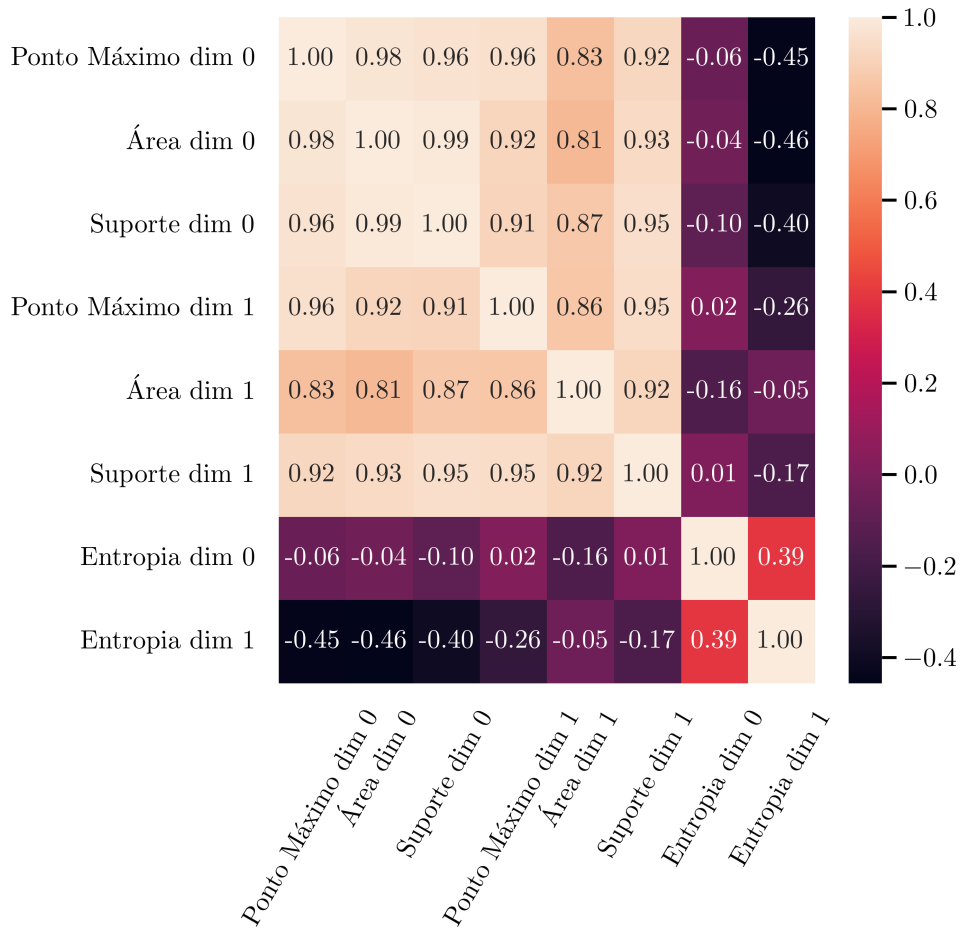


Figura 53: Matriz correlação dos dados extraídos das silhuetas e das curvas de entropia..

Observando a alta correlação entre os pontos máximos, o suporte e a área das silhuetas, para as dimensões 0 e 1, decidimos explorar cada uma dessas características, variando apenas suas dimensões.

	Grafo	pMax ₀	support ₀	area ₀	pMax ₁	support ₁	area ₁	entropia ₀	entropia ₁
Cluster 0	$\hat{G}^3$	3.19	3.75	6.95	3.56	3.90	3.51	10.98	8.05
	$\hat{G}^5$	2.55	3.24	4.72	2.92	3.29	2.71	9.60	7.15
	$\hat{G}^6$	2.28	3.06	3.93	2.32	2.97	1.84	9.46	5.92
	$\hat{G}^7$	2.93	3.52	6.08	3.31	3.73	3.22	12.12	8.22
	$\hat{G}^9$	2.63	3.37	5.25	2.96	3.37	2.60	10.87	6.02
	$\hat{G}^{10}$	2.69	3.27	5.04	3.04	3.46	3.03	9.24	7.50
	$\hat{G}^{15}$	3.23	3.77	7.19	3.40	3.91	3.54	9.57	7.33
	$\hat{G}^{17}$	3.01	3.60	6.37	3.28	3.70	2.92	9.35	5.97
Cluster 1	$\hat{G}^1$	4.06	4.65	10.18	4.057	4.623	6.79	7.82	5.80
	$\hat{G}^2$	3.81	4.34	9.53	3.81	4.18	5.14	9.00	5.14
	$\hat{G}^4$	3.70	4.23	8.97	3.95	4.34	5.04	10.89	6.90
	$\hat{G}^8$	4.48	4.96	12.75	4.22	4.49	4.32	9.72	8.26
	$\hat{G}^{11}$	3.61	4.83	10.63	3.79	4.81	5.74	10.60	7.59
	$\hat{G}^{12}$	3.26	3.86	7.34	3.62	3.97	4.00	9.27	1.32
	$\hat{G}^{13}$	3.85	4.43	9.79	4.27	4.66	5.60	8.91	7.07
	$\hat{G}^{14}$	3.49	4.12	8.42	3.70	4.41	4.87	10.28	6.50
	$\hat{G}^{16}$	3.66	4.12	8.76	3.88	4.20	3.69	8.88	5.82
	$\hat{G}^{18}$	3.73	4.35	9.56	4.05	4.46	5.86	10.83	6.39
	$\hat{G}^{19}$	3.86	4.38	8.95	4.21	4.50	6.03	9.66	7.41
	$\hat{G}^{20}$	4.08	4.58	11.37	4.36	4.72	5.31	8.72	6.37

Tabela 4: Valores extraídos das silhuetas e das curvas de entropia. Grafos enumerados: $\hat{G}^1$ (A Fazenda); $\hat{G}^2$ (Buraco Negro); $\hat{G}^3$ (Copa do Mundo); $\hat{G}^4$ (Dia das Crianças); $\hat{G}^5$ (Corrida F1); $\hat{G}^6$ (Futebol); $\hat{G}^7$ (Marília Mendonça); $\hat{G}^8$ (Nossa Senhora Aparecida); $\hat{G}^9$ (Tapa no Oscar); $\hat{G}^{10}$ (Novela Pantanal); $\hat{G}^{11}$ (Deputada Federal armada); $\hat{G}^{12}$ (Coronel Benjamin); $\hat{G}^{13}$ (Deputado Federal preso); $\hat{G}^{14}$ (Eleições 2022); $\hat{G}^{15}$ (Deputado Estadual e ucranianas); $\hat{G}^{16}$ (Presidente Brasileiro e Biden); $\hat{G}^{17}$ (Visita a Putin); $\hat{G}^{18}$ (Monark); $\hat{G}^{19}$ (ONU); $\hat{G}^{20}$ (Política Argentina).

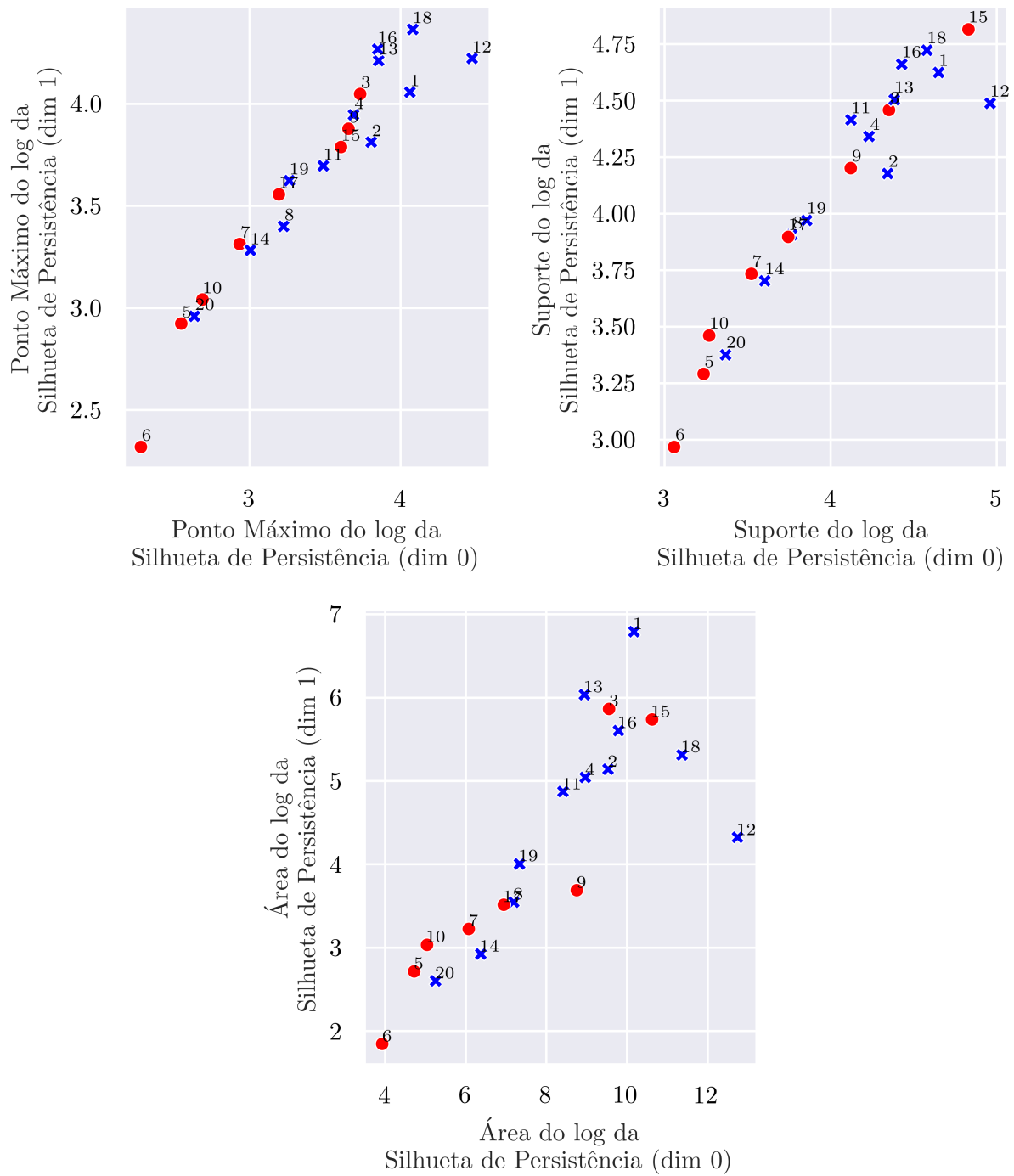


Figura 54: Propriedades de grafos do cluster 0 são coloridos de vermelho, enquanto que as propriedades de grafos do cluster 1 são coloridos com azul. Na primeira linha, da esquerda para à direita: Pontos máximos das silhuetas de persistência de cada grafo, de dimensão 0 e 1; Suporte das silhuetas de persistência de cada grafo, de dimensão 0 e 1. Na segunda linha: Área das silhuetas de persistência de cada grafo, de dimensão 0 e 1. Grafos enumerados: G^1 (A Fazenda); G^2 (Buraco Negro); G^3 (Copa do Mundo); G^4 (Dia das Crianças); G^5 (Corrida F1); G^6 (Futebol); G^7 (Marília Mendonça); G^8 (Nossa Senhora Aparecida); G^9 (Tapa no Oscar); G^{10} (Novela Pantanal); G^{11} (Deputada Federal armada); G^{12} (Coronel Benjamin); G^{13} (Deputado Federal preso); G^{14} (Eleições 2022); G^{15} (Deputado Estadual e ucranianas); G^{16} (Presidente Brasileiro e Biden); G^{17} (Visita a Putin); G^{18} (Monark); G^{19} (ONU); G^{20} (Política Argentina).

Dos gráficos da Figura 54, podemos observar que tanto os pontos máximos das silhuetas de persistência, quanto seus suportes, tem comportamento médio aproximadamente linear, para as dimensões zero e um. Quando olhamos para a área destas curvas, esse comportamento não fica tão claro.

4.6.5 Conclusão

Quando aplicamos a técnica de agrupamento *K-means* aos dados da subseção anterior, temos que as características de dimensão 0 e 1 das propriedades extraídas das curvas das silhuetas de persistência, e das entropias de persistência, fornecem os seguintes agrupamentos (que apresentamos no início da seção):

Cluster 0	Corrida F1, Novela Pantanal, Futebol, Marília Mendonça, Copa do Mundo, Tapa no Oscar, Visita a Putin, Deputado Estadual e ucranianas
Cluster 1	A Fazenda, Buraco Negro, Dia Das Crianças, Nossa Senhora Aparecida, Política na Argentina, Coronel Benjamin, Eleições 2022, Monark, ONU, Deputada Federal armada, Deputado Federal preso, Presidente Brasileiro e Biden

Tabela 5: Grafos classificados segundo suas propriedades topológicas pelo K-means.

Avaliando o agrupamento acima conseguimos, com o auxílio dos gráficos das subseções anteriores, detectar alguns padrões, são eles:

- (i) As áreas das curvas das silhuetas de persistência, para as dimensões zero e um, tem uma grande influência no agrupamento fornecido pela *K-means*, conforme vemos no gráfico da Figura 54. Deste gráfico, conseguimos separar os grafos G^5 (Corrida F1), G^6 (Futebol), G^7 (Marília Mendonça), G^9 (Tapa no Oscar), G^{10} (Novela Pantanal) e G^{17} (Visita a Putin), classificados como pertencentes ao cluster 0, dos grafos G^1 (A Fazenda), G^2 (Buraco Negro), G^4 (Dia das Crianças), G^{11} (Deputada Federal armada), G^{12} (Coronel Benjamin), G^{13} (Deputado Federal armado) e G^{18} (Monark), classificados como pertencentes ao cluster 1;

- (ii) Quando olhamos para as curvas das entropias de persistência, Figura 51 e 52, percebemos que o agrupamento, resultante do algoritmo *K-means*, captura diferenças na curva média das entropias. Todavia, isso só se observa, mais claramente, quando olhamos para a dimensão 1. Já para as entropias de dimensão zero não percebemos diferenças nos agrupamentos.

Vale a pena salientarmos, também, que a classificação de alguns grafos fogem do que o nosso senso comum imaginaria. Por exemplo, no começo deste estudo, não imaginávamos que grafos cujos tweets publicados, ligados a esporte, teriam comportamento semelhante a grafos associados a tópicos sobre política.

4.7 CLASSIFICAÇÃO DOS TREND TOPICS USANDO O BOTÔMETRO

Após estabelecermos a metodologia de agrupamento, dos grafos de interação dos usuários da Seção 4.3, onde as propriedades topológicas eram pontos centrais para a caracterização dos clusters, apresentaremos nesta seção uma outra metodologia, cujo objetivo é servir de comparativo aos resultados obtidos a partir da Homologia Persistente, Tabela 3. Esta metodologia alternativa será baseada no seguinte critério: inicialmente, para cada grafo amostrado $T_{(t_0^i, t_f^i, K^i)}$, $i \in \{1, \dots, 20\}$, selecionaremos 200 tweets ao acaso (seguindo uma uniforme), que servirão para resumirmos o perfil dos usuários do grafo em consideração; em seguida, atribuiremos aos criadores destes tweets uma nota, variando de 0 a 5, onde notas próximas de zero indicam que o usuário tem baixa chance de ser, de fato, um robô, e notas altas indicam maior chance do usuário ser uma máquina, programada pra fingir interações na rede social.

A partir do momento que para cada grafo $T_{(t_0^i, t_f^i, K^i)}$, $i \in \{1, \dots, 20\}$, somos capazes de estimar uma distribuição de probabilidade para as notas dos perfis dos usuários, por meio de seus histogramas, utilizaremos os vetores de frequências destes histogramas para compararmos cada grafo. Considerando que cada grafo pode ser visto como um ponto do espaço uclideo $\mathbb{R}^n$, agruparemos os grafos em dois clusters distintos, utilizando, tal qual fizemos na seção anterior, o método de *K-means* para as classificações. Posteriormente, utilizaremos os agrupamentos obtidos nesta seção para avaliarmos as similaridades entre os resultados obtidos desta seção, com os da seção anterior.

4.7.1 Resultados

Para atribuírmos notas a cada usuário, dos 200 amostrados por grafo, necessitaríamos de um estudo do comportamento histórico deles, o qual deveríamos levar em consideração o conteúdo dos tweets publicados por eles nessas janelas de tempo. Todavia, esse tipo de abordagem divergeria do nosso propósito inicial. Ao invés disso, usaremos para nosso auxílio um projeto já consolidada no mercado, o *Botômetro* (em inglês *Botometer*⁸). Este projeto nos fornecerá exatamente o que precisamos, uma plataforma que, dado um usuário qualquer, retorna uma nota indicando a chance do usuário ter perfil humano ou robô. Dos dados obtidos pelo Botômetro, apresentados na próxima seção, somos capazes de propor o agrupamento em dois clusters (usando o método *K-means*), conforme podemos ver na Tabela 6.⁹:

Cluster 0	Corrida F1, Novela Pantanal, Futebol, Buraco Negro, Marília Mendonça, Tapa no Oscar Política na Argentina, Monark
Cluster 1	Copa do Mundo, Visita a Putin, Deputado Estadual e ucranianas A Fazenda, Dia Das Crianças, Nossa Senhora Aparecida Coronel Benjamin, Eleições 2022, Presidente Brasileiro e Biden ONU, Deputada Federal armada, Deputado Federal preso

Tabela 6: Grafos classificados pelo *K-means*, tomando como base as distribuições de probabilidade estimadas pelos histogramas dos perfis dos usuários, medidos segundo o Botômetro.

4.7.2 Dados do Botômetro

Nesta seção apresentamos os vetores de frequência dos histogramas ajustados para perfis dos usuários amostrados para cada grafo, seguindo os critérios do *Botômetro*. Notemos que as notas atribuídas, a cada usuário, são números reais variando no intervalo $[0, 5]$.

⁸ <https://botometer.osome.iu.edu/>

⁹ Tal qual fizemos na exposição dos resultados da Homologia Persistente de Caminhos, começaremos apresentando os resultados finais dos agrupamentos pelo *K-Means*. Fazemos isso pois, dessa maneira, conseguimos apresentar os tratamentos dos dados de uma maneira a elucidar melhor os padrões dos clusters.

Do intervalo, em que as notas de perfil dos usuários podem variar, convencionamos de particionar este intervalo em 15 partes de comprimento igual, medindo a quantidade de notas em cada intervalo (procedimento para gerar os histogramas). Na Tabela 7 apresentamos os vetores de frequência de cada grafo nosso.

Observação 4.7.1. Dos vetores de frequência da Tabela 7, podemos observar que alguns casos a soma das frequências nem sempre contabiliza o valor total de usuários amostrados, valor este igual a 200. Isto ocorre devido ao usuário amostrado ter tido sua conta excluída de alguma forma, acarretando na falta de uma nota para seu perfil. Isto não acaba sendo um problema, pois depois normalizamos estes valores.

Grafo	vetor de frequências $\mathbb{R}^{15}$
G^1 (A Fazenda)	(4, 0, 0, 13, 21, 14, 18, 10, 10, 7, 20, 29, 31, 0, 0)
G^2 (Buraco Negro)	(2, 0, 0, 12, 24, 18, 20, 8, 18, 10, 14, 26, 24, 6, 1)
G^3 (Copa do Mundo)	(1, 0, 0, 20, 13, 13, 22, 8, 12, 16, 13, 23, 39, 4, 2)
G^4 (Dia das Crianças)	(1, 0, 0, 4, 12, 16, 16, 9, 16, 7, 19, 40, 25, 7, 1)
G^5 (Corrida F1)	(6, 0, 0, 25, 31, 15, 15, 12, 14, 10, 6, 17, 8, 2, 0)
G^6 (Futebol)	(4, 0, 0, 22, 12, 20, 19, 10, 12, 14, 10, 17, 15, 2, 0)
G^7 (Marília Mendonça)	(2, 0, 0, 11, 24, 27, 18, 6, 8, 4, 9, 13, 13, 2, 3)
G^8 (Nossa Senhora Aparecida)	(4, 0, 0, 16, 13, 12, 20, 19, 12, 13, 14, 29, 27, 2, 3)
G^9 (Tapa no Oscar)	(0, 0, 0, 13, 24, 22, 34, 16, 19, 18, 13, 11, 12, 2, 1)
G^{10} (Novela Pantanal)	(3, 0, 0, 23, 16, 13, 32, 6, 16, 16, 18, 21, 20, 1, 0)
G^{11} (Deputada Federal armada)	(0, 0, 0, 5, 12, 8, 14, 13, 16, 13, 19, 45, 38, 2, 0)
G^{12} (Coronel Benjamin)	(1, 0, 0, 4, 7, 11, 16, 6, 20, 11, 13, 46, 38, 7, 1)
G^{13} (Deputado Federal preso)	(2, 0, 0, 8, 18, 11, 21, 7, 13, 12, 11, 44, 41, 5, 1)
G^{14} (Eleições 2022)	(1, 0, 0, 10, 14, 15, 15, 11, 17, 12, 18, 34, 32, 4, 0)
G^{15} (Deputado Estadual e ucranianas)	(4, 0, 0, 9, 13, 14, 24, 12, 10, 18, 19, 25, 37, 4, 0)
G^{16} (Presidente Brasileiro e Biden)	(2, 0, 0, 4, 7, 7, 14, 9, 16, 12, 10, 37, 56, 6, 1)
G^{17} (Visita a Putin)	(0, 0, 0, 8, 15, 9, 12, 8, 13, 12, 16, 23, 47, 3, 1)
G^{18} (Monark)	(5, 0, 0, 20, 26, 12, 22, 7, 19, 12, 9, 21, 26, 2, 0)
G^{19} (ONU)	(1, 0, 0, 11, 11, 14, 22, 9, 7, 12, 11, 25, 32, 5, 2)
G^{20} (Política Argentina)	(2, 0, 0, 10, 15, 17, 27, 8, 24, 17, 12, 18, 23, 1, 0)

Tabela 7: Dados do Botômetro. Para cada linha temos o vetor de frequência dos histogramas das notas atribuídas aos usuários amostrados de cada grafo.

4.7.3 Histogramas

Seguindo os dados da Tabela 7, apresentamos os histogramas da Figura 55, de todos os grafos agrupados no cluster 0, e apresentamos os histogramas da Figura 56, de todos os grafos agrupados no cluster 1. Para cada histograma também ilustramos suas respectivas funções de densidade de probabilidade, aproximadas usando o método do kernel.

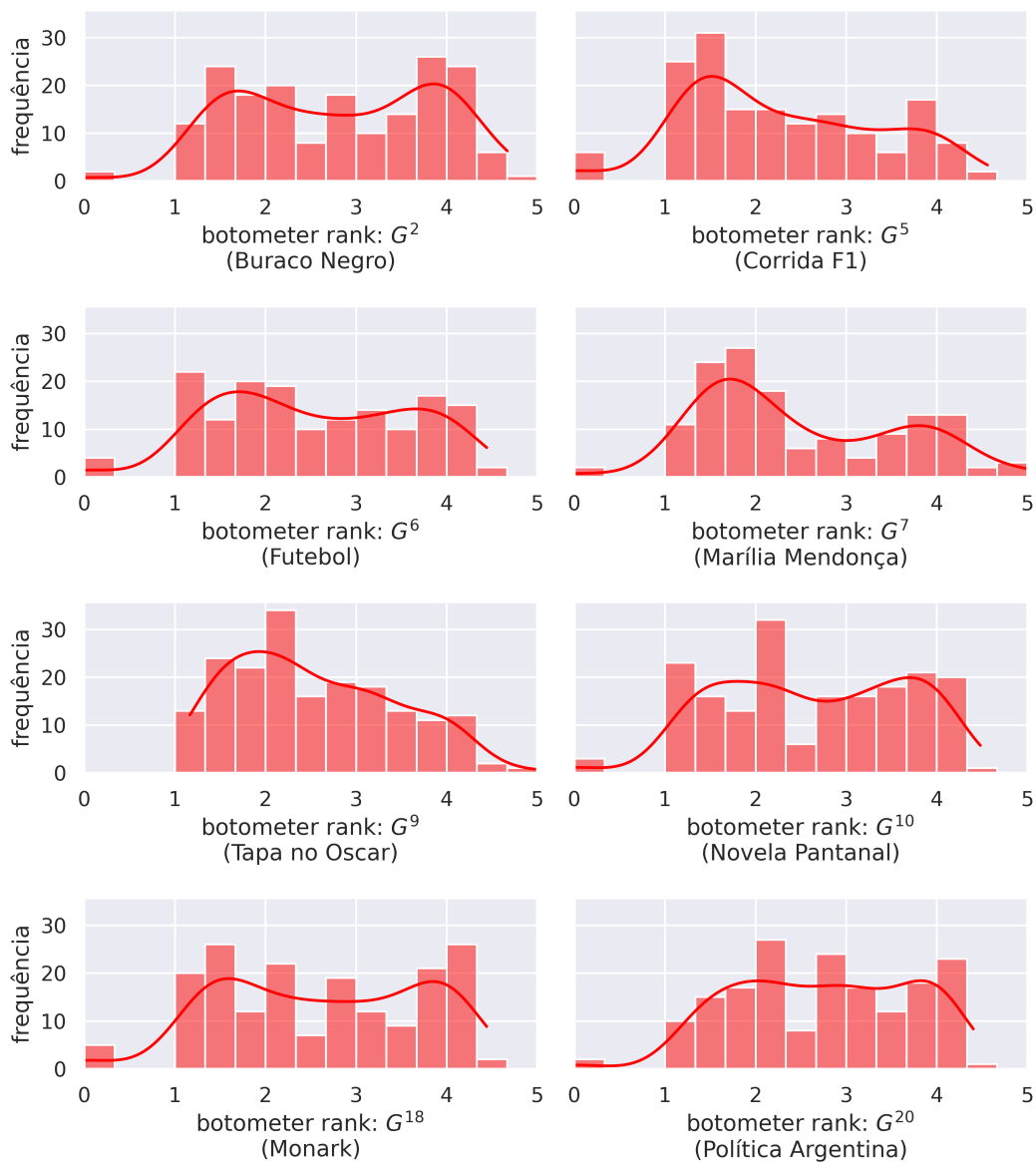


Figura 55: Histogramas das classificações dos usuários de cada grafo, seguindo o critério de classificação do Botômetro. (Cluster 0)

4.7.4 Conclusão

Encerraremos esta seção discutindo o resultado final, vide Tabela 6, do agrupamento resultante da aplicação do método K-means aos vetores de frequência dos histogramas das Figuras 55 e 56. Deste agrupamento, salientamos alguns pontos, começando pela curva da densidade estimada para os perfis de usuários:

- (i) Os histogramas dos perfis de usuários, dos grafos do cluster 0, constituem-se de três grafos que claramente possuem concentração de massa em perfis com nota menor que 2,5. Estes grafos são: G^4 (Corrida 1); G^7 (Marília Mendonça); G^9 (Tapa no Oscar). Os demais histogramas deste agrupamento, a priori, não aparentam ter uma concentração de massa bem definida, como os outros três histogramas citados;
- (ii) Já quando olhamos os histogramas dos perfis de usuários, dos grafos do cluster 1, conseguimos destacar oito grafos com concentração de massa em perfis com nota maior que 2,5, são eles: G^{11} (Deputada Federal armada); G^{12} (Coronel Benjamin); G^{13} (Deputada Federal preso); G^{14} (Eleições 2022); G^{15} (Deputada Estadual e ucranianas); G^{16} (Presidente Brasileiro e Biden); G^{17} (Visita a Putin); Finalmente, para os demais histogramas deste agrupamento, não é aparente uma concentração de massa bem definida, como os outros histogramas citados.

4.8 COMPARAÇÃO

Chegamos a parte final deste capítulo, onde avaliaremos como a classificação de agrupamento, utilizando a Homologia Persistente, se compara ao agrupamento de controle, resultante da análise das notas de perfil dos usuários, obtidas a partir do projeto *Botometer*.

Nossa primeira medida de comparação será uma medida de similaridade, que comparará

$$\mathcal{C}_{\text{Hom}} : \{\widehat{G}^i; i = 1, 2, \dots, 20\} \rightarrow \mathbb{Z}_2 \quad \text{e} \quad \mathcal{C}_{\text{Bot}} : \{\widehat{G}^i; i = 1, 2, \dots, 20\} \rightarrow \mathbb{Z}_2,$$

com $\mathcal{C}_{\text{Hom}}$ o classificador dos dados da Homologia Persistente, e $\mathcal{C}_{\text{Bot}}$ o classificador baseado nos dados de perfil dos usuários, obtido a partir do Botômetro.

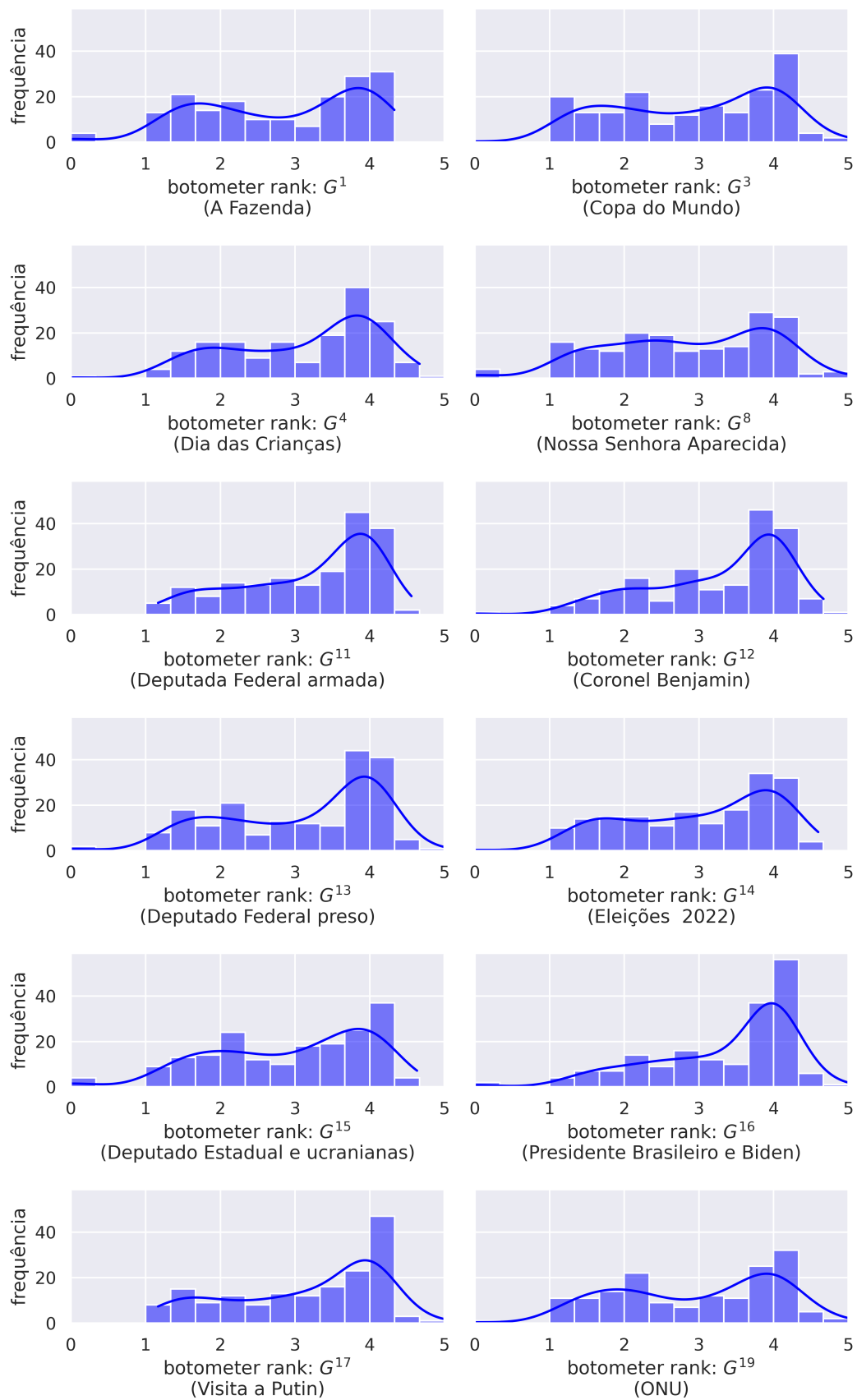


Figura 56: Histogramas das classificações dos usuários de cada grafo, seguindo o critério de classificação do Botômetro. (Cluster 1)

Os classificadores $\mathcal{C}_{\text{Hom}}$ e $\mathcal{C}_{\text{Bot}}$, definidos acima, seguem o já esperado, tendo seus valores de acordo com as duas seções anteriores: o classificador baseado na homologia persistente, $\mathcal{C}_{\text{Hom}}$, assumirá os valores da Tabela 5; já o classificador baseado no Botômetro, $\mathcal{C}_{\text{Bot}}$, assumirá os valores da Tabela 6. Para estes dois classificadores, definiremos a seguinte medida de *similaridade*:

$$\text{similaridade}(\mathcal{C}_{\text{Hom}}, \mathcal{C}_{\text{Bot}}) = \frac{\max_{\pi \in S^2} \left\{ \sum_{i=1}^{20} \mathbb{1} \left(\mathcal{C}_{\text{Hom}}(\widehat{G}^i) = \pi(\mathcal{C}_{\text{Hom}}(\widehat{G}^i)) \right) \right\}}{20},$$

com S^2 o grupo das simetrias. A medida de similaridade entre os nossos dois classificadores é exibida na Tabela 8, com valor geral igual a

$$\text{similaridade}(\mathcal{C}_{\text{Hom}}, \mathcal{C}_{\text{Bot}}) = 14/20 = 70\%.$$

Este valor de similaridade nos garante que, dentro de uma margem de segurança, os métodos coincidem entre si, e os classificadores estão detectando estruturas dos grafos em comum.

Continuando com nosso comparativo, avaliaremos quais classificadores apresentam maior assertividade em suas decisões de agrupamento. Para isto, recordemos o seguinte ponto: nossos classificadores $\mathcal{C}_{\text{Hom}}$ e $\mathcal{C}_{\text{Bot}}$ utilizam-se do mesmo algoritmo de clusterização, o K-means, vindo a divergir apenas na construção de seus espaços amostrais. Desta forma, seja $x \in \mathbb{R}^8$ um ponto qualquer, nossos classificadores funcionam da seguinte maneira:

$$\mathcal{C}_{\text{Hom}}(x) = \arg \min_{i=0,1} d \left(x, \text{centróide}_{\text{Hom}}^i \right)$$

e para $x \in \mathbb{R}^{15}$

$$\mathcal{C}_{\text{Bot}}(x) = \arg \min_{i=0,1} d \left(x, \text{centróide}_{\text{Bot}}^i \right),$$

onde o centróides de cada cluster, para cada classificador, são resultantes do K-means aplicado a nossa amostra.

Das distâncias dos pontos amostrais, aos respectivos centróides de seus clusters, temos que quanto maior a proximidade de um ponto ao centróide, maior a nossa *certeza* de que o classificador está tomando a decisão correta. Dos gráficos da Figura 57, apresentamos as distâncias dos pontos aos centróides 0 e 1, de cada um dos nossos classificadores.

Grafo	Classificação Homologia	Classificação Botômetro
G^1 (A Fazenda)	1	1
G^2 (Buraco Negro)	0	1
G^3 (Copa do Mundo)	1	0
G^4 (Dia das Crianças)	0	0
G^5 (Corrida F1)	1	1
G^6 (Futebol)	1	1
G^7 (Marília Mendonça)	1	1
G^8 (Nossa Senhora Aparecida)	0	0
G^9 (Tapa no Oscar)	1	1
G^{10} (Novela Pantanal)	1	1
G^{11} (Deputada Federal armada)	0	0
G^{12} (Coronel Benjamin)	0	0
G^{13} (Deputado Federal preso)	0	0
G^{14} (Eleições 2022)	0	0
G^{15} (Deputado Estadual e ucranianas)	1	0
G^{16} (Presidente Brasileiro e Biden)	0	0
G^{17} (Visita a Putin)	1	0
G^{18} (Monark)	0	1
G^{19} (ONU)	0	0
G^{20} (Política Argentina)	0	1
similaridade($\mathcal{C}_{\text{Hom}}, \mathcal{C}_{\text{Bot}}$) = 14/20		

Tabela 8: Medida de similaridade das duas metodologias de agrupamento de grafos: $\mathcal{C}_{\text{Hom}}$ (Homologia Persistente) e $\mathcal{C}_{\text{Bot}}$ (Botômetro). Linhas em cinza representam as classificações discordantes entre os dois classificadores.

Observação 4.8.1. Antes de rodarmos os algoritmos de aprendizagem não supervisionada, para obtermos os classificadores $\mathcal{C}_{\text{Hom}}$ e $\mathcal{C}_{\text{Bot}}$, realizamos a normalização dos nossos dados. Consequentemente, os pontos resultantes da metodologia de transformação dos dados, da Homologia Persistente, são elementos do espaço euclidiano satisfazendo

$$x \in \mathbb{R}^8, \text{ tal que } x \in [0, 1]^8,$$

enquanto que, para os pontos resultantes da análise de perfil dos usuários, utilizando-se o Botômetro, vale

$$x \in \mathbb{R}^{15}, \text{ tal que } x \in [0, 1]^{15}.$$

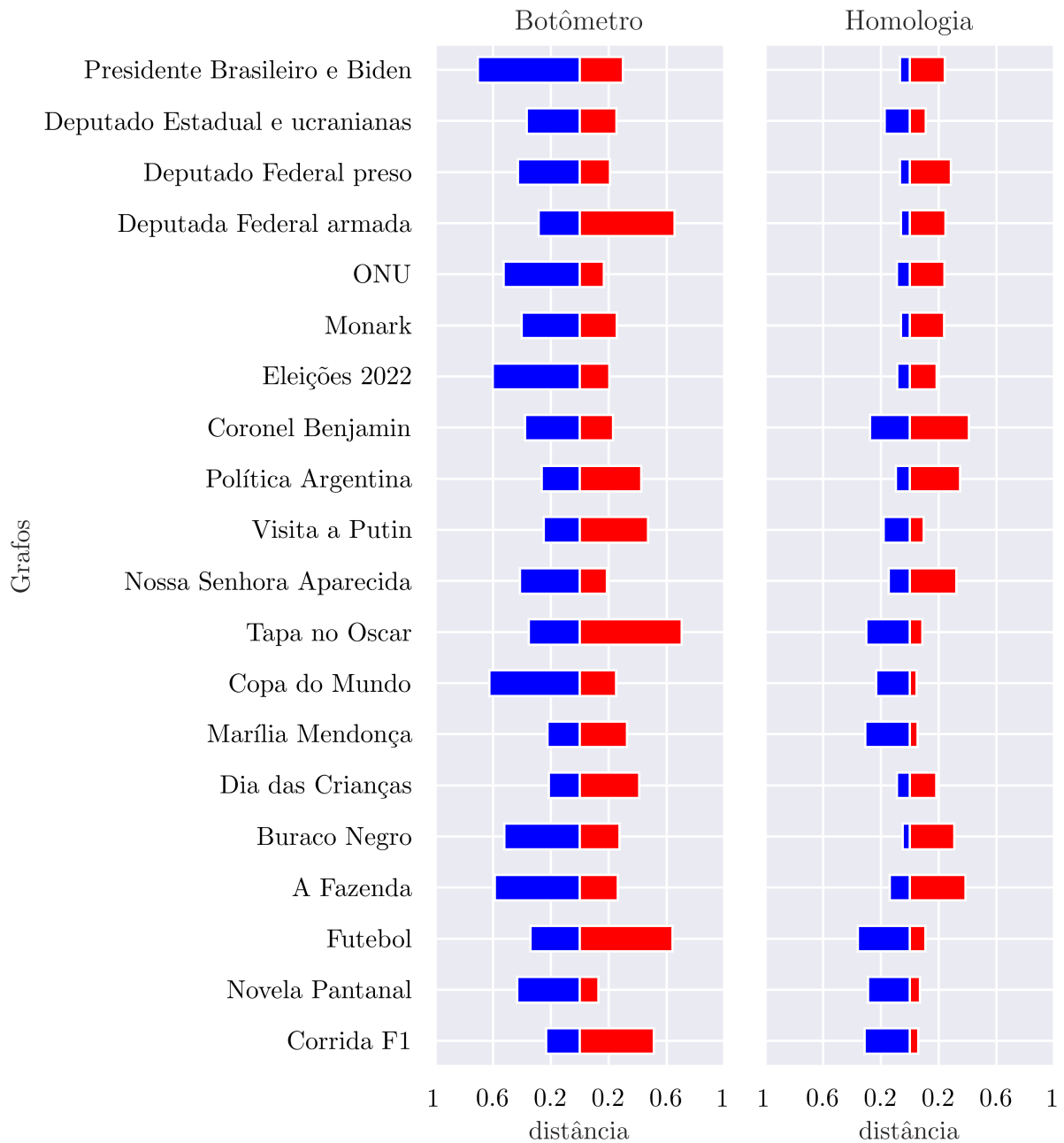


Figura 57: Distâncias dos dados aos centróides dos clustes 0 e 1, tomando em conta os dois classificadores: $\mathcal{C}_{\text{Hom}}$ e $\mathcal{C}_{\text{Bot}}$. O cluster 0 é representado pela barra vermelha ■, enquanto que o cluster 1 é representado pela barra azul ■.

Do observado em 4.8.1, e dos gráficos da Figura 57, observamos que os pontos normalizados da homologia estão, em sua maioria, mais próximos do centróide de seu cluster.

Alguns parágrafos acima havíamos afirmamos que nossos classificadores detêm a seguinte característica: seja x um ponto do espaço amostral qualquer, então o *grau de certeza* que o classificador atribuí um cluster, ao ponto x , está intimamente associado a distância do ponto ao centróide do cluster. A seguir construiremos uma medida para avaliarmos o grau de certeza dos nossos classificadores, que chamaremos de *grau de pertencimento*.

Seja $x \in \Omega$ um ponto qualquer, do espaço vetorial e normado Ω , e $\mathcal{C} : \Omega \rightarrow \mathbb{Z}_2$ um classificador, escolhido seguindo o algoritmo do K-means, i.e.,

$$\mathcal{C}(x) = \arg \min_{i=0,1} d(x, C^i),$$

com C^0 e C^1 os centróides estimados pelo K-means. Então, definimos a medida *grau de pertencimento* do ponto x ao cluster escolhido pelo classificador $\mathcal{C}$, para este ponto, $\mathcal{C}(x)$, como sendo a função

$$\begin{aligned} \omega : \Omega &\rightarrow \mathbb{R} \\ x &\mapsto \frac{|\|x - C^0\| - \|x - C^1\||}{\max \{\|x - C^0\|, \|x - C^1\|\}}. \end{aligned}$$

Observação 4.8.2. Como esperado, a função grau de pertencimento satisfaz a condição

$$\|x - C^0\| \approx \|x - C^1\| \text{ então } \omega(x) \approx 0,$$

enquanto que

$$\|x - C^0\| \gg \|x - C^1\| \text{ então } \omega(x) \approx 1.$$

Ou seja, o classificador $\mathcal{C}$ não tem uma boa assertividade para pontos equidistantes dos centróides. Todavia, conforme mais próximo o ponto é, de um determinado centróide, mais seguro é a classificação de $\mathcal{C}$.

Na Figura 58 apresentamos o grau de pertencimento para os dois classificadores, $\mathcal{C}_{\text{Hom}}$ e $\mathcal{C}_{\text{Bot}}$. Destes valores observamos que, para aproximadamente 15 casos, o classificador baseado na homologia $\mathcal{C}_{\text{Hom}}$ tem uma *assertividade* significativamente maior que o classificador $\mathcal{C}_{\text{Bot}}$. De fato, em 14 vezes o classificador $\mathcal{C}_{\text{Hom}}$ tem uma *certeza de classificação* maior que 60%, o que acontece em apenas 3 casos para o classificador do Botômetro.

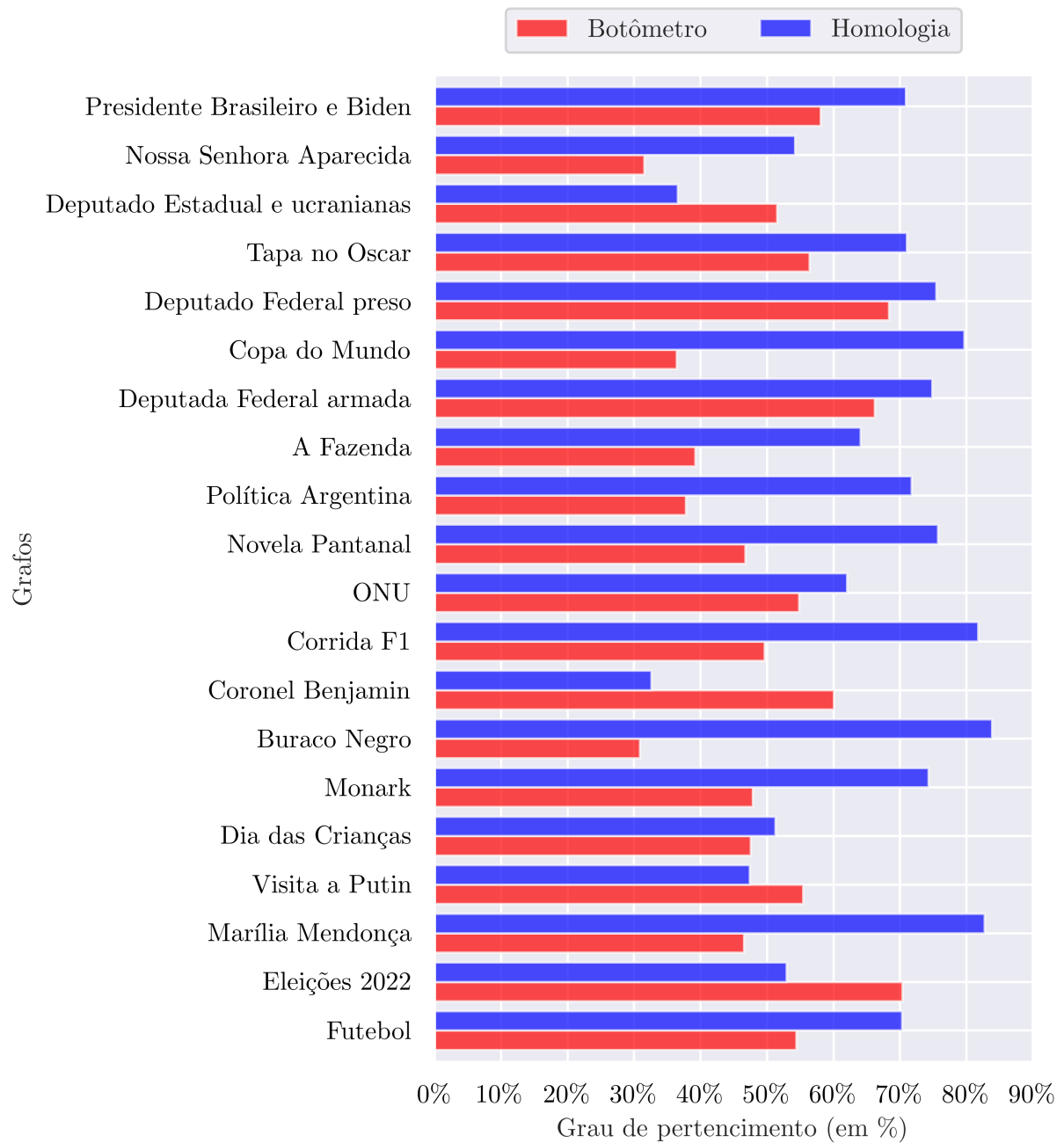
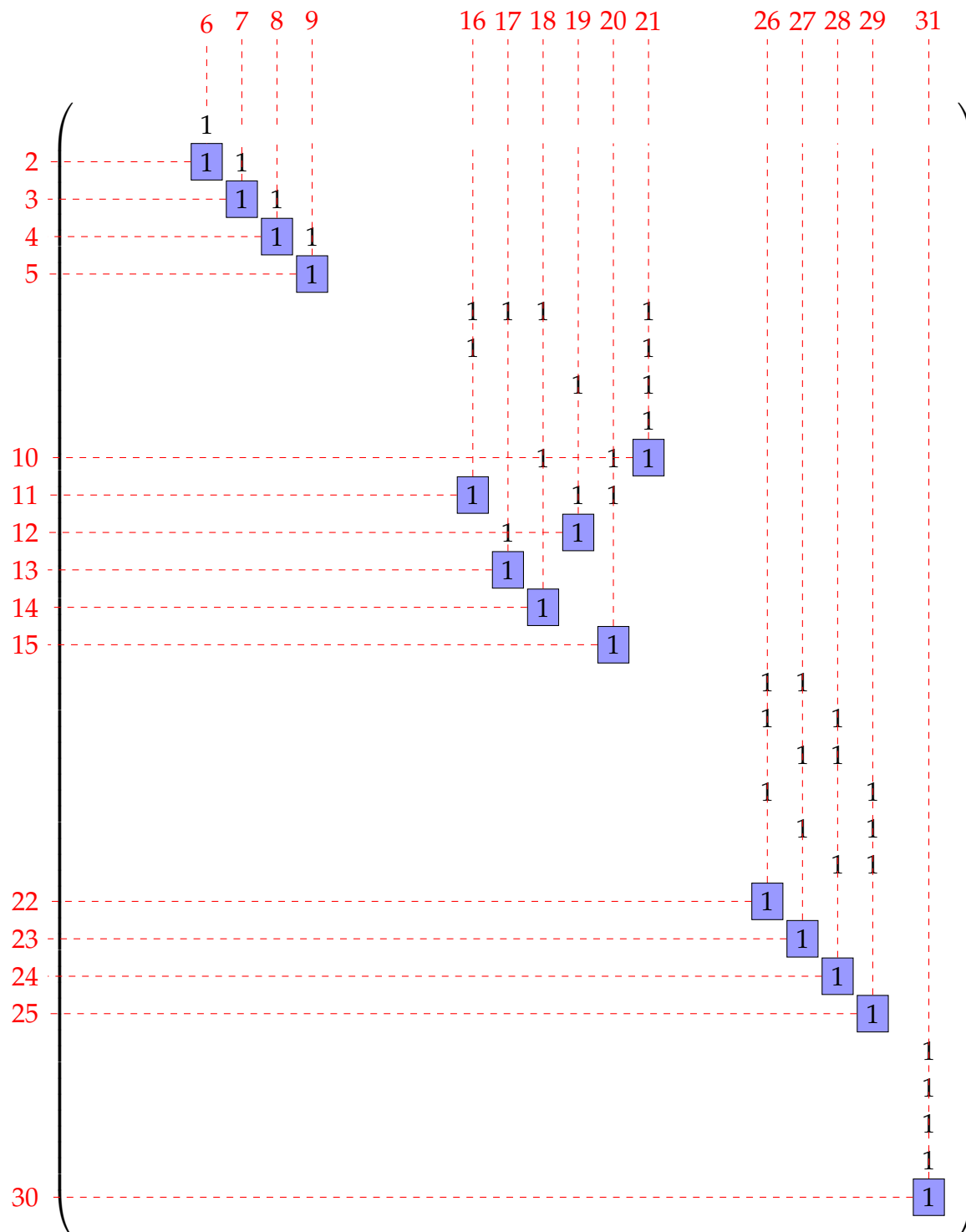


Figura 58: Graus de pertencimento de cada grafo, medidos a partir dos classificadores $\mathcal{C}_{\text{Hom}}$ e $\mathcal{C}_{\text{Bot}}$. Quanto maior o grau de pertencimento medido para um dado grafo G , maior certeza o classificador tem sobre o cluster escolhido para G .

Apêndice

A RESULTADO DO ALGORITMO 2, CAPÍTULO 2, APLICADO AO EXEMPLO 2.5.1



BIBLIOGRAFIA

- [1] Yasuaki H Iraoka, Tomoyuki Shirait, and Khanh Duy Trinh, Limit Theorems For Persistence Diagrams. *The Annals of Applied Probability*, 28(5):2740-2780, 2018.
- [2] D. Yogeshwaran, Eliran Subag, and Robert J. Adler, Random geometric complexes in the thermodynamic regime. *Probab. Theory Related Fields*, 167(1-2):107-142, 2017.
- [3] T. Schreiber and J. E. Yukich, Limit theorems for geometric functionals of Gibbs point processes *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques*, 49(4): 1158–1182, 2013
- [4] Fernández, R., Ferrari, P. A., Garcia, N. L., Perfect simulation for interacting point processes, loss networks and Ising models, *Stoch. Process. Appl*, 102(1): 63-88, 2002.
- [5] Nancy L. Garcia, Nevena Marić, *Improved bounds for perfect simulation of a continuous one-dimensional loss network*, 2018 (arXiv)
- [6] Christophe A.N. Biscio, Nicolas Chenavier, Christian Hirsch, and Anne Marie Svane, *Testing Goodness of Fit for Point Processes via Topological Data Analysis*, 2019 (arXiv)
- [7] Dereudre, David, Introduction to the theory of Gibbs point processes. *Stochastic Geometry. Lecture Notes in Mathematics*, vol 2237. Springer, Cham.
- [8] Samir Chowdhury and Facundo Mémoli, Persistent Path Homology of Directed Networks. *arXiv*, 2017.
- [9] Samir Chowdhury and Facundo Mémoli, Persistent Path Homology of Directed Networks. *Proceedings of the 2018 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2018.
- [10] Chazal, F., Fasy, B. T., Lecci, F., Rinaldo, A. e Wasserman, L., Stochastic convergence of persistence landscapes and silhouettes, *arXiv*, 2018.

- [11] Eric Berry, Yen-Chi Chen, Jessi Cisewski-Kehe, Brittany Terese Fasy, Functional Summaries of Persistence Diagrams, *Proceedings of the Thirtieth Annual Symposium on Computational Geometry*, ACM, 2014.
- [12] Nieves Atienza, Rocio Gonzalez-Diaz, Matteo Rucco Persistent Entropy for Separating Topological Features from Noise in Vietoris-Rips Complexes, *Journal of Intelligent Information Systems*, 52, 637–655, 2019.
- [13] Sebastian Raschka Naive Bayes and Text Classification I. Introduction and Theory *arXiv*, 2014.
- [14] Elie Azeraf, Emmanuel Monfrini and Wojciech Pieczynski Using the Naive Bayes as a Discriminative Classifier *arXiv*, 2021.
- [15] Thomas M. J. Fruchterman and Edward M. Reingold, Graph drawing by force-directed placement. *Software—Practice & Experience*, 21(11):1129-1164, 1991.
- [16] Günter Last and Mathew Penrose, *Lectures on the Poisson Process*. Cambridge University Press, 2017.
- [17] P. Hall and C. C. Heyde, *Martingale Limit Theory and Its Application*. Academic Press, 2014.
- [18] Elon Lages Lima, *Homologia Básica*. IMPA, Rio de Janeiro, 2^a edição, 2012.
- [19] Elon Lages Lima, *Grupo Fundamental e Espaços de Recobrimento*, IMPA, Rio de Janeiro, 5^a edição, 2018.
- [20] Garret Birkhoff e Saunders MacLane, *Algebra*, Chelsea Publishing Company, New York, 3^a edição, 1988.
- [21] Allen Hatcher, *Algebraic Topology*, Cambridge University Press, 1^a edição, 2001
- [22] John L. Kelley, *General Topology*, Dover Publication, Inc., New York, Reimpressão da 1^a edição de 1955, 2017,
- [23] J.P. May, *A Concise Course in Algebraic Topology*, University of Chicago Press, 1^a edição, 1999
- [24] César Polcino Milies, *Anéis e Módulos*, Editora Livraria da Física, São Paulo, 2^a edição, 2018

- [25] Thomas D. Ahle, Sharp and Simple Bounds for the raw Moments of the Binomial and Poisson Distributions. *Statistics & Probability Letters*, 182, 2022.
- [26] Steve Y. Oudot, *Persistence Theory: From Quiver Representations to Data Analysis*, Editora American Mathematical Society, 1^a edição, 2016
- [27] Jean-Daniel Boissonnat, Frédéric Chazal, Mariette Yvinec, *Geometric and Topological Inference* Editora Cambridge University Press, 1^a edição, 2018
- [28] Frédéric Chazal, Vin de Silva, Marc Glisse, Steve Oudout, *The Structure and Stability of Persistence Modules*, Editora Springer, 1^a edição, 2016
- [29] Malors Emilio Espinosa Lara, *Homología Persistente*, CIMAT, Notas de Aula
- [30] Shai Shalev-Shwartz, Shai Ben-David, *Understanding Machine Learning: From Theory to Algorithms*, Editora Cambridge University, 1^a edição, 2014
- [31] Yifan Hu, Efficient and High Quality Force-Directed Graph Drawing, *Mathematica Journal*, 10.37-71
- [32] Ronald Meester and Rahul Roy, *Continuum Percolation*, Cambridge University Press, 1^a edição, 1996
- [33] The GUDHI Project, *GUDHI User and Reference Manual*, GUDHI Editorial Board, edição 3.3.0, 2020, website: <https://gudhi.inria.fr/doc/3.3.0/>