



UFRPE

**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA**

Laura Alves Pacifico da Silva

Dissertação

ANÁLISE TOPOLÓGICA DE DADOS PARA SÉRIES TEMPORAIS

Recife - PE

17 de janeiro de 2020



UFRPE

**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA**

Laura Alves Pacifico da Silva

Dissertação

ANÁLISE TOPOLÓGICA DE DADOS PARA SÉRIES TEMPORAIS

Trabalho apresentado ao Programa de Pós-Graduação em Biometria e Estatística Aplicada do Departamento de Estatística e Informática da Universidade Federal Rural de Pernambuco como requisito parcial para obtenção do grau de Mestre em Biometria e Estatística Aplicada.

Orientador: Prof. Dr. Wilson Rosa de Oliveira Júnior

Recife - PE

17 de janeiro de 2020

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

586a

Silva, Laura Alves Pacifico da Silva

Análise Topológica de Dados para Séries Temporais / Laura Alves Pacifico da Silva Silva. - 2020.
97 f. : il.

Orientador: Wilson Rosa de Oliveira Junior.
Inclui referências e apêndice(s).

Dissertação (Mestrado) - Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em Biometria e Estatística Aplicada, Recife, 2020.

1. Análise Topológica de Dados . 2. Homologia Persistente. 3. Análise Topológica de Dados para Séries Temporais. I. Junior, Wilson Rosa de Oliveira, orient. II. Título

CDD 519.5

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

Análise Topológica de Dados para Séries Temporais

Laura Alves Pacifico da Silva

Dissertação julgada adequada para obtenção do título de Mestre em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 17/01/2020 pela Comissão Examinadora.

Orientador:

Prof. Dr. Wilson Rosa de Oliveira Júnior
Universidade Federal Rural de Pernambuco

Banca Examinadora:

Profa. Dra. Tatijana Stosic
Universidade Federal Rural de Pernambuco

Prof. Dr. Leon Denis da Silva
Universidade Federal Rural de Pernambuco

A minha avó, Cleonice Pacifico (in memoriam).

Agradecimentos

Em primeiro lugar, agradeço a Deus pelo dom da vida, por me guiar nos dias felizes e também nos dias sombrios me fazendo nunca perder a fé e esperança.

Agradeço à minha mãe, Maria Joana por fazer dos meus sonhos os seus sonhos, por sempre acreditar no meu potencial e me incentivar a trilhar os caminhos que sempre sonhei, independente das dificuldades que possam aparecer. Estou aqui graças a você e por você.

A toda minha família, obrigada pela rede de carinho e incentivo.

Ao meu namorado, Renato Teixeira, obrigada por me apoiar até aqui nesses quase 9 anos de relacionamento. Nos dias mais felizes e mais tristes você esteve comigo me incentivando, amparando e protegendo, tudo o que sou hoje trás um pouco do que aprendi com você, obrigada por tudo!

Aos meus amigos da vida: Larissa, Carol, Gilson (e toda sua família), Yara, Mário, Lorena, Edgar, Tanaka, Alan, Barbara, Marcelo, Tarci, Clessius e Eudes pelas longas risadas e amizade verdadeira. Cada um de vocês é muito importante para mim.

Aos amigos que a Biometria me trouxe: Joelma, Jackson, Denise, Micka, Leika, Bárbara, Augusto, Luciano, Edgo, Fábio, Kerolly, Jucarlos, Patricia, Marcela e tantos outros que compartilharam das dificuldades e das felicidades que a vida acadêmica nos proporciona. Vocês foram essenciais.

Aos Professores do departamento, que além do papel de docente contribuem para a formação pessoal de cada um que por aqui passa.

Ao departamento de matemática da UFRPE, que me recebeu de braços abertos durante

o período em que fui professora substituta. Todo o trabalho produzido nesse tempo contribuiu muito para o meu crescimento como profissional. Obrigada por me fazerem sentir em casa.

Ao meu orientador, Prof. Wilson por ser um orientador excepcional, pela paciência e compreensão com todas as dificuldades. Deixo aqui meu agradecimento pelo total envolvimento com este trabalho.

A Leon e Tatijana por aceitarem dedicar o valioso tempo de vocês lendo e colaborando com este trabalho.

Por fim agradeço à CAPES e CNPq pelo apoio financeiro.

Tive a sorte de conhecer e conviver com todos vocês. Muito obrigada a todos que direta ou indiretamente contribuíram para a realização deste trabalho.

*“É preciso força pra sonhar e perceber
que a estrada vai além do que se vê...”*

Marcelo Camelo.

Resumo

Análise topológica de dados (TDA), que se refere a utilização de ferramentas topológicas em dados, ganhou impulso considerável nos últimos anos. O TDA combina elementos de topologia algébrica e outras ferramentas da matemática pura para permitir um estudo útil da forma dos dados através de homologia persistente. Mais recentemente TDA está sendo usado para estudar séries temporais. Neste trabalho, aplicamos técnicas da Análise Topológica de Dados, mais precisamente, usamos homologia persistente para calcular características topológicas mais persistentes em dados de séries de commodities agrícolas: açúcar, etanol e petróleo, no período de 25/01/2010 à 31/12/2018. Utilizamos diagramas de persistência para representar as saídas obtidas e utilizamos a distância de bottleneck para avaliar o grau de similaridade entre os pares de commodities estudados. Os resultados obtidos mostraram que maiores distâncias no diagrama de persistência foram obtidas nas séries de retorno e volatilidade para pares de petróleo/açúcar e petróleo/etanol, onde correlações menores evidenciam baixa similaridade entre as séries. Tais resultados são validados com o estudo de (LIMA *et al.*, 2019; SILVA *et al.*, 2019) e se justificam para o primeiro par, pela relação entre custo do transporte e a influencia que isso provoca no preço final, para o segundo par, pela concorrência no mercado, pois se tratam de combustíveis que compõem o mercado energético.

Palavras-chave: Análise Topológica de Dados, Homologia Persistente, Análise Topológica de Dados para Séries Temporais

Abstract

Topological Data Analysis (TDA), which refers to the use of topological data tools, has gained considerable momentum in recent years. TDA combines elements of algebraic topology and other tools of pure mathematics to enable useful study of data form through persistent homology. More recently TDA is being used to study time series. In this work, we apply Topological Data Analysis techniques, more precisely, we use persistent homology to calculate more persistent topological characteristics in data from series of agricultural commodities: sugar, ethanol and crude oil, from January 25, 2010 to December 31, 2018. We use persistence diagrams to represent the outputs obtained and we use the bottleneck distance to evaluate the degree of similarity between the studied commodity pairs. The results obtained showed that greater distances in the persistence diagram were obtained in the return and volatility series for crude oil / sugar and crude oil / ethanol pairs, where smaller correlations show low similarity between the series. These results are validated with the study of (LIMA *et al.*, 2019; SILVA *et al.*, 2019) and are justified for the first pair, by the relationship between transport cost and the influence that this has on the final price, for the second pair, by market competition, since These are fuels that make up the energy market.

Keywords:: Topological Data Analysis, Persistent Homology, Topological Data Analysis for time series.

Lista de Figuras

Figura 1	O conjunto $A \subseteq X$ e seus conjuntos relacionados.	24
Figura 2	Exemplos de simplices	27
Figura 3	(a)Exemplo de um complexo simplicial (b) exemplo de um complexo não simplicial	28
Figura 4	Correspondência entre os complexos simpliciais e os complexos simpliciais abstratos.	29
Figura 5	Exemplo da ação do operador fronteira.	31
Figura 6	A ação do operador fronteira.	32
Figura 7	A sequência de complexos de cadeia, ciclos e grupos de fronteira conectados por homomorfismos.	34
Figura 8	Exemplo de um diagrama de Persistência.	42
Figura 9	Série temporal de preços P para açúcar, etanol e petróleo no período de 25/01/2010 a 31/12/2018.	50
Figura 10	Série temporal de retornos R para açúcar, etanol e petróleo no período de 25/01/2010 a 31/12/2018.	50
Figura 11	Série temporal de volatilidades V para açúcar, etanol e petróleo no período de 25/01/2010 a 31/12/2018.	51
Figura 12	Informação Mútua e Método dos Falsos Vizinhos Próximos para série de Retorno do Açúcar	53
Figura 13	Informação Mútua e Método dos Falsos Vizinhos Próximos para série de Volatilidade do Açúcar	53
Figura 14	Informação Mútua e Método dos Falsos Vizinhos Próximos para série de Retorno do Etanol	54

Figura 15	Informação Mútua e Método dos Falsos Vizinhos Próximos para série de Volatilidade do Etanol	54
Figura 16	Informação Mútua e Método dos Falsos Vizinhos Próximos para série de Retorno do Petróleo	54
Figura 17	Informação Mútua e Método dos Falsos Vizinhos Próximos para série de Volatilidade do Petróleo	55
Figura 18	Reconstrução por coordenadas atrasadas para as séries de retorno das commodities.	55
Figura 19	Reconstrução por coordenadas atrasadas para as séries de volatilidade das commodities.	55
Figura 20	Diagramas de persistência para as séries de retorno das commodities. . .	56
Figura 21	Diagramas de persistência para as séries de volatilidade das commodities.	56
Figura 22	Distância de Bottleneck entre os pares de commodities para as séries de retorno (a)Açúcar/Etanol, (b)Petróleo/Açúcar, (c)Petróleo/Etanol.	57
Figura 23	Distância de Bottleneck entre os pares de commodities para as séries de volatilidade (a)Açúcar/Etanol, (b)Petróleo/Açúcar, (c)Petróleo/Etanol. .	58
Figura 24	Pipeline para TDA + Redes Neurais	61

Lista de Tabelas

Tabela 1	Números de Betti para um único ponto, uma circunferência, esfera, um toro e para a garrafa de Klein. Em um espaço k -dimensional, o n -ésimo número de Betti é sempre zero para qualquer $n \geq k$	36
Tabela 2	Correlação entre as Séries de Retorno versus Distância de Bottleneck. . .	57
Tabela 3	Correlação entre as Séries de Volatilidade versus Distância de Bottleneck. . .	57

Lista de Símbolos

\mathbb{N}	Conjunto dos números naturais
\mathbb{Z}	Conjunto dos números inteiros
\mathbb{Q}	Conjunto dos números racionais
\mathbb{R}	Conjunto dos números reais
\mathbb{R}^n	Espaço euclidiano n-dimensional
\mathbb{X}	Espaço topológico
\in	Pertence
\notin	Não pertence
\exists	Não pertence
Σ	Somatório
\cup	União de conjuntos
\cap	Interseção de conjuntos
\subseteq	Subconjunto
\emptyset	Conjunto vazio
\sim	Relação de equivalência
\bar{x}	Classe de equivalência
\equiv	Equivalente
\circ	Composição
$(G, *)$	Grupo
$\overset{\circ}{A}$	Interior de A
\bar{A}	Fecho de A
∂A	Fronteira de A
τ	Parâmetro de atraso
d	Dimensão da nuvem de pontos
β_k	k -ésimo número de Betti

Sumário

1	Introdução	1
1.1	Introdução	1
1.2	Objetivos	4
1.2.1	Objetivos Gerais	4
1.2.2	Objetivos Específicos	4
1.3	Estrutura do Trabalho	4
2	Conceitos Preliminares	7
2.1	Relações de equivalência	7
2.2	Teoria dos Grupos	10
2.3	Topologia	23
2.4	Homologia Simplicial	26
3	Análise Topológica de Dados	38
3.1	Homologia Persistente	38
3.1.1	Filtrações e Persistência	39
3.1.2	Diagramas de Persistência	41
4	Análise Topológica de Dados para Séries Temporais Agroambientais	46
4.1	TDA para Séries Temporais	46
4.2	TDA em Commodities Agrícolas	49
4.3	Dados	50
5	Resultados e Discussão	53

6	Conclusões e Perspectivas	60
	Referências Bibliográficas	64
A	Códigos Fonte	65

CAPÍTULO 1

Introdução

1.1 Introdução

Com o desenvolvimento da ciência e tecnologia, dados de mais diversos tipos estão sendo produzidos em larga escala. Com a grande produção desses dados, se fez necessário a adaptação de técnicas já existentes e criação de outras para analisá-los e avaliá-los pois as técnicas tradicionais de análise de dados nem sempre acompanham a quantidade e a complexidade explosivas destes. Para análise desse tipo de dado utilizar a metodologia de criação de modelo, simulação e avaliação não é viável. Essa execução é útil para resolver problemas clássicos, onde a teoria vem sendo amplamente estudada, de modo que pode ser adaptada ao modelo que será criado. Portanto, para processar tais dados precisaremos primeiro definir uma hipótese concreta das características dos dados que podem ser testadas (SNÁŠEL *et al.*, 2017; CARLSSON, 2009).

Neste sentido, utilizaremos a topologia como ferramenta matemática que possibilita a visualização e organização dos dados, de modo que estruturas que técnicas convencionais não capturam sejam percebidas. Com a Análise Topológica de Dados (Topological Data Analysis - TDA) é possível criar uma “representação comprimida” de todas as características dos dados para rapidamente ajudar a desvendar padrões e relacionamentos do conjunto dados. O formalismo matemático, que foi desenvolvido para a incorporação de técnicas geométricas e topológicas (CARLSSON, 2009), lida com nuvens de pontos, ou seja, conjuntos finitos de pontos imersos em um espaço multidimensional, geralmente Euclidiano. Então, são adaptadas ferramentas dos vários ramos da geometria e da topologia para o estudo destes conjuntos que são amostras finitas,

tomadas a partir de um objeto geométrico, possivelmente com ruído. A topologia fornece uma linguagem formal para a matemática qualitativa, onde as relações de proximidade (ou vizinhança) são estudadas, sem o uso de distâncias. A ideia de construção de resumos das características dos dados envolve a compreensão da relação entre objetos topológicos e geométricos, ([SNÁŠEL *et al.*, 2017](#)).

Devido ao rápido crescimento na gama de aplicações da topologia surge a necessidade de algoritmos eficientes para computar características dos dados tais como: grupos de homologia, homologia persistente e etc. Há uma variedade de estratégias que foram adotadas. Uma destas sugere ver a diagonalização de Smith ([CARLSSON, 2009](#)) como um problema puramente algébrico e, em seguida, buscar um algoritmo ideal. A análise de pior caso de tais algoritmos sugere uma complexidade supercúbica em relação ao tamanho do complexo, que é proibitivo para grandes conjuntos de dados. Uma estratégia alternativa é desenvolver algoritmos eficientes para problemas restritos, por exemplo, restringindo a dimensão do complexo ou restringindo a computação à dos números de Betti ([CARLSSON *et al.*, 2012](#)).

Embora a abordagem de usar ferramentas de topologia para resolver problemas dados por séries temporais tenha sido utilizada no passado ([MULDOON *et al.*, 1993](#)), a Análise Topológica de Dados (TDA) tornou-se famosa como um tópico de pesquisa através dos trabalhos de ([EDELBRUNNER; HARER, 2010](#)) e ([CARLSSON, 2009](#)). Atualmente, tem sido utilizada com sucesso em redes cósmicas ([WEYGAERT *et al.*, 2011](#)), análise de formas ([CARRIÈRE; OUDOT; OVSIANIKOV, 2015](#); [BONIS *et al.*, 2016](#)), análise de dados biológicos ([KOVACEV-NIKOLIC *et al.*, 2016](#); [WANG; OMBAO; CHUNG, 2018](#)), redes de sensores ([ADAMS; CARLSSON, 2015](#)), bem como outros campos.

A utilização de TDA para séries temporais é uma área relativamente nova e de rápido desenvolvimento, com aplicações a diversas áreas. [BERWALD; GIDEA; VEJDEMO-JOHANSSON](#) utilizaram TDA em análises climáticas. [KHASAWNEH; MUNCH](#) usaram noções de persistência em nuvens de pontos em várias janelas de séries temporais para rastrear a estabilidade de sistemas dinâmicos. [PEREA; HARER](#) utilizaram a noção de persistência máxima dos grupos de homologia para quantificar a periodicidade das séries temporais. [PEREIRA; MELLO](#) usaram técnicas advindas da homologia persistente para reunir populações de *Tribolium* (besou-

ros). [UMEDA](#) utilizou características topológicas de grupos de homologia uni e bidimensional como entrada em redes neurais convolucionais para classificação de séries temporais em três domínios diferentes. [GIDEA; KATZ](#) exploraram o uso do TDA em séries temporais financeiras.

Sabe-se que séries temporais não possuem originalmente representações em forma de nuvens de pontos. A transformação de uma série temporal para uma nuvem de pontos pode ser feita através da imersão de Takens, uma janela deslizante ([TAKENS, 1981](#)), garantindo a preservação das propriedades topológicas da série temporal. Esta técnica consiste em transformar uma série temporal em seu espaço de fase, ou seja, em uma nuvem de pontos ou um conjunto de pontos, com dois parâmetros, τ que é chamado parâmetro de atraso e d que especifica a dimensão da nuvem de pontos. Aqui o grande problema é a determinação dos valores desses parâmetros.

O uso de imersão de Takens é justificado matematicamente por um teorema ([TAKENS, 1981](#)), no qual é demonstrado que, para um d suficientemente grande, quase todas as opções de parâmetro de atraso τ permitem a detecção de quaisquer ciclos existentes, desde que haja um número infinito de dados sem ruídos. No entanto, os sinais do mundo real são dados finitos e com ruído. Assim, a falta de tais propriedades fazem com que com que ciclos significativos não sejam percebidos e/ou sejam observados falsos ciclos, dependendo da resolução assumida da amostragem. Como resultado, apenas algumas poucas opções de τ podem realmente fornecer resultados válidos sob amostragem ruidosa. Uma dificuldade primária no uso de imersão de Takens é prever quais são essas opções válidas de d quando os parâmetros do sinal ou sistema subjacente são desconhecidos.

Takens nada fala sobre como deve ser feita a escolha correta do delay. Aqui, a ideia é encontrar um valor τ em que $x(t + \tau)$ tenha alguma correlação com $x(t)$, mas esta não seja muito forte. Baseados nesta ideia, métodos como método da função de autocorrelação e método da informação mútua média são utilizados na determinação de tal parâmetro ([MALETIĆ; ZHAO; RAJKOVIĆ, 2016](#)).

Nesse sentido a proposta deste trabalho é fazer um levantamento teórico completo sobre a análise topológica de dados e aplica-la a séries agroambientais, mais especificamente séries de commodities agrícolas.

1.2 Objetivos

1.2.1 Objetivos Gerais

Utilizar Análise Topológica de Dados em Séries Temporais.

1.2.2 Objetivos Específicos

- Fazer um estudo teórico sobre a Geometria e Topologia para dados;
- Utilizar as ferramentas desenvolvidas no estudo teórico para:

Analisar a relação entre o preço do açúcar e etanol brasileiro e preços internacionais do petróleo, aplicando TDA e comparar a análise dos resultados com o grau de correlação das séries de retorno e volatilidade afim de mostrar como o mercado externo difere topologicamente do mercado interno.

1.3 Estrutura do Trabalho

Este trabalho esta dividido em 6 capítulos.

Neste capítulo introdutório, apresentamos uma visão geral do trabalho, mostrando uma breve apanhado histórico da teoria que será desenvolvida bem como os objetivos que pretendemos alcançar.

No capítulo 2, é feito um levantamento de conceitos matemáticos importantes para o entendimento de análise topológica de dados. Partimos de relação equivalência e em seguida definimos uma estrutura algébrica munida com uma operação, um grupo. Com o desenvolvimento de tal estrutura vamos agora olhar para a forma destas e com isso trataremos de sua topologia. Por fim, usaremos métodos de topologia algébrica para contar a ocorrência de padrões geométricos num dado espaço topológico.

No Capítulo 3, estudaremos a adaptação dos objetos definidos no capítulo 2, os grupos

de homologia, para nosso propósito que é verificar quais características topológicas persistem nos diferentes níveis de filtração.

No capítulo 4, desenvolvemos uma forma de analisar séries temporais via TDA. Para fins de aplicação da teoria desenvolvida analisaremos dados de commodities agrícolas. Este capítulo é encerrado com a proposta de comparação de commodities de açúcar, etanol e petróleo e relação dos pares no mercado financeiro.

No capítulo 5, discutimos os resultados obtidos nas comparações dos pares de commodities via TDA com os resultados obtidos anteriormente.

Por fim, são feitas considerações finais acerca do trabalho e propostas de trabalhos futuros.

CAPÍTULO 2

Conceitos Preliminares

Neste capítulo, apresentaremos algumas definições e exemplos sobre relações de equivalência, teoria dos grupos e topologia algébrica que serão usados neste trabalho. Discutiremos sobre grupos, complexos simpliciais sua realização geométrica e, posteriormente, introduziremos grupos de homologia. Esses conceitos formarão a base para nossa abordagem topológica dos dados.

2.1 Relações de equivalência

Nesta seção definiremos o que é uma relação de equivalência e apresentamos alguns exemplos. Através de tais relações é possível classificar e “agrupar” elementos de um certo conjunto em classes onde dois elementos estão na mesma classe se são equivalentes pela relação.

Definição 2.1.1. (*Relação de equivalência*) Seja S um conjunto não vazio e \sim uma relação em S . Dizemos que a relação \sim é de equivalência se, para quaisquer $x, y, z \in S$, são satisfeitas as seguintes propriedades:

- i. (*Reflexiva*) $x \sim x$;
- ii. (*Simétrica*) Se $x \sim y$, então $y \sim x$;
- iii. (*Transitiva*) Se $x \sim y$ e $y \sim z$, então $x \sim z$.

Exemplo 2.1.1. A relação de igualdade em \mathbb{R} é uma relação de equivalência. De fato, dados $x, y, z \in \mathbb{R}$, temos que

- i. $x = x$;
- ii. Se $x = y$, então $y = x$;
- iii. Se $x = y$ e $y = z$ então, $x = z$.

Definição 2.1.2. (*Classe de equivalência*) Seja S um conjunto não vazio, $x \in S$ e \sim uma relação de equivalência em S . Chamamos de classe de equivalência do elemento x em relação a \sim , o conjunto \bar{x} formado por todos os elementos de S que estão relacionados com x . Ou ainda

$$\bar{x} = \{y \in S \mid y \sim x\}$$

dizemos então que y é equivalente a x .

Proposição 2.1.1. *Sejam S um conjunto não vazio e \sim uma relação de equivalência em S . Para quaisquer $x, y \in S$ valem:*

- i. $\bar{x} = \bar{y}$ se e somente se $x \sim y$;
- ii. Se $\bar{x} \neq \bar{y}$, então $\bar{x} \cap \bar{y} = \emptyset$;
- iii. $\bigcup_{x \in S} \bar{x} = S$.

Demonstração. i. Vamos mostrar que se $\bar{x} = \bar{y}$ então $x \sim y$. De fato, se $\bar{x} = \bar{y}$, dado $x \in \bar{x}$ tem-se que $x \in \bar{y}$. Portanto, $x \sim y$.

Por outro lado, se $z \in \bar{x}$ tem-se que $z \sim x$. Dado que $x \sim y$, usando a transitividade temos que $z \sim y$ logo, $z \in \bar{y}$. E portanto, $\bar{x} \subset \bar{y}$. Agora, dado $w \in \bar{y}$, temos que $w \sim y$. Como por hipótese, $x \sim y$, então $y \sim x$. Pela transitividade $w \sim x$, ou seja, $w \in \bar{x}$ daí, $\bar{y} \subset \bar{x}$. Assim, $\bar{x} = \bar{y}$.

- ii. Se $\bar{x} \cap \bar{y}$ é não vazio, então existe $z \in \bar{x} \cap \bar{y}$. Daí, $z \in \bar{x}$ e $z \in \bar{y}$, logo, $z \sim x$ e $z \sim y$. Usando a simetria temos que $x \sim z$ e pela transitividade, $x \sim y$. Desta forma, usando o item i. podemos concluir que $\bar{x} = \bar{y}$.

iii. Note que $\bar{x} \subset S$, para qualquer $x \in S$. Logo, $\bigcup_{x \in S} \bar{x} \subset S$. Se $x \in S$ como \sim é uma relação de equivalência em S , então pela reflexividade temos que $x \sim x$, ou seja, $x \in \bar{x}$. Como, $x \in \bigcup_{x \in S} \bar{x}$, temos que $x \in \bigcup_{x \in S} \bar{x}$ daí, $S \subset \bigcup_{x \in S} \bar{x}$. Portanto, $\bigcup_{x \in S} \bar{x} = S$.

□

Definição 2.1.3. (*Conjunto Quociente*) Seja A um conjunto não vazio e \sim uma relação de equivalência em A . O conjunto formado por todas as classes de equivalência, denotado por

$$A / \sim = \{\bar{x} \mid x \in A\},$$

é denominado conjunto quociente de A pela relação de equivalência \sim .

Proposição 2.1.2. Seja \mathbb{Z} o conjunto dos números inteiros e $n \in \mathbb{Z}$. Considere a relação $\equiv \text{mod } (n)$, definida da seguinte forma: Dados x, y, z dizemos que $x \equiv y \text{ mod } (n)$ se existe $k \in \mathbb{Z}$ tal que $x - y = nk$. Então, $\equiv \text{mod } (n)$ é uma relação de equivalência.

Demonstração. Para todos $x, y, z \in \mathbb{Z}$ temos que

i. $x - x = 0 = n \cdot 0$, para qualquer $n \in \mathbb{Z}$, logo,

$$x \equiv x \text{ mod } (0),$$

portanto,

$$x \sim x \Rightarrow x \equiv x \text{ mod } (n).$$

Sendo assim, $\equiv \text{mod } (n)$ é reflexiva.

ii. Se $x \equiv y \text{ mod } (n)$, então existe $k' \in \mathbb{Z}$, tal que

$$x - y = nk'.$$

Portanto,

$$y - x = n(-k')$$

Daí,

$$y \equiv x \pmod{n}$$

e assim a relação é simétrica.

iii. Se $x \equiv y \pmod{n}$ e $y \equiv z \pmod{n}$, então existem w e $w' \in \mathbb{Z}$ tais que

$$x - y = nw \quad \text{e} \quad y - z = nw'.$$

Assim

$$(x - y) + (y - z) = nw + nw',$$

logo

$$x + (-y + y) - z = n(w + w'),$$

portanto

$$x - z = n(w + w').$$

O que nos dá a transitividade da relação. Logo, $x \equiv z \pmod{n}$ é relação de equivalência.

□

Observação 2.1.1. A relação de equivalência mostrada acima é chamada de congruência módulo n . O conjunto quociente

$$\mathbb{Z} / \equiv \pmod{n} = \{\bar{x} \mid x \in \mathbb{Z}\},$$

é denotado por \mathbb{Z}_n .

2.2 Teoria dos Grupos

Nesta seção estudaremos algumas estruturas algébricas munidas com uma operação, estas estruturas são chamadas grupos. Uma outra forma de definir um grupo é dizer que estudaremos simetrias, ou seja, o grupo consiste no conjunto de transformações que mantêm o

objeto inalterado e na operação de combinar duas dessas transformações executando uma após a outra. Estes serão de fundamental importância para o trabalho, pois são usados para descrever os invariantes de espaços topológicos.

Definição 2.2.1. (*Grupo*) Seja G um conjunto não vazio munido de uma operação

$$\begin{aligned} * : G \times G &\longrightarrow G \\ (a, b) &\longmapsto a * b. \end{aligned}$$

Dizemos que $(G, *)$ é um grupo se para quaisquer $a, b, c \in G$ são válidas:

- i. $(a * b) * c = a * (b * c)$;
- ii. Existe $e \in G$ tal que $a * e = e * a = a$;
- iii. Para todo $a \in G$, com $a \neq e$ existe $a^{-1} \in G$ tal que $a * a^{-1} = a^{-1} * a = e$.

Observação 2.2.1. Afim de tornar a leitura menos carregada, deixaremos de explicitar a operação, escrevendo somente G para designar o grupo $(G, *)$, assim como quando não houver imprecisão utilizaremos ab no lugar de $a * b$.

Exemplo 2.2.1. O conjunto dos números inteiros \mathbb{Z} munido com a operação de soma usual $(\mathbb{Z}, +)$ é um grupo. De fato, dados $a, b, c \in \mathbb{Z}$

- i. $(a + b) + c = a + (b + c)$;
- ii. Existe $0 \in \mathbb{Z}$, tal que $a + 0 = 0 + a = a$;
- iii. Para todo $a \in \mathbb{Z}$, sendo $a \neq 0$, existe um único elemento $(-a) \in \mathbb{Z}$ tal que $a + (-a) = (-a) + a = 0$.

Portanto, $(\mathbb{Z}, +)$ é um grupo.

Exemplo 2.2.2. (\mathbb{Z}, \cdot) não é um grupo. De fato, dado, $m \in \mathbb{Z} - \{-1, 1\}$, não existe $n \in \mathbb{Z}$ tal que $m \cdot n = 1$.

Exemplo 2.2.3. O conjunto dos números racionais sem o zero \mathbb{Q}^* munido da operação de multiplicação usual é um grupo. Com efeito, sejam $x, y, z \in \mathbb{Q}^*$, tem-se

$$i. (x \cdot y) \cdot z = x \cdot (y \cdot z);$$

$$ii. \text{ Existe } 1 \in \mathbb{Q}^*, \text{ tal que } x \cdot 1 = 1 \cdot x = x;$$

$$iii. \text{ Para todo } x \in \mathbb{Q}^*, \text{ com } x \neq 1, \text{ existe um \uacute;nico } x^{-1}, \text{ tal que, } x \cdot x^{-1} = x^{-1} \cdot x = 1.$$

Portanto, (\mathbb{Q}^*, \cdot) é um grupo.

Definição 2.2.2. (Grupo finito) Um grupo $(G, *)$ é finito quando o conjunto G for finito.

Exemplo 2.2.4. (Grupo das permutações) Seja S um conjunto não vazio e

$$G = \{f : S \longrightarrow S; f \text{ é bijetiva} \}$$

munido da operação.

$$\circ : G \times G \longrightarrow G$$

$$(f, g) \longmapsto f \circ g$$

onde $(f \circ g)(x) = f(g(x))$, para $x \in S$. Tem-se que (G, \circ) é um grupo.

Demonstração. Note que o conjunto G é não vazio, pois a função identidade, dada por

$$i : S \longrightarrow S$$

$$x \longmapsto i(x) = x$$

é uma função bijetiva. Logo, $i \in G$. Agora, tome $f, g, h \in G$.

i. Para todo $x \in S$

$$\begin{aligned} [(f \circ g) \circ h](x) &= f \circ g(h(x)) \\ &= f(g(h(x))) \\ &= f[(g \circ h)(x)] \\ &= [f \circ (g \circ h)](x), \end{aligned}$$

assim, $(f \circ g) \circ h = f \circ (g \circ h)$ então esta operação é associativa.

ii. Note que,

$$\begin{aligned} f \circ i(x) &= f(i(x)) \\ &= f(x). \end{aligned}$$

Analogamente, $i \circ f(x) = f(x)$. Assim, a função identidade é o elemento neutro de G .

iii. Como f é bijetiva, então f admite o elemento inverso que neste caso é a função inversa, ou seja, existe $f^{-1} \in G$ tal que

$$f \circ f^{-1} = i \quad \text{e} \quad f^{-1} \circ f = i.$$

□

Proposição 2.2.1. (*Propriedade do Cancelamento*) *Sejam G um grupo e $x, y, z \in G$, com $z \neq e$, tais que $x * z = y * z$, então $x = y$.*

Demonstração. Seja $x * z = y * z$. Observe que

$$x = x * e, \tag{2.1}$$

como $z \neq e$, então existe $z^{-1} \in G$, tal que $z * z^{-1} = e$. Assim, usando 2.1 temos que

$$\begin{aligned}
 x &= x * e \\
 &= x * (z * z^{-1}) \\
 &= (x * z) * z^{-1} \\
 &= (y * z) * z^{-1} \\
 &= y * (z * z^{-1}) \\
 &= y * e \\
 &= y.
 \end{aligned}$$

Assim, $x = y$. □

Proposição 2.2.2. *Seja $(G, *)$ um grupo e $x \in (G, *)$. Então, o elemento neutro e o elemento inverso de G são únicos.*

Demonstração. Primeiro mostraremos que o elemento neutro é único. De fato, sejam e e e' os elementos neutros do grupo $(G, *)$. Assim,

$$e' = e' * e = e * e' = e.$$

Daí, $e' = e$.

Agora, mostraremos que o elemento inverso é único. De fato, seja $x \in G$ e suponhamos que existem $x', x^{-1} \in G$, tais que x' e x^{-1} sejam inversos de x . Daí,

$$x' = x' * e = x' * (x * x^{-1}) = (x' * x) * x^{-1} = e * x^{-1} = x^{-1}.$$

Portanto, temos que $x' = x^{-1}$, isto é, o inverso também é único. □

Proposição 2.2.3. *Sejam (G, \cdot) um grupo e $a, b \in G$. Então, são válidos:*

i. $(a^{-1})^{-1} = a$

ii. $(a \cdot b)^{-1} = b^{-1} \cdot a^{-1}$

Demonstração. **i.** Note que,

$$a \cdot a^{-1} = e = (a^{-1})^{-1} \cdot a^{-1}$$

Daí, pela propriedade do cancelamento, temos que $a = (a^{-1})^{-1}$.

ii. Utilizando as propriedades associativa, elemento neutro e elemento inverso, segue que

$$\begin{aligned} (a \cdot b) \cdot (b^{-1} \cdot a^{-1}) &= a \cdot (b \cdot b^{-1}) \cdot a^{-1} \\ &= a \cdot e \cdot a^{-1} \\ &= (a \cdot e) \cdot a^{-1} \\ &= a \cdot a^{-1} \\ &= e. \end{aligned}$$

Como o elemento inverso é único, temos que $(a \cdot b)^{-1} = b^{-1} \cdot a^{-1}$.

□

Alguns grupos satisfazem uma quarta propriedade chamada comutatividade com relação a operação $*$ em G , tais grupos são chamados grupos abelianos.

Definição 2.2.3. (*Grupo Abeliano*) Dizemos que $(G, *)$ é um grupo abeliano, quando a operação que define G for comutativa, isto é, para $a, b \in G$

$$a * b = b * a$$

Exemplo 2.2.5. O grupo $(\mathbb{Z}, +)$ é um grupo abeliano, pois, quaisquer que sejam $x, y \in \mathbb{Z}$ tem-se

$$x + y = y + x$$

Exemplo 2.2.6. O conjunto $G = GL(n, K)$, $n = 2$, de todas as matrizes 2×2 invertíveis com coeficientes em um corpo K é um grupo **não** abeliano, em relação a operação de produto de

matrizes. Com efeito, sejam $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ e $\begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}$ elementos de G . Note que

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 3 & 1 \end{pmatrix} \neq \begin{pmatrix} 2 & 0 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

Alguns subconjuntos de um grupo G quando satisfazem algumas condições, também são grupos com a operação de G , tais subconjuntos são denominados subgrupos de G .

Definição 2.2.4. (Subgrupo) Seja $(G, *)$ um grupo e J um subconjunto não vazio de G . Dizemos que J é um subgrupo de G se $(J, *)$ tem estrutura de grupo com a mesma operação de G neste caso escrevemos a notação $J \leq G$ para dizer que J é subgrupo de G .

Observação 2.2.2. Todo subgrupo G possui dois subgrupos G e $\{e\}$, estes são chamados subgrupos triviais de G .

Proposição 2.2.4. Seja G um grupo e J um subgrupo não vazio. J será subgrupo de G se, e somente se são válidas:

- i. $e \in J$;
- ii. Para todo $x, y \in J$ temos que $xy \in J$;
- iii. Para todo $x \in J$, temos que $x^{-1} \in J$.

Demonstração. Se J é um subgrupo de G então é grupo com a mesma operação de G . Sejam $x, y \in J$, valem

- i. $e \in J$;
- ii. $xy \in J$;
- iii. Existe $x^{-1} \in J$ tal que $xx^{-1} = e$.

Por outro lado, como $e \in J$, então J é não vazio. Para todos $x, y, z \in J$, como J é um subconjunto de G , então $x, y, z \in G$. Como G é grupo vale

$$(xy)z = x(yz),$$

e por iii. todo elemento de J possui inverso em J . Desta maneira J é grupo e pela definição anterior, J é subgrupo de G . \square

Seja G um grupo e seja J um subgrupo de G . Sobre G , defina a relação de equivalência \sim_E da seguinte maneira:

$$y \sim_E x \Leftrightarrow \exists j \in J \text{ tal que } y = xj.$$

Por definição, a classe de equivalência que contém x é o conjunto $\{y \in G \mid y \sim_E x\} = \{xj \mid j \in J\}$ denotaremos este conjunto por xJ e o chamaremos de **classe lateral à esquerda** de J em G que contém x . Quando não houver confusão, chamaremos simplesmente de classe lateral de x à esquerda. Em particular, J é a classe lateral do elemento neutro e à esquerda. Note que, $y \in xJ \Leftrightarrow yJ = xJ$.

Analogamente, definindo

$$y \sim_D x \Leftrightarrow \exists j \in J \text{ tal que } y = jx$$

obtemos as **classes laterais à direita** de J em G como sendo a classe lateral de x à direita $Jx = \{jx \mid j \in J\}$.

Agora, seja G um grupo e seja J um subgrupo de G . Gostaríamos de saber se a operação de G induz de maneira natural uma operação sobre o conjunto das classes laterais à esquerda de J em G , isto é, se a operação

$$(xJ, yJ) \mapsto xyJ$$

é bem definida, no sentido de não depender da escolha dos representantes de x e y . Dados $x, y \in G$ e $h, k \in J$ arbitrários, então x e xh são representantes da mesma classe xJ e y e yk são da mesma classe yJ . Assim, a operação induzida sobre as classes laterais à esquerda é bem definida

se e só se

$$xyJ = xhykJ, \quad \forall x, y \in G, \quad \forall h, k \in J$$

logo, se e só se

$$y^{-1}x^{-1}xyJ = y^{-1}x^{-1}xhykJ,$$

ou seja,

$$y_j y^{-1} \in J, \quad \forall y \in G \quad \forall j \in J$$

Proposição 2.2.5. *Seja J um subgrupo de G . As afirmações abaixo são equivalentes:*

- i.** *a operação induzida sobre as classes laterais à esquerda de J em G é bem definida;*
- ii.** $gJg^{-1} \subseteq J, \quad \forall g \in G;$
- iii.** $gJg^{-1} = J, \quad \forall g \in G;$
- iv.** $gJ = Jg, \quad \forall g \in G$

Definição 2.2.5. *Um subgrupo J é um subgrupo normal de G e escrevemos $H \triangleleft G$ se ele satisfaz as afirmações equivalentes da proposição anterior.*

Neste caso, as classes laterais à esquerda de J são iguais as classes laterais à direita de J , então chamaremos apenas de **classes laterais de J** .

Agora, veremos funções entre grupos que preservam algumas propriedades algébricas, estes serão chamados Homomorfismos.

Definição 2.2.6. *(Homomorfismos de Grupos) Sejam $(G, *)$ e (G', \cdot) grupos. Uma função $\psi : G \rightarrow G'$ que satisfaz*

$$\psi(x * y) = \psi(x) \cdot \psi(y),$$

*para quaisquer $x, y \in G$ é chamada **homomorfismo**.*

Exemplo 2.2.7. *Sejam os grupos (\mathbb{R}_+, \cdot) e $(\mathbb{R}, +)$. A função*

$$\begin{aligned} f : \mathbb{R}_+ &\longrightarrow \mathbb{R} \\ x &\longmapsto f(x) = \log(x) \end{aligned}$$

é um homomorfismo. Com efeito, sejam $a, b \in \mathbb{R}_+$ temos que

$$\begin{aligned} f(a \cdot b) &= \log(ab) \\ &= \log(a) + \log(b) \\ &= f(a) + f(b) \end{aligned}$$

Observação 2.2.3. Quando $\psi : G \rightarrow G'$ é um homomorfismo bijetivo, dizemos que ψ é um **isomorfismo** e dizemos que G e G' são isomorfos e denotamos por $G \simeq G'$. Quando o isomorfismo for $\psi : G \rightarrow G$ ele é chamado de **automorfismo** e é denotado por $\text{Aut}G$.

Definição 2.2.7. Sejam G e G' grupos, $\psi : G \rightarrow G'$ um homomorfismo e e' o elemento neutro de G' . O conjunto definido por

$$\text{Ker}(\psi) = \{g \in G; \psi(g) = e', e' \in G'\}$$

é chamado núcleo do homomorfismo ψ .

Proposição 2.2.6. Se $\psi : (G, *) \rightarrow (G', \cdot)$ é um homomorfismo então:

- i. $\psi(e) = e'$, onde e é o elemento neutro de G e e' é o elemento neutro de G' ;
- ii. $\psi(a^{-1}) = \psi(a)^{-1}$ para todo $a \in G$;
- iii. ψ é injetiva se e somente se, $\text{ker}(\psi) = \{e\}$

Demonstração. **i.** Tem-se

$$\psi(e) = \psi(e * e) = \psi(e) \cdot \psi(e)$$

multiplicando por $\psi(e)^{-1}$ ficamos com

$$e' = \psi(e)$$

ii. Pelo item i., tem-se

$$\begin{aligned}\psi(a)\psi(a^{-1}) &= \psi(aa^{-1}) \\ &= \psi(e) \\ &= e'.\end{aligned}$$

Assim, conclui-se que $\psi(a^{-1}) = \psi(a)^{-1}$.

iii. Seja ψ injetiva e $g \in \ker(\psi)$. Logo, $\psi(g) = e'$. Usando o item i. temos que $\psi(e) = e'$. Logo, $\psi(e) = \psi(g)$ e assim, $e = g \in \{e\}$ assim, conclui-se que $\ker(\psi) \subset \{e\}$. Como por definição $\{e\} \in \ker(\psi)$ temos então que $\ker(\psi) = \{e\}$. Por outro lado, suponha que $\ker(\psi) = \{e\}$. Se $g_1, g_2 \in \ker(\psi)$ então $\psi(g_1) = e' = \psi(g_2)$. Logo,

$$\begin{aligned}e' &= \psi(g_1)\psi(g_2)^{-1} \\ &= \psi(g_1)\psi(g_2^{-1}) \\ &= \psi(g_1g_2^{-1}).\end{aligned}$$

Portanto, $g_1g_2^{-1} \in \ker(\psi) = \{e\}$. Daí,

$$g_1g_2^{-1} = e.$$

Assim, $g_1 = g_2$ e portanto ψ é injetiva.

□

Lema 2.2.1. *Sejam G e G' e $\psi : G \rightarrow G'$ um homomorfismo. O conjunto*

$$\text{Im}(\psi) = \psi(G) = \{\psi(g) \mid g \in G\},$$

é chamado imagem do homomorfismo ψ .

O teorema a seguir nos diz que sobre determinadas condições o quociente $G/\ker(\psi) \simeq \text{Im}(\psi)$.

Teorema 2.2.1. (Teorema do Isomorfismo) *Seja $\psi : G \longrightarrow G'$ um homomorfismo de grupos. Então*

$$G/\ker(\psi) \simeq \text{Im}(\psi).$$

Uma forma de entender objetos complexos é entender objetos simples primeiro. Grupos cíclicos são grupos simples que podem ser facilmente classificados. Podemos usar grupos cíclicos como blocos de construção para formar grupos maiores. Por outro lado, podemos dividir grupos maiores em coleções de grupos cíclicos. Grupos cíclicos são fundamentais para a compreensão dos grupos abelianos.

Teorema 2.2.2. *Seja G um grupo e seja $a \in G$. Então, $J = \{a^n \mid n \in \mathbb{Z}\}$ é um subgrupo de G e é o menor subgrupo de G que contém a , ou seja, todo subgrupo que contém a contém J .*

Definição 2.2.8. (Grupo Cíclico) *O grupo J do Teorema 2.2.2 é o subgrupo cíclico de G gerado por a e será denotado por $\langle a \rangle$. Se $\langle a \rangle$ é finito, a ordem de a é $|\langle a \rangle|$. Um elemento a de um grupo G gera G e é um gerador para G se $\langle a \rangle = G$. Um grupo G é cíclico se tiver um gerador.*

Exemplo 2.2.8. $\mathbb{Z} = \langle 1 \rangle$ em adição e, portanto, é cíclico.

Definição 2.2.9. \mathbb{Z}_n é um grupo cíclico.

Podemos classificar completamente os grupos cíclicos usando o seguinte teorema

Teorema 2.2.3. *Qualquer grupo cíclico infinito é isomorfo a \mathbb{Z} sob adição. Qualquer grupo cíclico finito de ordem n é isomorfo a \mathbb{Z}_n sob adição modulo n .*

Consequentemente, podemos usar \mathbb{Z} e \mathbb{Z}_n como grupos cíclicos prototípicos.

Definição 2.2.10. *Seja G um grupo e seja $a_i \in G$ para $i \in I$. O menor subgrupo de G contendo $\{a_i \mid i \in I\}$ é o **subgrupo gerado** por $\{a_i \mid i \in I\}$. Se este subgrupo é todo G , então $\{a_i \mid i \in I\}$ gera G e a_i são os **geradores** de G . Se existe um conjunto finito $\{a_i \mid i \in I\}$ que gera G , então G é **finamente gerado**.*

Com a definição acima conseguimos classificar tais grupos através do seguinte teorema

Teorema 2.2.4. (Teorema Fundamental dos Grupos Abelianos Finitamente Gerados) *Todo grupo abeliano G finitamente gerado é isomorfo para um produto direto de grupos cíclicos da forma*

$$\mathbb{Z}_{(p_1^{r_1})} \times \mathbb{Z}_{(p_2^{r_2})} \times \cdots \times \mathbb{Z}_{(p_n^{r_n})} \times \mathbb{Z} \times \mathbb{Z} \times \cdots \times \mathbb{Z}$$

em que os p_i são primos, não necessariamente distintos. O produto direto é único exceto pelo possível arranjo de fatores; isto é, o número de fatores de \mathbb{Z} é único e as potências primárias $(p_i)^{r_i}$ são únicas.

Agora, nos aprofundamos na álgebra avançada observando estruturas algébricas cada vez mais ricas que encontraremos em nosso estudo de homologia.

Vimos que um grupo abeliano finitamente gerado é isomorfo a um produto de grupos cíclicos infinitos e finitos. Agora, caracterizaremos fatores infinitos usando a noção de grupos abelianos livres. Como trataremos apenas de grupos abelianos, usaremos $+$ para denotar a operação do grupo e 0 para o elemento de identidade. Para $n \in \mathbb{Z}^+$, $a \in G$, usamos $na = a + a + \cdots + a$ e $-na = (-a) + (-a) + \cdots + (-a)$ para denotar a soma de n fatores de a e seu inverso, respectivamente. Finalmente, $0a = 0$, onde o primeiro 0 está em \mathbb{Z} e o segundo em G . É importante perceber que G ainda é um grupo com uma única operação, adição, mesmo que usemos multiplicação em nossa notação.

Teorema 2.2.5. *Seja X um subconjunto não vazio de um grupo abeliano. As condições a seguir em X são equivalentes.*

- i.** *Cada elemento diferente de zero a em G pode ser expresso exclusivamente na forma $a = n_1x_1 + n_2x_2 + \cdots + n_rx_r$ para $n_i \neq 0$ em \mathbb{Z} e $x_i \in X$ distintos.*
- ii.** *X gera G , e $n_1x_1 + n_2x_2 + \cdots + n_rx_r = 0$ para $n_i \in \mathbb{Z}$ e $x_i \in X$ se e somente se, $n_1 = n_2 = \cdots = n_r = 0$.*

As condições acima remetem a definição de vetores linearmente independentes. Como descobriremos em breve, essa semelhança não é acidental.

Definição 2.2.11. (Grupo abeliano livre) *Um grupo abeliano que possui um grupo gerador não*

vazio X que satisfaz as condições do Teorema 2.2.5 é um grupo abeliano livre e X é uma base para o grupo.

Já vimos um grupo abeliano livre: o produto direto finito do grupo \mathbb{Z} em si é um grupo abeliano livre com uma base natural. De fato, podemos usar esse grupo como um protótipo.

Teorema 2.2.6. *Se G é um grupo abeliano livre diferente de zero com r elementos na base, então G é isomorfo a $\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z} \times \cdots \times \mathbb{Z}$, r vezes.*

2.3 Topologia

Nesta seção daremos a um conjunto à estrutura usando uma topologia para obter um espaço topológico. Uma característica importante no estudo dos espaços topológicos são as funções conhecidas como homeomorfismos. Trata-se de funções que preservam a “estrutura topológica” do seu espaço. Assim, se entre dois espaços existe um homeomorfismo, então esses espaços são topologicamente iguais.

Definição 2.3.1. (Topologia) *Uma topologia em um conjunto X é um subconjunto $T \subseteq 2^X$, tal que:*

- i. *Se $S_1, S_2 \in T$, então $S_1 \cap S_2 \in T$;*
- ii. *Se $\{S_j \mid j \in J\} \subseteq T$, então $\bigcup_{j \in J} S_j \in T$;*
- iii. *$\emptyset, X \in T$.*

A definição 2.3.1 afirma implicitamente que apenas interseções finitas e uniões infinitas dos conjuntos abertos estão abertas. Uma topologia é simplesmente um sistema de conjuntos que descreve a conectividade do conjunto. Esses conjuntos têm nomes:

Definição 2.3.2. *Seja X um conjunto e T uma topologia. $S \in T$ é um **conjunto aberto**. Os **conjuntos fechados** são $X - S$, onde $S \in T$.*

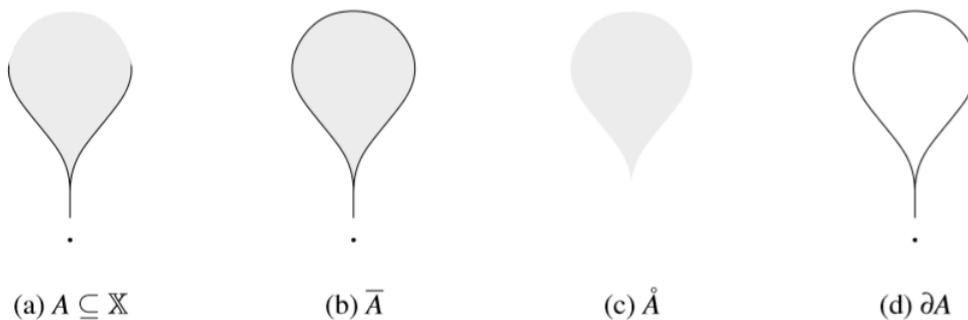
Um conjunto pode ser apenas fechado, apenas aberto, aberto e fechado, ou nenhum dos dois. Por exemplo, o conjunto vazio é aberto e fechado por definição. Combinamos um conjunto com uma topologia para obter os espaços nos quais estamos interessados.

Definição 2.3.3. (*Espaço Topológico*) O par (X, \mathcal{T}) de um conjunto X e uma topologia \mathcal{T} é um espaço topológico.

Geralmente usamos apenas X como notação para o espaço topológico (X, \mathcal{T}) quando a topologia \mathcal{T} é assumida ser conhecida.

Definição 2.3.4. (*Interior, Fecho, Fronteira*) O **interior** \mathring{A} do conjunto $A \subseteq X$ é a união de todos os conjuntos abertos contidos em A . O **fecho** \bar{A} do conjunto $A \subseteq X$ é a interseção de todos os conjuntos fechados contendo A . A **fronteira** de um conjunto A é $\partial A = \bar{A} - \mathring{A}$.

Figura 1: O conjunto $A \subseteq X$ e seus conjuntos relacionados.



Fonte: (ZOMORODIAN, 2005)

Definição 2.3.5. Uma **vizinhança** de $x \in X$ é qualquer $A \subseteq X$ tal que $x \in \mathring{A}$. Uma base de vizinhanças em $x \in X$ é uma coleção de vizinhanças de x de modo que cada vizinhança de x contenha uma das bases de vizinhança.

Uma base de vizinhanças é uma ferramenta para definir topologias, utilizaremos este fato nas definições a seguir.

Definição 2.3.6. Uma métrica ou função distância $d : X \times X \rightarrow \mathbb{R}$ é uma função que satisfaz os seguintes axiomas:

i. Para todo $x, y \in X$, temos que $d(x, y) \geq 0$;

ii. Se $d(x, y) = 0$, então $x = y$;

iii. Para todo $x, y \in X$, $d(x, y) = d(y, x)$;

iv. Para todo $x, y, z \in X$, $d(x, y) + d(y, z) \geq d(x, z)$.

Definição 2.3.7. (*Bola aberta*) A bola aberta $B(x, r)$ com o centro x e raio $r > 0$ em relação a métrica d é definida como $B(x, r) = \{y \mid d(x, y) < r\}$.

Podemos mostrar que bolas abertas podem servir de vizinhança para uma topologia de um conjunto X com uma métrica.

Definição 2.3.8. (*Espaço Métrico*) Um conjunto X com uma função métrica d é chamado de *espaço métrico*.

Um espaço métrico é um espaço topológico. A maioria dos espaços em que estamos interessados são subconjuntos de espaços métricos em particular o espaço euclidiano.

Definição 2.3.9. O produto cartesiano de \mathbb{R} n vezes, juntamente com a métrica euclidiana

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

é o espaço euclidiano n -dimensional, \mathbb{R}^n .

Agora, definimos uma relação de equivalência nos espaços topológicos.

Definição 2.3.10. (*homeomorfismo*) Um homeomorfismo $f : X \rightarrow Y$ é uma função 1 – 1, de modo que f, f^{-1} são contínuos. Dizemos que X é homeomórfico a Y , $X \approx Y$ e que X e Y têm o mesmo tipo topológico.

A seguir veremos que variedades são um tipo de espaço topológico em que estamos interessados. Eles correspondem bem aos espaços com os quais estamos mais familiarizados, os espaços euclidianos. Intuitivamente, uma variedade é um espaço topológico que se parece localmente com \mathbb{R}^n . Em outras palavras, cada ponto admite um sistema de coordenadas, consistindo em funções de coordenadas nos pontos da vizinhança, determinando a topologia da vizinhança.

Definição 2.3.11. (Carta) Uma carta em $p \in X$ é uma função $\phi : U \rightarrow \mathbb{R}^d$, onde $U \subseteq X$ é um conjunto aberto contendo p e ϕ é um homeomorfismo sobre um subconjunto aberto de \mathbb{R}^d . A dimensão da carta ϕ é d . As funções coordenadas de $x^i = u^i \circ \phi; U \rightarrow \mathbb{R}$, em que $u^i : \mathbb{R}^n \rightarrow \mathbb{R}$ são as coordenadas padrão \mathbb{R}^d .

Definição 2.3.12. (Hausdorff) Um espaço topológico X é Hausdorff se, para cada $x, y \in X, x \neq y$, houver vizinhanças U, V de x, y , respectivamente, de modo que $U \cap V = \emptyset$.

Definição 2.3.13. Um espaço topológico X é separável se possui uma base enumerável de vizinhanças.

Finalmente, definamos o que é uma variedade

Definição 2.3.14. (Variedade) Um espaço Hausdorff separável X é chamado de d -variedade se houver uma carta d -dimensional em cada ponto $x \in X$, ou seja, se $x \in X$ tiver uma vizinhança homeomórfica para \mathbb{R}^n . É chamado de d -variedade com fronteira se $x \in X$ tiver uma vizinhança homeomórfica para \mathbb{R}^d ou o meio espaço euclidiano $\mathbb{H}^d = \{x \in \mathbb{R}^d \mid x_1 \geq 0\}$. A fronteira de X é o conjunto de pontos com vizinhança homeomórfica para \mathbb{H}^d . A variedade tem dimensão d .

Definição 2.3.15. Uma imersão $f : X \rightarrow Y$ é um mapa cuja restrição à sua imagem $f(Y)$ é um homeomorfismo.

2.4 Homologia Simplicial

Nesta seção vamos utilizar métodos da topologia algébrica que permitem a contagem de ocorrências de padrões geométricos em um espaço topológico, tentaremos distinguir as ocorrências uma das outras ou encontrar características similares entre elas. Discutiremos sobre complexos simpliciais e sua realização geométrica e, posteriormente, introduziremos grupos de homologia. Esses conceitos formarão a base para nossa abordagem topológica dos dados.

Começemos olhando para a definição geométrica.

Seja $S = \{x_0, x_1, x_2, \dots, x_k\}$ um conjunto de pontos no espaço Euclidiano \mathbb{R}^d . Dizemos que $x \in \mathbb{R}^d$ é **combinação afim** de S se existem $\alpha_0, \alpha_1, \dots, \alpha_k$, onde $\sum_{i=0}^k \alpha_i = 1$ e x pode ser

escrito como

$$x = \sum_{i=0}^k \alpha_i \cdot x_k$$

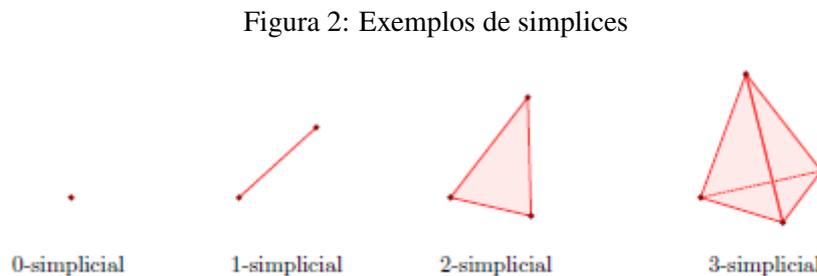
Uma combinação afim é dita uma **combinação convexa** se $\alpha_i \geq 0$ para cada i . Os números $\alpha_0, \alpha_1, \dots, \alpha_k$ são chamados de **coordenadas baricêntricas** de x . Para o conjunto de todos os pontos que são combinações convexas de S , denominamos **cobertura convexa** de S . Se nenhum ponto x_i de S é uma combinação de $S - \{x_i\}$, dizemos que S é um **conjunto linearmente independente**.

Exemplo 2.4.1. ($k=1$) Suponha que $S = \{x_0, x_1\} \subset \mathbb{R}^2$ onde $x_0 \neq x_1$. O segmento que passa por x_0 e x_1 é o conjunto de pontos que podem ser escritos em função de um parâmetro α como

$$x_0 + \alpha(x_1 - x_0) = (1 - \alpha)x_0 + \alpha x_1$$

Observe que os pontos no segmento que conectam x_0 a x_1 são obtidos para $0 \leq \alpha \leq 1$. Se chamarmos $\alpha_0 = 1 - \alpha$ e $\alpha_1 = \alpha$, fica claro que $\alpha_0, \alpha_1 \geq 0$ e $\alpha_0 + \alpha_1 = 1$. Portanto, a cobertura convexa de S é o segmento fechado que conecta x_0 a x_1 .

Definição 2.4.1. Um *simplex* (ou *simplicial*) de dimensão k ou k -simplex σ é a cobertura convexa de $(n + 1)$ pontos independentes em \mathbb{R}^n .



Fonte: Próprio Autor

Em outras palavras, um k -simplex é o menor conjunto convexo do espaço euclidiano \mathbb{R}^n que contém $n + 1$ pontos que não pertencem a um hiperplano k -dimensional. Um k -simplex

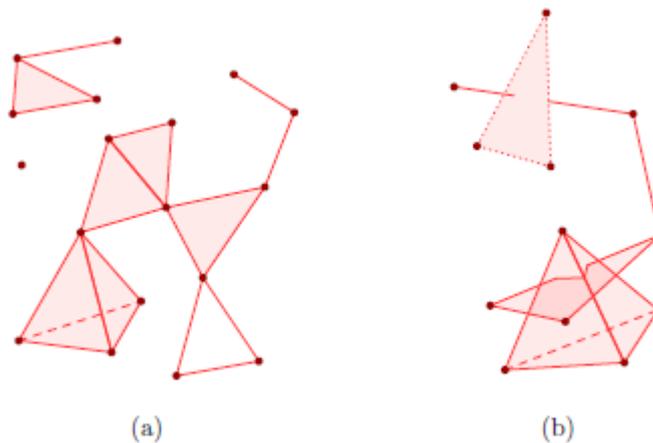
é um subespaço topológico k -dimensional de um espaço euclidiano, isto é, a dimensão de um k -simplex σ é $\dim\sigma = k$.

Definição 2.4.2. Um complexo simplicial K , é um conjunto finito de simplices que satisfazem:

- i. Se $\sigma \in K$ e τ é face de σ , então $\tau \in K$.
- ii. Se $\sigma, \tau \in K$ e $\sigma \cap \tau = \emptyset$, então $\sigma \cap \tau$ é uma face de σ e τ

Ou seja, um complexo simplicial S é uma coleção finita de simplices tais que cada face de um simplex de S é um simplex de S e a intersecção de dois simplices de S é vazia ou uma face de ambos.

Figura 3: (a) Exemplo de um complexo simplicial (b) exemplo de um complexo não simplicial



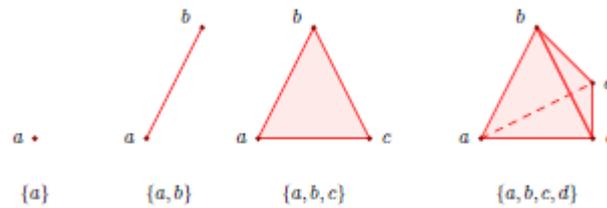
Fonte: Próprio Autor

Note que qualquer complexo simplicial determina um objeto combinatório consistindo de subconjuntos do conjunto de todos os vértices, motivando a seguinte definição:

Definição 2.4.3. Por um complexo simplicial abstrato K , entenderemos o par $K = (V(K), \Sigma(K))$, onde $V(K)$ é um conjunto finito chamado de vértices de K onde $\Sigma(K)$ é um subconjunto (chamado de simplexos) da coleção de todos os subconjuntos não-vazios de $V(K)$, satisfazendo as seguintes condições:

- i. Cada elemento de $V(K)$ pertence a algum elemento de $\Sigma(K)$ e
- ii. Se σ é um elemento em $\Sigma(K)$, então todo subconjunto não vazio de σ também é um elemento de $\Sigma(K)$.

Figura 4: Correspondência entre os complexos simpliciais e os complexos simpliciais abstratos.



Fonte: Próprio Autor

Definição 2.4.4. (*Esquema de Vértices*) Seja K um complexo simplicial com os vértices V e \mathcal{K} seja a coleção de todos os subconjuntos $\{v_0, v_1, \dots, v_k\}$ de V , de modo que os vértices v_0, v_1, \dots, v_k abranjam um simplex de K . A coleção \mathcal{K} é chamada de esquema de vértices de K .

A coleção \mathcal{K} é um complexo simplicial abstrato assim podemos comparar complexos simpliciais facilmente, usando isomorfismos.

Definição 2.4.5. (*Isomorfismo*) Seja K_1, K_2 complexos simpliciais abstratos com conjuntos de vértices V_1, V_2 , respectivamente. Um isomorfismo entre K_1, K_2 é uma bijeção $\varphi : V_1 \rightarrow V_2$, de modo que os conjuntos em K_1 e K_2 são os mesmos sob a renomeação dos vértices por φ e seu inverso.

O próximo teorema, cuja prova pode ser encontrada em (MUNKRES, 2018), nos permitirá abandonar os termos “geométrico” e “abstrato” e chamar apenas de complexo simplicial.

Teorema 2.4.1. (*Realização Geométrica*) Todo complexo simplicial abstrato S é isomorfo ao esquema de vértices de algum complexo simplicial geométrico K . E dois complexos simpliciais geométricos são isomorfos se, e somente se, seus esquemas de vértices são isomorfos como complexos simpliciais abstratos.

Se um complexo simplicial abstrato S é isomorfo ao esquema de vértices de algum complexo simplicial geométrico K , dizemos que K é uma **realização geométrica** de S .

Podemos usar complexos simpliciais para representar variedades e outros espaços topológicos por seu espaço subjacente. O espaço subjacente $|X|$ de um complexo simplicial X é $|X| = \bigcup_{\sigma \in K} \sigma$. Por si só, $|X|$ é um espaço topológico. Quando um espaço topológico \mathbb{X} é homeomórfico ao espaço subjacente de um complexo simplicial X , dizemos que $|X|$ é uma **triangulação** de \mathbb{X} . Um exemplo clássico de triangulação é o da esfera que pode ser triangulada como um tetraedro.

Antes de introduzir grupos de homologia, precisamos primeiro definir a orientação no contexto de complexos simpliciais.

Definição 2.4.6. *Um simplex orientado é uma ordenação dos vértices do simplex. Dois simplex tem a mesma orientação se a permutação de seus índices tem sinal positivo.*

Definição 2.4.7. *Seja K um complexo simplicial e p uma dimensão. Uma p -cadeia c é a combinação linear de simplex orientados, $c = \sum a_i \sigma_i$ onde σ_i são p -simplexos orientados e $a_i \in \mathbb{Z}$ são coeficientes. Denotamos $C_p = C_p(K)$ o grupo de p -cadeias em K .*

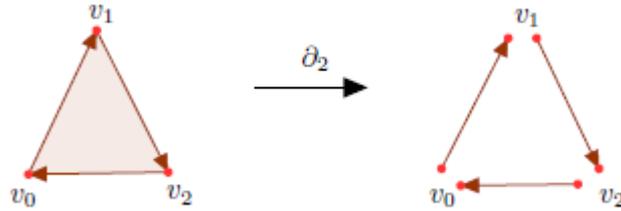
Definição 2.4.8. *(Operador Fronteira) Definimos o operador $\partial_p : C_p(K) \rightarrow C_{p-1}(K)$ por:*

$$\partial[v_0, v_1, \dots, v_p] = \sum_{i=0}^p (-1)^i [v_0, \dots, \hat{v}_i, \dots, v_p]$$

para um simplex orientado $[v_0, v_1, \dots, v_p]$ onde \hat{v}_i indica que v_i foi retirado da sequência. O operador fronteira define um homomorfismo e por este motivo o denotamos homomorfismo de fronteira.

Exemplo 2.4.2. *Seja λ um complexo simplicial orientado $[v_0, v_1, v_2]$ o operador fronteira ∂_2 é dado como:*

Figura 5: Exemplo da ação do operador fronteira.



Fonte: Próprio Autor.

$$\partial_2[v_0, v_1, v_2] = \sum_{i=0}^2 (-1)^i [v_0, \dots, \hat{v}_i, \dots, v_2] = [v_1, v_2] - [v_0, v_2] + [v_0, v_1]$$

Exemplo 2.4.3. Na figura 6 temos um losango composto por dois triângulos, cada um marcado para indicar a ordem dos vértices correspondentes à orientação positiva, adotamos a convenção de que uma orientação positiva corresponde à ordem em que os vértices seriam visitadas por um caminho no sentido anti-horário ao redor da fronteira do simplex. Se interpretarmos o 1-simplex (v_j, v_k) como o segmento orientado do vértice v_j para o vértice v_k temos que

$$A = (v_3, v_0) + (v_0, v_1) + (v_1, v_3) \quad (2.2)$$

corresponde a fronteira de (v_0, v_1, v_3) . De maneira análoga,

$$B = (v_3, v_1) + (v_1, v_2) + (v_2, v_3) \quad (2.3)$$

corresponde a fronteira de (v_1, v_2, v_3) e a soma

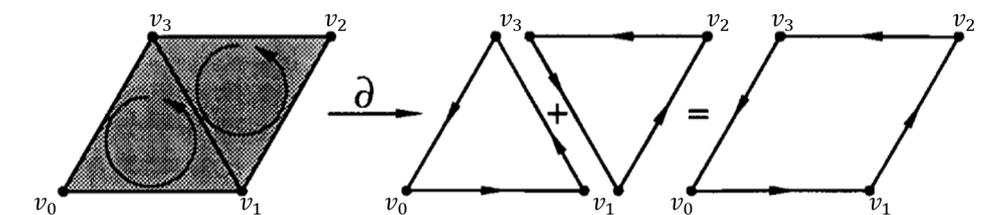
$$\begin{aligned} A + B &= (v_3, v_0) + (v_0, v_1) + (v_1, v_3) + (v_3, v_1) + (v_1, v_2) + (v_2, v_3) \\ &= (v_3, v_0) + (v_0, v_1) + (v_1, v_3) - (v_1, v_3) + (v_1, v_2) + (v_2, v_3) \\ &= (v_3, v_0) + (v_0, v_1) + (v_1, v_2) + (v_2, v_3) \end{aligned}$$

corresponde a fronteira de toda a região. Observe que usamos

$$(v_1, v_3) + (v_3, v_1) = (v_1, v_3) - (v_1, v_3) = 0,$$

uma relação que faz sentido geometricamente e algebricamente. Na interpretação anterior, $(v_1, v_3) + (v_3, v_1)$ corresponde a um caminho que se inverte e, portanto, não leva a lugar algum. Neste último, definimos $(v_3, v_1) = -(v_1, v_3)$: uma relação algébrica que diz que um simplex de orientação negativa é o inverso aditivo de seu par de orientação positiva.

Figura 6: A ação do operador fronteira.



Fonte: Adaptado de (MULDOON *et al.*, 1993)

Observe que a fronteira de um p -simplex é exatamente as $(p - 1)$ faces dimensionais dele, com uma orientação que é dada naturalmente pela fórmula. Se tomarmos a fronteira da fronteira determinamos o seguinte resultado:

Teorema 2.4.2. *Seja $\sigma = [v_0, v_1, \dots, v_p]$ um simplex orientado então*

$$\partial^2 \sigma = 0$$

Demonstração. Seja $\sigma = [v_0, v_1, \dots, v_p]$. Da definição temos que

$$\begin{aligned}
\partial^2 &= \partial \left(\sum_i (-1)^i [v_0, \dots, \hat{v}_i, \dots, v_p] \right) \\
&= \sum_{j < i} (-1)^i (-1)^j [v_0, \dots, \hat{v}_j, \dots, \hat{v}_i, \dots, v_p] \\
&\quad + \sum_{j > i} (-1)^i (-1)^{j-1} [v_0, \dots, \hat{v}_i, \dots, \hat{v}_j, \dots, v_p]
\end{aligned}$$

Como alternamos i e j na segunda soma, todos os termos que estão na primeira também estarão na segunda com o sinal negativo. Assim $\partial^2 \sigma = 0$. \square

Definição 2.4.9. (*Complexo de Cadeias*) Definimos o complexo de cadeias como sendo uma sequência de grupos de cadeia conectados por homomorfismos de fronteira.

$$\dots \xrightarrow{\partial_{p+1}} C_p \xrightarrow{\partial_p} C_{p-1} \xrightarrow{\partial_{p-1}} \dots \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0$$

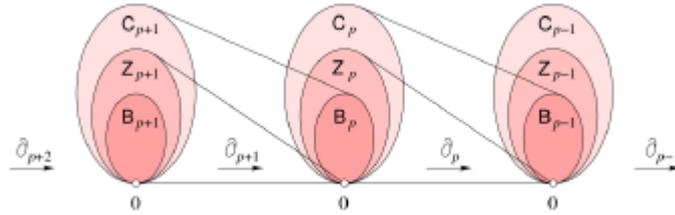
Definição 2.4.10. (*p-ciclo*) Um p -ciclo c é uma p -cadeia com fronteira vazia, ou seja, $\partial c = 0$. Denotamos por $Z_p = Z_p(K)$ o grupo de p -ciclos.

Definição 2.4.11. (*p-fronteira*) Uma p -fronteira c é uma p -cadeia que é uma fronteira de uma $(p + 1)$ cadeia, ou seja, $\partial c = \partial d$, com $d \in C_{p+1}$. Denotamos por $B_p = B_p(K)$ o grupo de fronteiras.

Definição 2.4.12. Definimos $Z_p = \text{Ker}(\partial_p)$ e $B_p = \text{Im}(\partial_{p+1})$.

Toda p -fronteira é também um p -ciclo, ou equivalentemente, B_p é um subgrupo de Z_p .

Figura 7: A seqüência de complexos de cadeia, ciclos e grupos de fronteira conectados por homomorfismos.



Fonte: (ZOMORODIAN, 2005)

Teorema 2.4.3. Os conjuntos $Z_p = \text{Ker}(\partial_p)$ e $B_p = \text{Im}(\partial_{p+1})$ são subgrupos normais abelianos livres de C_p . $\text{Im}(\partial_{p+1})$ também é um subgrupo normal de $\text{Ker}(\partial_p)$.

Demonstração. A demonstração deste resultado foi omitida mas pode ser encontrada em (MUNKRES, 2018) □

Definição 2.4.13. (Grupo de Homologia) O p -ésimo grupo de homologia é o grupo quociente

$$H_p(X) := \frac{\text{ker}(\partial_p)}{\text{im}(\partial_{p+1})} = \frac{Z_p}{B_p}$$

Se os p -ciclos z_1 e z_2 pertencem a mesma classe de Z_p , dizemos que são **homólogos** e denotamos $z_1 \sim z_2$.

Exemplo 2.4.4. Podemos representar o círculo como um complexo simplicial de dimensão 1 de vértices $\{v_0, v_1, v_2\}$. Escolhendo a orientação $[v_0, v_1], [v_1, v_2], [v_2, v_0]$,

$$C_2 \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0$$

Então,

$$\begin{aligned} \partial_1(c) &= \partial_1(\lambda_0[v_0, v_1] + \lambda_1[v_1, v_2] + \lambda_2[v_2, v_0]) \\ &= \lambda_0(v_1 - v_0) + \lambda_1(v_2 - v_1) + \lambda_2(v_0 - v_2) \\ &= (\lambda_2 - \lambda_0)v_0 + (\lambda_0 - \lambda_1)v_1 + (\lambda_1 - \lambda_2)v_2 \end{aligned} \tag{2.4}$$

e isso é 0 se e somente se $\lambda_0 = \lambda_1 = \lambda_2$. Portanto,

$$\ker(\partial_1) = \{\lambda([v_0, v_1] + [v_1, v_2] + [v_2, v_0]) : \lambda \in \mathbb{Z} \approx \mathbb{Z}\}$$

Por outro lado, $\text{Im}(\partial_2) = 0$. Logo,

$$H_1(K) = \frac{Z_1}{B_1} = \frac{\ker(\partial_1)}{\text{Im}(\partial_2)} \approx \mathbb{Z}$$

Dado que $\partial_0 = 0$ então $\ker(\partial_0) \approx \mathbb{Z}^3$ e

$$\text{Im}(\partial_1) = \{\mu_0 v_0 + \mu_1 v_1 + \mu_2 v_2 : \mu_2 = -(\mu_0 + \mu_1) \text{ e } \mu_0, \mu_1 \in \mathbb{Z}\}$$

Portanto, $\text{Im}(\partial_1) \approx \mathbb{Z}^2$. Logo,

$$H_0(K) = \frac{Z_0}{B_0} = \frac{\ker(\partial_0)}{\text{Im}(\partial_1)} \approx \mathbb{Z}$$

Assim, podemos concluir que os grupos de homologia do círculo são:

$$\begin{aligned} H_0(K) &= \mathbb{Z} \\ H_1(K) &= \mathbb{Z} \\ H_p(K) &= 0, \quad p \geq 2 \end{aligned} \tag{2.5}$$

Como são grupos de fatores de dois grupos abelianos livres, os grupos de homologia são abelianos finamente gerados. Portanto, o teorema fundamental dos grupos abelianos finamente gerados se aplica. Assim, podemos usar os grupos de homologia para descrever espaços através de invariantes topológicos chamados números de Betti, mas antes disso definamos o que é um invariante topológico:

Definição 2.4.14. (*Invariante Topológico*) Um invariante é um mapa que atribui o mesmo objeto a espaços topológicos do mesmo tipo.

Definição 2.4.15. (*Números de Betti*) O k -ésimo número de Betti β_k de um complexo simplicial K é $\beta(H_p)$, o rank da parte livre de H_p .

Ou seja, os Números de Betti são usados para distinguir espaços topológicos com base na conectividade de complexos simpliciais n - dimensionais. Informalmente, o de k -ésimo número Betti refere-se ao número de furos k - dimensionais em uma superfície topológica.

Exemplo 2.4.5. *Abaixo os números de Betti de um ponto, disco, esfera, toro, e garrafa de klein.*

Tabela 1: Números de Betti para um único ponto, uma circunferência, esfera, um toro e para a garrafa de Klein. Em um espaço k -dimensional, o n -ésimo número de Betti é sempre zero para qualquer $n \geq k$.

	β_0	β_1	β_2	β_3
Ponto	1	0	0	0
Circunferência	1	1	0	0
Esfera	1	0	1	0
Toro	1	2	1	0
Garrafa de Klein	1	2	1	0

Fonte: Próprio Autor

Outro famoso invariante é a característica de Euler

Definição 2.4.16. *Seja K um complexo simplicial e $s_i = \text{card}\{\sigma \in K \mid \dim\sigma = i\}$. A característica de Euler $\chi(K)$ é*

$$\chi(K) = \sum_{i=0}^{\dim K} (-1)^i s_i = \sum_{\sigma \in K - \{\emptyset\}} (-1)^{\dim\sigma} \quad (2.6)$$

Embora seja definida para um complexo simplicial, a característica de Euler é um invariante para $|K|$, o espaço subjacente de K . Dada qualquer triangulação de um espaço \mathbb{M} , sempre obteremos o mesmo número inteiro, que chamaremos de característica de Euler desse espaço $\chi(\mathbb{M})$.

CAPÍTULO 3

Análise Topológica de Dados

Neste capítulo desenvolveremos a teoria que nos dará condição de analisar as séries temporais topologicamente. Para isso faremos adaptações do que foi apresentado no capítulo 2.

A Análise Topológica de Dados (TDA) é uma área emergente para a análise de dados complexos. TDA refere-se a uma classe de métodos que obtém informações de estruturas topológicas em dados que pertencem a um espaço topológico, ou seja, um espaço matemático que permite continuidade, conexão e convergência (CARLSSON, 2009). A saída do TDA pode então ser usada para aprendizado estatístico eficaz sobre os dados. O TDA combina topologia algébrica e outras ferramentas matemáticas para permitir um estudo útil da forma dos dados. As topologias mais discutidas dos dados incluem componentes conectados, buracos, etc., de um espaço topológico. Topologia Computacional (ou algorítmica), é uma sobreposição entre os fundamentos matemáticos da topologia com ciência da computação e consiste em duas partes, isto é, obter a topologia de um espaço e uma homologia persistente (CHAZAL *et al.*, 2009). Usando topologia computacional, o TDA visa analisar recursos topológicos de dados e representar esses recursos usando representações de baixa dimensão (CARLSSON, 2009).

3.1 Homologia Persistente

Os grupos de homologia, em geral nos dão uma boa descrição da geometria dos espaços, porém para nossos problemas onde os espaços são finitos, pouca informação podemos obter a partir deles. Para nossos propósitos, que é a aplicação de TDA, utilizaremos elementos de

homologia simplicial. Faremos uma adaptação da teoria da homologia usual, mais precisamente, usaremos a homologia persistente.

Homologia persistente refere-se a uma classe de métodos para medir características topológicas de formas e funções. Ele converte os dados em complexos simpliciais e descreve a estrutura topológica de um espaço em diferentes resoluções espaciais. As topologias mais persistentes são detectadas em uma ampla gama de escalas espaciais e são consideradas mais propensas a representar características verdadeiras do espaço subjacente do que as variações de amostragem, ruído etc. A homologia persistente, portanto, provoca a persistência de topologias essenciais nos dados e gera o nascimento e a morte de tais topologias por meio de um diagrama de persistência, que é uma estatística resumida popular no TDA. As entradas de dados para homologia persistente são geralmente representadas como nuvens de pontos ou funções, enquanto as saídas dependem da natureza da análise e geralmente consiste em um diagrama de persistência. Uma nuvem de pontos representa uma amostra de pontos de uma variedade subjacente e sua homologia persistente aproxima as informações topológicas da variedade. Se os dados são representados como uma função Morse (isto é, uma função suave na variedade, de modo que todos os pontos críticos não sejam degenerados), a homologia persistente da função é matematicamente equivalente à análise das informações topológicas da variedade.

No capítulo 2, desenvolvemos uma abordagem para explorar a topologia de um espaço. Estudamos um método algébrico para calcular a topologia usando grupos de homologia. Aplicando homologia às “filtrações”, obtemos caracterizações para um espaço. Nos interessa os intervalos que persistem por muito tempo, uma vez que tendem a refletir as verdadeiras propriedades homológicas do objeto subjacente.

3.1.1 Filtrações e Persistência

Consideremos um complexo simplicial K , e uma função $f : K \rightarrow \mathbb{R}$. Seja f uma função monótona, ou seja, para σ, τ temos que $f(\sigma) \leq f(\tau)$ sempre que σ é face de τ .

A monotonicidade de f implica que o conjunto $K(a) = f^{-1}(-\infty, a]$ é um subcomplexo de K para cada $a \in \mathbb{R}$. Seja m o número de simplex em K e n um número de valores de f , então

obtemos $n + 1 \leq m + 1$ diferentes subcomplexos que ordenamos de maneira crescente e que nos permite fazer a seguinte definição:

Definição 3.1.1. (*Filtração*) Uma filtração de K é uma sequência aninhada de subcomplexos que começa com o complexo vazio e termina com o complexo completo,

$$\emptyset = K_0 \subseteq K_1 \subseteq \dots \subseteq K_{n-1} \subseteq K_n = K$$

Intuitivamente, podemos pensar numa filtração como a construção de um complexo adicionando um simplex de cada vez. Mais que a sequência de complexos, estamos interessados no comportamento topológico que é dado pela sequência de grupos de homologia associados a cada filtração.

Para cada $i \leq j$ temos uma aplicação de inclusão natural do espaço subjacente de K_i para K_j :

$$\emptyset = K_0 \hookrightarrow K_1 \hookrightarrow \dots \hookrightarrow K_{n-1} \hookrightarrow K_n = K$$

Portanto, temos um homomorfismo induzido entre os grupos de homologia associados a cada complexo na filtração para cada dimensão p :

$$\emptyset = H_p(K_0) \xrightarrow{\varphi_p^{0,1}} H_p(K_1) \xrightarrow{\varphi_p^{1,2}} \dots \xrightarrow{\varphi_p^{n-2,n-1}} H_p(K_{n-1}) \xrightarrow{\varphi_p^{n-1,n}} H_p(K_n) = H_p(K)$$

Assim, temos que quando vamos de K_{i-1} a K_i obtemos novas classes de homologia e perdemos algumas que se convertem na classe de equivalência trivial ou se juntam com outras classes. Assim, agrupamos essas classes que nascem com ou antes de um determinado valor e morrem após outro valor. As seguintes definições formalizam esta ideia

Definição 3.1.2. (*Grupos de Homologia Persistente*) Os p -ésimos grupos de homologia persistente são as imagens dos homomorfismos induzidos pela inclusão $H_p^{i,j} = \text{Im}\varphi_p^{i,j}$ para $0 \leq i \leq j \leq n$. Os correspondentes p -ésimos números de Betti persistentes são os ranks destes grupos, $\beta_p^{i,j} = \text{rank}H_p^{i,j}$.

Observação 3.1.1. Note que $H_p^{i,i} = H_p(K_i)$ e $H_p^{i,j} \subseteq H_p(K_j)$

Os grupos de homologia consistem nas classes de homologia de K_i que ainda estão presentes ou “vivem” no grupo de homologia de K_j , ou seja

$$H_p^{i,j} = \frac{Z_p(K_i)}{B_p(K_j) \cap Z_p(K_i)}$$

Temos que estes grupos para cada dimensão p e cada par de índices $i \leq j$. Podemos ser mais concretos a cerca das classes contabilizadas pelos grupos de homologia persistente.

Definição 3.1.3. (*Nascimento e morte de classes*) Seja γ uma classe em $H_p(K_i)$, dizemos que nasce em K_i se $\gamma \notin H_p^{i-1,i}$. Se γ nasce em K_i então morre em K_j caso se una com uma classe anterior indo de K_{j-1} a K_j , ou seja,

$$\varphi_p^{i,j}(\gamma) \in H_p^{i-1,j} \quad e \quad \varphi_p^{i,j-1}(\gamma) \notin H_p^{i-1,j-1}$$

Definição 3.1.4. (*Persistência*) Se γ nasce em K_i e morre entrando em K_j então chamamos a diferença, no valor da função, a persistência de γ , definida como $\text{pers}(\gamma) = a_j - a_i$. A diferença entre os índices i e j é chamada índice de persistência. Se γ nasce em K_i mas não morre, então temos que sua persistência é infinita.

Observe que o nascimento e a morte também podem ser vistos como uma sequência de espaços vetoriais que não necessariamente são os grupos de homologia. Precisamos apenas de uma sequência finita e homomorfismos de esquerda e direita que para espaços vetoriais são chamados aplicações lineares.

3.1.2 Diagramas de Persistência

Aqui, introduziremos os diagramas de persistência que nos permitem visualizar toda a informação dos grupos de homologia no processo de filtração.

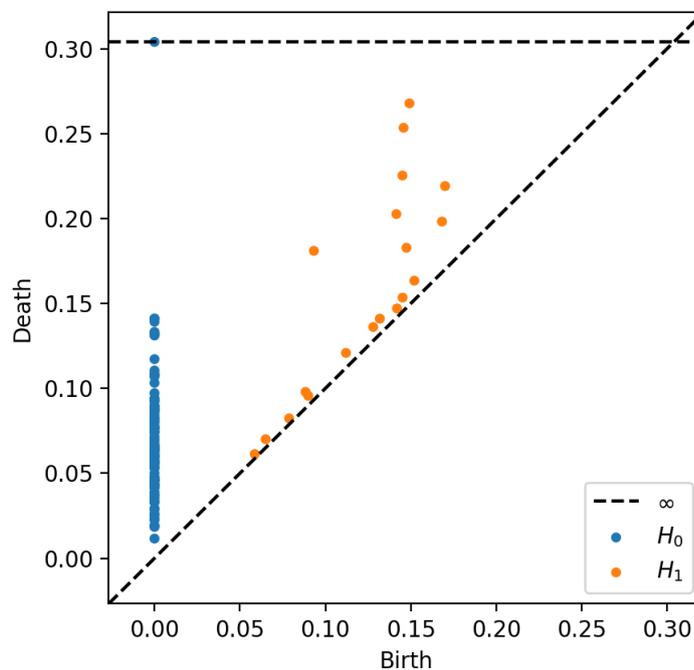
Definição 3.1.5. Definimos $\mu_p^{i,j}$ o número de classes p -dimensionais que nascem em K_i e morrem em K_j , ou seja,

$$\mu_p^{i,j} = (\beta_p^{i,j-1} - \beta_p^{i,j}) - (\beta_p^{i-1,j-1} - \beta_p^{i-1,j})$$

para todo $i < j$ e todo p .

Definição 3.1.6. (Diagrama de persistência) Marcando cada ponto (a_i, a_j) com multiplicidade $\mu_p^{i,j}$, obtemos o p -ésimo diagrama de persistência da filtração denotado por $Dgm_p(f)$.

Figura 8: Exemplo de um diagrama de Persistência.



Fonte: Próprio Autor

A figura acima ilustra um diagrama de persistência. Por construção, todos os valores estão acima da diagonal e as classes mais próximas a diagonal são as classes que persistem por pouco tempo, por outro lado as classes mais próximas a horizontal superior representam as classes que persistem por um maior tempo após filtração.

O próximo resultado nos diz que o diagrama traz consigo toda a informação dos grupos de homologia persistente. Assim, teremos uma maneira direta de calcular o diagrama de persistência e então teremos toda a informação topológica.

Teorema 3.1.1. (Lema Fundamental da Homologia Persistente) Seja $\emptyset = K_0 \subseteq K_1 \subseteq \dots \subseteq K_n = K$ uma filtração. Para cada par de índices $0 \leq k \leq l \leq n$ e cada dimensão p , o p -ésimo

Número de Betti Persistente é dado

$$\beta_p^{k,l} = \beta_p^{k,n} + \sum_{i \leq k} \sum_{j > l} \mu_p^{i,j} \quad (3.1)$$

Demonstração. Observe que de acordo com a definição 3.1.5 temos:

$$\begin{aligned} \sum_{i \leq k} \sum_{j > l} \mu_p^{i,j} &= \sum_{i \leq k} \sum_{j > l} (\beta_p^{i,j-1} - \beta_p^{i,j}) - (\beta_p^{i-1,j-1} - \beta_p^{i-1,j}) \\ &= \sum_{i \leq k} (\beta_p^{i,l} - \beta_p^{i,n}) - (\beta_p^{i-1,l} - \beta_p^{i-1,n}) \\ &= \sum_{i \leq k} (\beta_p^{i,l} - \beta_p^{i-1,l}) - (\beta_p^{i,n} - \beta_p^{i-1,n}) \\ &= (\beta_p^{k,l} - \beta_p^{0,l}) - (\beta_p^{k,n} - \beta_p^{0,n}) \end{aligned}$$

Mas, $\beta_p^{0,j} = 0$ para todo j pois $K_0 = \emptyset$ não tem nenhuma classe de homologia. Portanto,

$$\beta_p^{k,l} = \beta_p^{k,n} + \sum_{i \leq k} \sum_{j > l} \mu_p^{i,j}$$

□

A seguir definimos algumas métricas para o espaço dos diagramas de persistência

Definição 3.1.7. (*Distância de Wasserstein*) A distância de Wasserstein entre dois diagramas de persistência X e Y é

$$W_p(X, Y) = \inf_{\varphi: X \rightarrow Y} \left[\sum_{x \in X} \|x - \varphi(x)\|_q \right]^{1/p} \quad (3.2)$$

em que $1 \leq p, q \leq \infty$, φ varia sobre todas as bijeções e $\|(x, y)\|_\infty = \max\{|x|, |y|\}$ é a norma L_∞ usual.

A distância a seguir é um caso especial da distância de Wasserstein, permitindo que $p = \infty$.

Definição 3.1.8. (*Distância de Bottleneck*) A distância de Bottleneck entre dois diagramas de

persistência X e Y é definida como

$$W_\infty(X, Y) = \inf_{\varphi: X \rightarrow Y} \sup_{x \in X} \|x - \varphi(x)\|_q \quad (3.3)$$

Ou seja, a distância acima mede o grau de similaridade entre dois diagramas de persistência.

CAPÍTULO 4

Análise Topológica de Dados para Séries Temporais Agroambientais

Neste capítulo utilizaremos TDA para avaliar séries temporais de commodities agrícolas. Para isso inicialmente faremos um estudo de como converter uma série temporal em uma nuvem de pontos. Ao final do capítulo apresentamos as séries que serão estudadas e aplicaremos TDA a elas.

4.1 TDA para Séries Temporais

Métodos inspirados nas propriedades topológicas e geométricas adotam o seguinte pipeline:

- i. Supõe-se que os dados sejam uma nuvem de pontos em um espaço métrico distribuídos por pontos ou com conectividade intrínseca aos pares, definida por uma matriz de incidência. A métrica definida para os dados geralmente é fornecida experimentalmente ou construída por algum procedimento.
- ii. Uma cadeia de complexos simpliciais chamada filtração, é construída sobre os dados. Essa estrutura reflete os dados em diferentes escalas. A principal dificuldade aqui é como definir a filtração para obter informações relevantes sobre os dados de uma maneira que possa ser fácil de construir e manipular computacionalmente.

iii. Invariantes topológicos e informações geométricas são extraídos dos dados, e esse conhecimento pode ser visto como impressão digital dos dados. Eles podem ser usados para entender melhor os dados ou ser combinados com outros recursos para análise posterior.

Como as entradas para análise utilizando a abordagem de TDA são nuvem de pontos e como nossos dados são descritos por séries temporais precisamos fazer a “conversão” de modo que características topológicas importantes não sejam perdidas. Para isso utilizaremos alguns métodos que reconstruam o espaço de fase onde a variedade está inserida.

Inspirado por um trabalho inicial de (MULDOON *et al.*, 1993) utilizamos um método geral para a construção de complexo simplicial a partir de séries temporais observadas de sistemas dinâmicos com base no procedimento de reconstrução por coordenadas com atraso. Assim, o complexo simplicial obtido preservará todas as características topológicas pertinentes do espaço de fase reconstruído e pode ser analisado a partir de aspectos topológicos, combinatórios e algébricos.

Uma das descobertas importantes da teoria dos sistemas dinâmicos é a existência de atratores estranhos nos espaços de fase dos sistemas não lineares. Geralmente, as informações sobre a dinâmica são obtidas a partir de observações diretas de um número relativamente pequeno de variáveis dinâmicas, e a importante contribuição nessa direção é a da “embedologia” (SAUER; YORKE; CASDAGLI, 1991), um ramo da teoria de sistemas dinâmicos que aborda um problema importante de determinar certas características de espaço de fase apenas a partir de séries temporais de um única observação.

No entanto, o campo da “embedologia” não é um campo de pesquisa completo, apesar das técnicas confiáveis que foram desenvolvidas e usadas com sucesso por algum tempo. No procedimento padrão, o ponto de partida é uma série temporal escalar observada, digamos $X = \{x_1, x_2, \dots, x_n\}$ de alguma variável X medida em intervalos de tempo pares, e o sinal é elevado para d dimensões usando a imersão atrasada para que o sinal se torne um caminho em um espaço de fase de maior dimensão, de dimensão e forma desconhecidas. Os pontos no espaço reconstruído geralmente convergem para uma variedade ou outro subespaço do espaço euclidiano n -dimensional ou atratores da dimensão fractal.

O teorema de Takens assume condições idealizadas no sentido de que as séries temporais observadas não são contaminadas por ruído. Do ponto de vista prático, o procedimento de mapeamento do vetor de estado de um sistema dinâmico para um ponto no espaço reconstruído consiste em coletar m amostras uniformemente espaçadas das séries temporais observadas da variável apropriada e concatená-las em um único vetor.

$$\vec{v}_i(t) = [x_t, x_{t-\tau}, x_{t-2\tau}, \dots, x_{t-(d-1)\tau}]$$

Os principais parâmetros no procedimento de reconstrução são τ , o atraso e d , a dimensão de imersão. O teorema de imersão de Takens (TAKENS, 1981) garante a preservação das propriedades topológicas do atrator, mas não de suas propriedades geométricas. Isso implica que a escolha do atraso e a dimensão de imersão têm um impacto importante na qualidade dos resultados obtidos nas aplicações. Um requisito do teorema de Takens é que a dimensão de incorporação deve satisfazer $d > 2d_{frac} + 1$, onde d_{frac} é a dimensão fractal do sistema. O primeiro passo no processo de reconstrução é a estimativa do atraso (ideal) e, uma vez atingido esse objetivo, é determinada a dimensão de imersão apropriada d . O aspecto não trivial da reconstrução do espaço de fase é refletido no grande número de técnicas sugeridas para estimativas de τ e d . Os dois métodos mais utilizados para estimar o atraso baseiam-se na determinação do primeiro mínimo da informação mútua média (FRASER; SWINNEY, 1986) ou o primeiro valor zero da função de autocorrelação (GRASSBERGER; PROCACCIA, 1983). A vantagem do primeiro é que as correlações lineares e não lineares são levadas em consideração, enquanto o segundo inclui apenas correlações lineares. Para os fins deste estudo, o primeiro mínimo local da informação mútua é usado para determinar o melhor tempo de atraso.

O próximo passo é a estimativa da dimensão de imersão. Neste caso temos duas possibilidades para a determinação do parâmetro d , aplicar o Teorema de Takens ou Método dos falsos vizinhos mais próximos (KENNEL; BROWN; ABARBANEL, 1992; TRUONG, 2017), no primeiro, calculamos a dimensão fractal do espaço (utilizando o coeficiente de hurst ou fda) e posteriormente utilizamos o resultado do teorema. No segundo, fazemos a reconstrução do atrator a cada valor $m = 2, 3, 4, \dots$ da dimensão de incorporação, é calculado a porcentagem

de pontos vizinhos em cada dimensão, para isso usa-se uma hipersfera de raio ϵ e calcula-se a distância m -dimensional entre pares de pontos. Se d, ϵ os pontos são vizinhos na dimensão m . Se na dimensão $m + 1$ os pontos deixam de ser vizinhos devido ao atrator ter se desdobrado então eles eram falsos vizinhos produzidos pela projeção do atrator numa dimensão menor que a necessária.

A seguir veremos uma aplicação de TDA em séries temporais de commodities agrícolas.

4.2 TDA em Commodities Agrícolas

Os preços de commodities agrícolas sofrem variações de preço por influência das condições climáticas, pelo biocombustível e pelo alto consumo de energia, através de fertilizantes, processamento de alimentos. A relação entre os preços do petróleo e das commodities alimentícias aumenta o custo da produção de alimentos e induz políticas relacionadas à produção de biocombustíveis. (LIMA *et al.*, 2019; BAFFES, 2013).

Muito tem se discutido sobre os preços do petróleo e os preços das commodities agrícolas, especialmente durante a última década devido a crises de alimentos em 2006/2008 e de 2010, onde os preços eram muito altos, seguidos por um rápido declínio, criando uma grande flutuação e por consequência, grande instabilidade (ABBOTT; BATTISTI, 2011).

O biocombustível é o produto que conecta os mercados de energia e alimentos. Como este é produzido a partir de culturas agrícolas existe a preocupação de como a produção e os preços dos alimentos são afetados pelo aumento da produção de biocombustíveis (KRISTOUFEK; JANDA; ZILBERMAN, 2016).

Os maiores produtores de etanol do mundo são os Estados Unidos e Brasil, que representam 56% e 28% da produção mundial, respectivamente (RFA, 2019). O etanol brasileiro é obtido pelo processamento da cana-de-açúcar, também utilizada na produção de açúcar, importante produto agrícola brasileiro (DRABIK ET AL, 2015).

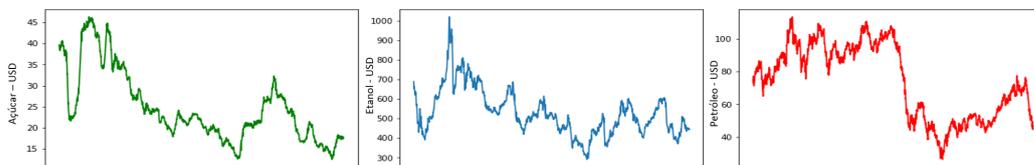
A associação entre preços de commodities agrícola e os preços de biocombustíveis tem despertado bastante interesse, estão sendo utilizadas diferentes metodologias para avaliação de

tais dados. [Lima et al. \(2019\)](#) estudaram a relação entre açúcar e etanol usando os métodos econofísicos para analisar correlações cruzadas de longo prazo entre essas commodities, ([SILVA et al., 2019](#)) utilizou o método Transfer Entropy para identificar em quais direções os fluxos de informação são mais fortes. Diante disto, utilizaremos TDA como uma medida de similaridade entre as commodities: Açúcar, Etanol e Petróleo, nas séries de Retorno e Volatilidade para investigar as possíveis influências sob preços destas commodities.

4.3 Dados

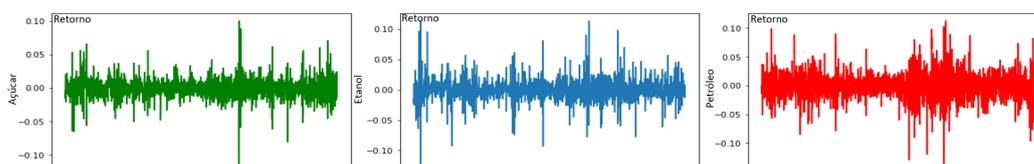
Os dados foram obtidos das seguintes fontes: Preços diários de açúcar e etanol (em USD) do Centro de Estudos Avançados em Economia Aplicada / Faculdade de Agricultura Luiz de Queiroz / Universidade de São Paulo - CEPEA / ESALQ / USP ([CEPEA/ESALQ/USP, 2019](#)) e Preços diários de petróleo bruto (em USD) da US Energy Information Administration - EIA ([EIA, 2019](#)), no período de 25/01/2010 a 31/12/2018. Os retornos logarítmicos são dados por $R_t = \ln(P_t) - \ln(P_{t-1})$ onde P_t são os preços diários observados. A volatilidade é dada por $V_t = |R_t|$. A série original, de retorno e volatilidade são mostradas nas figuras 9,10, 11.

Figura 9: Série temporal de preços P para açúcar, etanol e petróleo no período de 25/01/2010 a 31/12/2018.



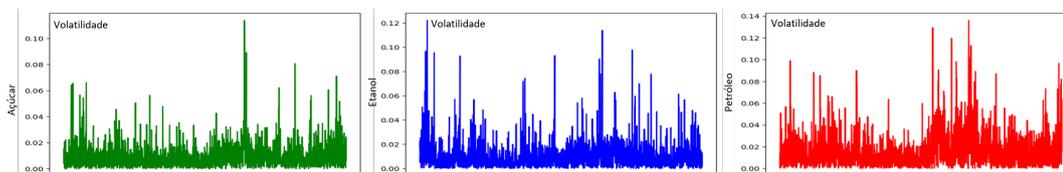
Fonte: Próprio Autor

Figura 10: Série temporal de retornos R para açúcar, etanol e petróleo no período de 25/01/2010 a 31/12/2018.



Fonte: Próprio Autor

Figura 11: Série temporal de volatilidades V para açúcar, etanol e petróleo no período de 25/01/2010 a 31/12/2018.



Fonte: Próprio Autor

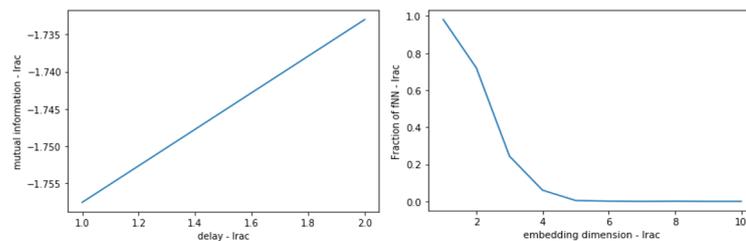
Para análise dos dados foram utilizados o software Python 3.7 e os pacotes: Ripser, Persim, TDAsets, MognuTDA, Nolds, Hurst.

CAPÍTULO 5

Resultados e Discussão

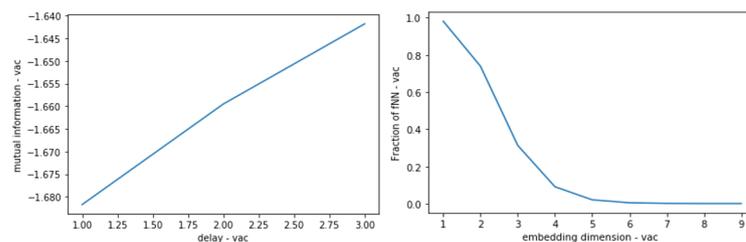
Como discutimos na seção 4.1 para reconstruirmos a série temporal através de coordenadas atrasadas, precisamos determinar os parâmetros de atraso (τ) e de dimensão de imersão (d). Para isso foram utilizados os métodos de informação mútua, para determinação do atraso, e método dos falsos vizinhos próximos para determinação da dimensão em cada série das commodities estudadas. Estes resultados podem ser vistos nas figuras 12,13, 14, 15,16 e 17.

Figura 12: Informação Mútua e Método dos Falsos Vizinhos Próximos para série de Retorno do Açúcar



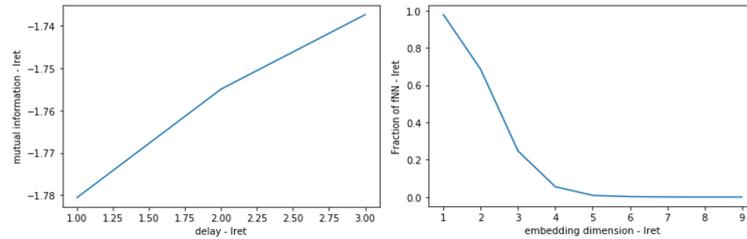
Fonte: Próprio Autor

Figura 13: Informação Mútua e Método dos Falsos Vizinhos Próximos para série de Volatilidade do Açúcar



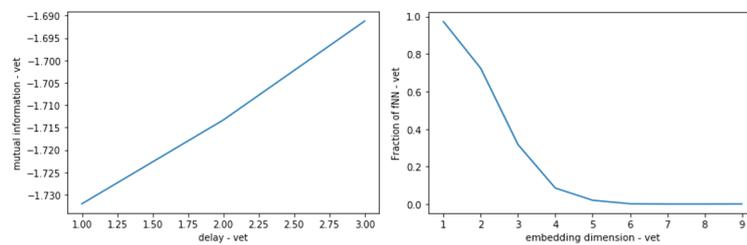
Fonte: Próprio Autor

Figura 14: Informação Mútua e Método dos Falsos Vizinhos Próximos para série de Retorno do Etanol



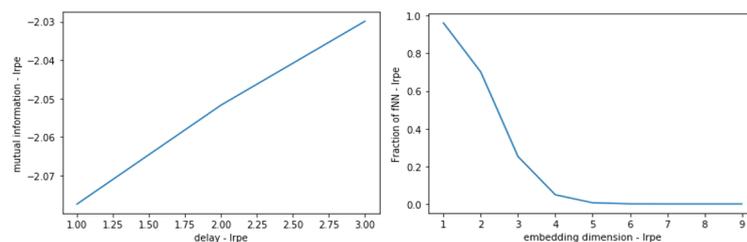
Fonte: Próprio Autor

Figura 15: Informação Mútua e Método dos Falsos Vizinhos Próximos para série de Volatilidade do Etanol



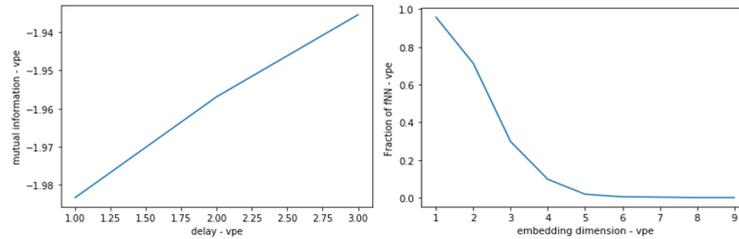
Fonte: Próprio Autor

Figura 16: Informação Mútua e Método dos Falsos Vizinhos Próximos para série de Retorno do Petróleo



Fonte: Próprio Autor

Figura 17: Informação Mútua e Método dos Falsos Vizinhos Próximos para série de Volatilidade do Petróleo

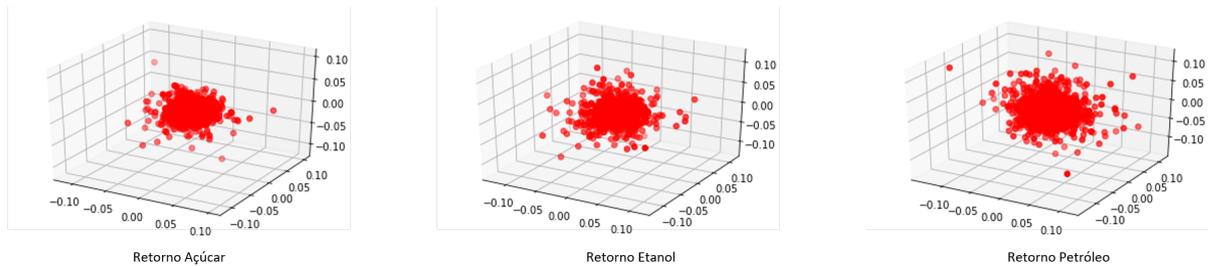


Fonte: Próprio Autor

Os resultados obtidos mostram que o atraso ideal para as séries estudadas são $\tau = 1$ e $d = 4$. Estes resultados corroboram com o sugerido por (ZBILUT, 2005).

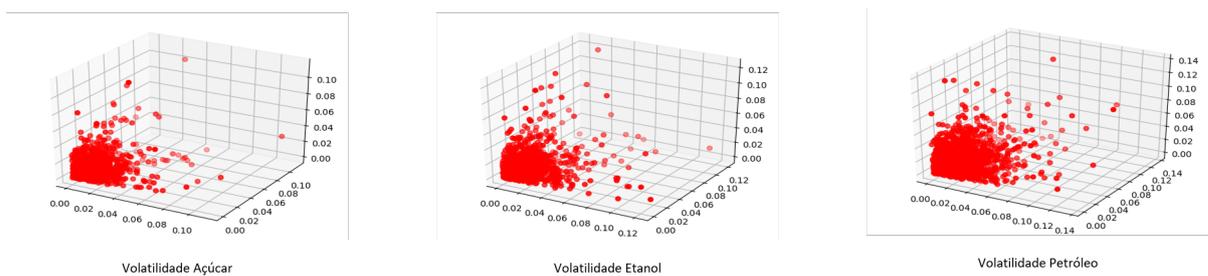
Assim determinados os parâmetros, obtemos as seguintes configurações para a reconstrução por coordenadas atrasadas vistas nas figuras 18 e 19.

Figura 18: Reconstrução por coordenadas atrasadas para as séries de retorno das commodities.



Fonte: Próprio Autor

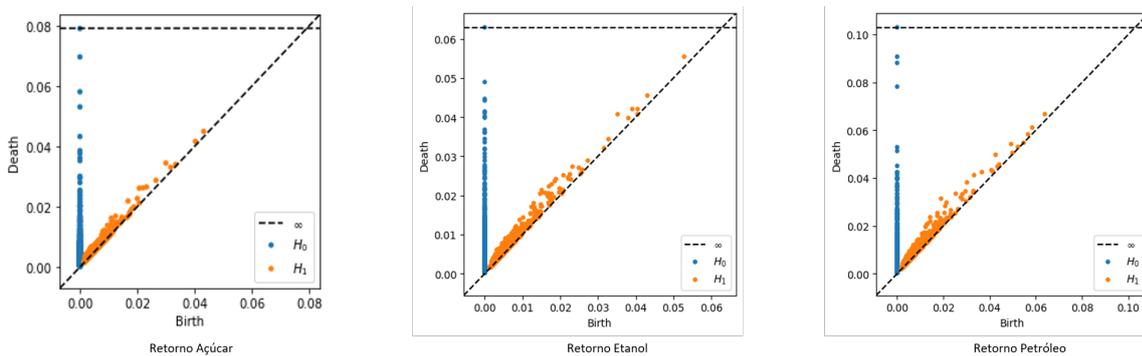
Figura 19: Reconstrução por coordenadas atrasadas para as séries de volatilidade das commodities.



Fonte: Próprio Autor

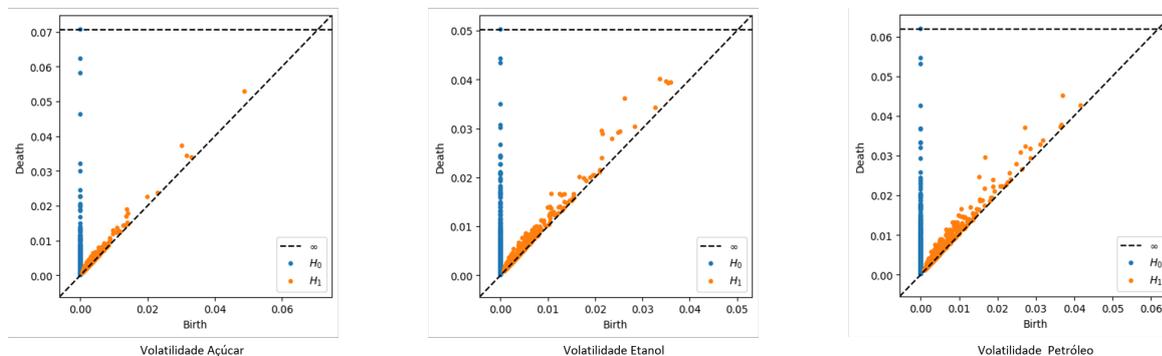
Foram obtidos os diagramas de persistência de cada série afim de usar a Distância de Bottleneck (3.3) para comparar o grau de similaridade entre os diagramas e concluir sobre as séries.

Figura 20: Diagramas de persistência para as séries de retorno das commodities.



Fonte: Próprio Autor

Figura 21: Diagramas de persistência para as séries de volatilidade das commodities.



Fonte: Próprio Autor

Os resultados dos cálculos da distância de Bottleneck (22, 23) e correlações entre os diferentes pares de commodities são mostrados nas tabelas 2 e 3.

Observa-se na tabela 2 que menores distâncias de Bottleneck são obtidas para séries de retornos de açúcar e etanol, onde há maior correlação devido à razão de mercado na produção de duas commodities produzidas pela mesma matéria-prima, a cana-de-açúcar. A maior distância do diagrama de persistência é observada nas séries de retorno para pares de petróleo/açúcar e petróleo/etanol, onde correlações menores evidenciam baixa similaridade entre as séries. No

primeiro par, a relação entre os produtos está associada ao custo do transporte, que influencia no preço final do açúcar, sendo esperado que o aumento do preço do petróleo resulte em aumento do preço do açúcar. Para o par petróleo/etanol ocorre a concorrência no mercado, ambos são combustíveis e compõem o mercado energético.

Resultados semelhantes foram obtidos para as séries de volatilidade, pois a relação de preços descrita anteriormente influenciou a variabilidade dos preços das commodities vide tabela 3.

Tabela 2: Correlação entre as Séries de Retorno versus Distância de Bottleneck.

	CORRELAÇÃO	DISTÂNCIA DE BOTTLENECK
Etanol-Açúcar	0.5601247	0.002765
Petróleo-Açúcar	0.2051491	0.004819
Petróleo-Etanol	0.1901008	0.004241

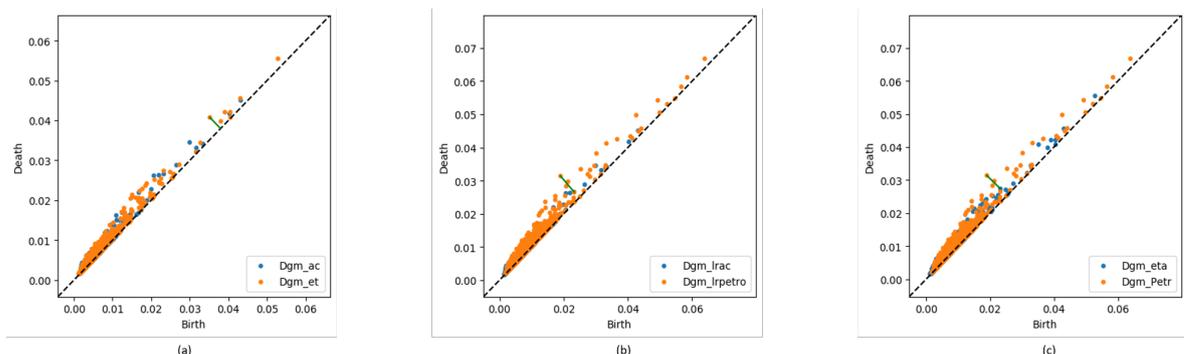
Fonte: Próprio Autor

Tabela 3: Correlação entre as Séries de Volatilidade versus Distância de Bottleneck.

	CORRELAÇÃO	DISTÂNCIA DE BOTTLENECK
Etanol-Açúcar	0.436171	0.004037
Petróleo-Açúcar	0.078786	0.006198
Petróleo-Etanol	0.073744	0.004567

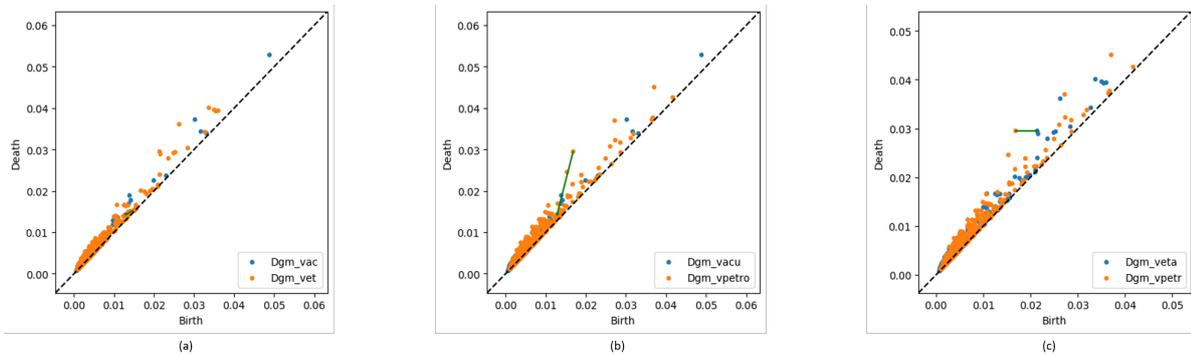
Fonte: Próprio Autor

Figura 22: Distância de Bottleneck entre os pares de commodities para as séries de retorno (a)Açúcar/Etanol, (b)Petróleo/Açúcar, (c)Petróleo/Etanol.



Fonte: Próprio Autor

Figura 23: Distância de Bottleneck entre os pares de commodities para as séries de volatilidade (a) Açúcar/Etanol, (b) Petróleo/Açúcar, (c) Petróleo/Etanol.



Fonte: Próprio Autor

Tais resultados corroboram com os resultados de (LIMA *et al.*, 2019; SILVA *et al.*, 2019) que encontraram a existência de correlações de longo prazo entre essas mercadorias, fornecendo as informações sobre qual mercadoria em cada par de mercadorias analisado contribui mais para as correlações, através das medidas de similaridades entre os diagramas de persistência.

CAPÍTULO 6

Conclusões e Perspectivas

A motivação central para este estudo foi produzir uma interpretação confiável da dinâmica subjacente dos dados de séries temporais de commodities agrícolas usando homologia persistente. Os resultados mostraram que maiores distâncias no diagrama de persistência foram obtidas nas séries de retorno e volatilidade para pares de petróleo/açúcar e petróleo/etanol, onde correlações menores evidenciam baixa similaridade entre as séries. Tais resultados são validados com o estudo de (LIMA *et al.*, 2019; SILVA *et al.*, 2019) e se justificam para o primeiro par, pela relação entre custo do transporte e a influencia que isso provoca no preço final, para o segundo par, pela concorrência no mercado, pois se tratam de combustíveis que compõem o mercado energético.

Como proposta futura pretendemos avaliar outros fenômenos agroambientais descritos por séries temporais e acrescentar uma análise que combina informações quantificadas por homologias persistente com aprendizado de máquina.

A aplicação de aprendizagem de máquina (ML) para as saídas de TDA é uma área recente e com amplas possibilidades de estudo/descoberta (UMEDA, 2017). O TDA tem sido utilizado com sucesso em diferentes tópicos, apesar de ainda não serem encontradas pesquisas suficientes sobre o TDA combinadas com técnicas de ML. Por outro lado, verifica-se que a maioria destes trabalhos de pesquisa é destinada ao ML baseado em kernels via TDA (MEJIARUDA; JIMENEZ-MORENO; HERNANDEZ, 2018).

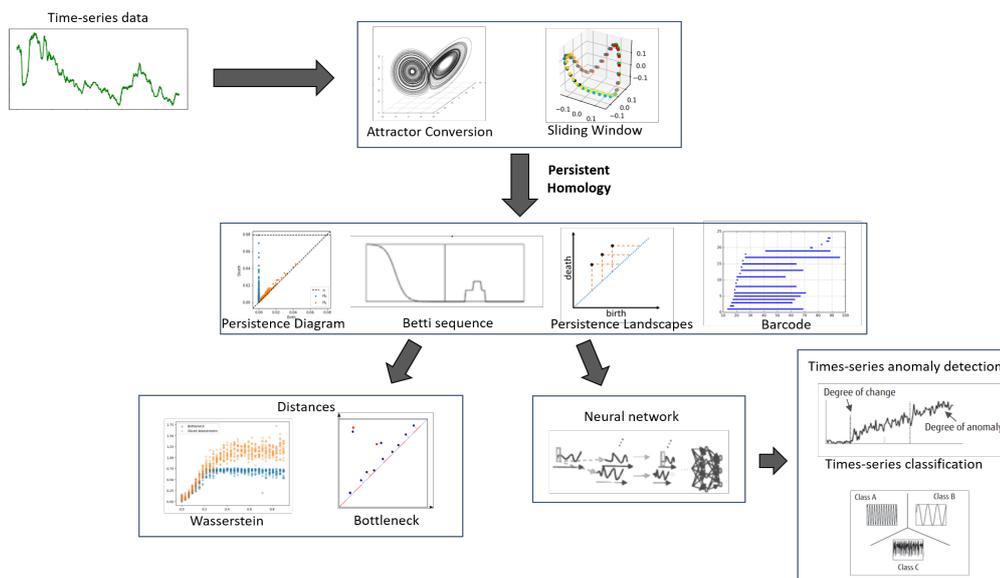
O ML (mais especificamente Deep Learning) permite que modelos computacionais compostos por várias camadas de processamento aprendam representações de dados com vários

níveis de abstração (LECUN; BENGIO; HINTON, 2015). Para séries temporais é efetivo se a classificação for clara e se houver dados supervisionados rotulados de acordo com a classe. No entanto, ao executar a análise de séries temporais na realidade, os critérios para classificação podem não ser claros; portanto, o objetivo geralmente se torna detectar quando ocorre uma alteração de um estado normal para um estado anormal ou quando ocorrem sinais dessa alteração. Por isso, utilizaremos TDA para detecção de anomalias (UMEDA, 2017).

Diante dessa perspectiva daremos continuidade a este trabalho fazendo Análise Topológica de Dados para diversos fenômenos que são descritos através de séries temporais. Estas séries devem ser transformadas em uma nuvem de pontos em um espaço com dimensão parametrizada por dois parâmetros, afim de obter uma superfície difeomorfa a variedade de onde a série foi amostrada, serão analisadas as diferentes possibilidades de saídas (diagramas de persistência, números de Betti, característica de Euler, Barcode etc..) em seguida será avaliado o impacto que as diversas formas de obtenção de uma topologia de uma nuvem de pontos têm na classificação e na complexidade computacional e por fim serão utilizadas redes neurais e distâncias para inferir sobre as saídas obtidas.

A nossa proposta de trabalhos futuros pode ser sintetizada segundo a pipeline abaixo:

Figura 24: Pipeline para TDA + Redes Neurais



Fonte: Próprio Autor

Referências Bibliográficas

ABBOTT, P.; BATTISTI, A. Borot de. Recent global food price shocks: Causes, consequences and lessons for african governments and donors. **Journal of African Economies**, Oxford University Press, v. 20, n. suppl_1, p. i12–i62, 2011.

ADAMS, H.; CARLSSON, G. Evasion paths in mobile sensor networks. **The International Journal of Robotics Research**, SAGE Publications Sage UK: London, England, v. 34, n. 1, p. 90–104, 2015.

BAFFES, J. A framework for analyzing the interplay among food, fuels, and biofuels. **Global Food Security**, Elsevier, v. 2, n. 2, p. 110–116, 2013.

BERWALD, J.; GIDEA, M.; VEJDEMO-JOHANSSON, M. Automatic recognition and tagging of topologically different regimes in dynamical systems. **arXiv preprint arXiv:1312.2482**, 2013.

BONIS, T. *et al.* Persistence-based pooling for shape pose recognition. In: SPRINGER. **International Workshop on Computational Topology in Image Context**. [S.l.], 2016. p. 19–29.

CARLSSON, G. Topology and data. **Bulletin of the American Mathematical Society**, v. 46, n. 2, p. 255–308, 2009.

CARLSSON, G. *et al.* Topological data analysis and machine learning theory. In: **BIRS Workshop, Alberta**. [S.l.: s.n.], 2012.

CARRIÈRE, M.; OUDOT, S. Y.; OVSJANIKOV, M. Stable topological signatures for points on 3d shapes. In: WILEY ONLINE LIBRARY. **Computer Graphics Forum**. [S.l.], 2015. v. 34, n. 5, p. 1–12.

CEPEA/ESALQ/USP. **Centro de Estudos Avançados em Economia Aplicada - CEPEA-Esalq/USP**. 2019. <https://www.cepea.esalq.usp.br/br>. December 05, 2019. Disponível em: [⟨{https://www.cepea.esalq.usp.br/b}⟩](https://www.cepea.esalq.usp.br/b)

CHAZAL, F. *et al.* Gromov-hausdorff stable signatures for shapes using persistence. In: WILEY ONLINE LIBRARY. **Computer Graphics Forum**. [S.l.], 2009. v. 28, n. 5, p. 1393–1403.

EDELSBRUNNER, H.; HARER, J. **Computational topology: an introduction**. [S.l.]: American Mathematical Soc., 2010.

EIA. US Energy Information Administration - EIA. 2019. <https://www.eia.gov/dnav/pet/hist/RWTCD.htm>. December 05, 2019. Disponível em: [⟨{https://www.eia.gov/dnav/pet/hist/RWTCD.htm}⟩](https://www.eia.gov/dnav/pet/hist/RWTCD.htm)

FRASER, A. M.; SWINNEY, H. L. Independent coordinates for strange attractors from mutual information. **Physical review A**, APS, v. 33, n. 2, p. 1134, 1986.

GIDEA, M.; KATZ, Y. Topological data analysis of financial time series: Landscapes of crashes. **Physica A: Statistical Mechanics and its Applications**, Elsevier, v. 491, p. 820–834, 2018.

GRASSBERGER, P.; PROCACCIA, I. Measuring the strangeness of strange attractors. **Physica D: Nonlinear Phenomena**, Elsevier, v. 9, n. 1-2, p. 189–208, 1983.

KENNEL, M. B.; BROWN, R.; ABARBANEL, H. D. Determining embedding dimension for phase-space reconstruction using a geometrical construction. **Physical review A**, APS, v. 45, n. 6, p. 3403, 1992.

KHASAWNEH, F. A.; MUNCH, E. Chatter detection in turning using persistent homology. **Mechanical Systems and Signal Processing**, Elsevier, v. 70, p. 527–541, 2016.

KOVACEV-NIKOLIC, V. *et al.* Using persistent homology and dynamical distances to analyze protein binding. **Statistical applications in genetics and molecular biology**, De Gruyter, v. 15, n. 1, p. 19–38, 2016.

KRISTOUFEK, L.; JANDA, K.; ZILBERMAN, D. Comovements of ethanol-related prices: evidence from brazil and the usa. **Gcb Bioenergy**, Wiley Online Library, v. 8, n. 2, p. 346–356, 2016.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **nature**, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015.

LIMA, C. R. A. *et al.* Cross-correlations between brazilian biofuel and food market: Ethanol versus sugar. **Physica A: Statistical Mechanics and its Applications**, Elsevier, v. 513, p. 687–693, 2019.

MALETIĆ, S.; ZHAO, Y.; RAJKOVIĆ, M. Persistent topological features of dynamical systems. **Chaos: An Interdisciplinary Journal of Nonlinear Science**, AIP Publishing, v. 26, n. 5, p. 053105, 2016.

MEJIA-RUDA, E.; JIMENEZ-MORENO, R.; HERNANDEZ, R. D. Topological data analysis for machine learning based on kernels: A survey. **International Journal of Applied Engineering Research**, v. 13, n. 17, p. 13268–13271, 2018.

MULDOON, M. *et al.* Topology from time series. **Physica D: Nonlinear Phenomena**, Elsevier, v. 65, n. 1-2, p. 1–16, 1993.

MUNKRES, J. R. **Elements of algebraic topology**. [S.l.]: CRC Press, 2018.

- PEREA, J. A.; HARER, J. Sliding windows and persistence: An application of topological methods to signal analysis. **Foundations of Computational Mathematics**, Springer, v. 15, n. 3, p. 799–838, 2015.
- PEREIRA, C. M.; MELLO, R. F. de. Persistent homology for time series and spatial data clustering. **Expert Systems with Applications**, Elsevier, v. 42, n. 15-16, p. 6026–6038, 2015.
- RFA. **Renewable Fuels Association-RFA**. 2019. <https://ethanolrfa.org/statistics/1549569130196-da23898a-53d8>. December 05, 2019. Disponível em: [⟨{https://ethanolrfa.org/statistics/1549569130196-da23898a-53d8}⟩](https://ethanolrfa.org/statistics/1549569130196-da23898a-53d8).
- SAUER, T.; YORKE, J. A.; CASDAGLI, M. Embedology. **Journal of statistical Physics**, Springer, v. 65, n. 3-4, p. 579–616, 1991.
- SILVA, J. M. da *et al.* Using transfer entropy to measure the information flow in sugar, ethanol and crude oil price series. **Sigmae**, v. 8, n. 2, p. 405–410, 2019.
- SNÁŠEL, V. *et al.* Geometrical and topological approaches to big data. **Future Generation Computer Systems**, Elsevier, v. 67, p. 286–296, 2017.
- TAKENS, F. Detecting strange attractors in turbulence. In: **Dynamical systems and turbulence, Warwick 1980**. [S.l.]: Springer, 1981. p. 366–381.
- TRUONG, P. **An exploration of topological properties of high-frequency one-dimensional financial time series data using TDA**. 2017.
- UMEDA, Y. Time series classification via topological data analysis. **Information and Media Technologies**, Information and Media Technologies Editorial Board, v. 12, p. 228–239, 2017.
- WANG, Y.; OMBAO, H.; CHUNG, M. K. Topological data analysis of single-trial electroencephalographic signals. **The annals of applied statistics**, NIH Public Access, v. 12, n. 3, p. 1506, 2018.
- WEYGAERT, R. Van de *et al.* Alpha, betti and the megaparsec universe: on the topology of the cosmic web. In: SPRINGER-VERLAG. **Transactions on computational science XIV**. [S.l.], 2011. p. 60–101.
- ZBILUT, J. P. Use of recurrence quantification analysis in economic time series. In: **Economics: Complex Windows**. [S.l.]: Springer, 2005. p. 91–104.
- ZOMORODIAN, A. J. **Topology for computing**. [S.l.]: Cambridge university press, 2005. v. 16.

APÊNDICE A

Códigos Fonte

Sliding Window:

```
import matplotlib.pyplot as plt
import numpy as np
import networkx as nx
from mpl_toolkits.mplot3d import Axes3D
import math
import pylab

def sliding_windows(M,t, series):
    M-=1
    coords = []
    for i in range(0, len(series)-1):
        if i + M*t > len(series)-1:
            return coords
        else:
            point = ()
            for j in range(0,M+1):
                point += (series[i + j*t],)
            coords.append(point)
    return coords
```

```
##plot sliding windows with matplotlib (uses scatter plot)
```

```
def plot_3d(window):
```

```
    fig = plt.figure()
```

```
    ax = fig.add_subplot(111, projection='3d')
```

```
    x = []
```

```
    y = []
```

```
    z = []
```

```
    for i in window:
```

```
        x.append(i[0])
```

```
        y.append(i[1])
```

```
        z.append(i[2])
```

```
    ax.scatter3D(x,y,z,s = 30, c = 'r')
```

```
    ax.text2D(0.05, 0.95,"", transform=ax.transAxes)
```

```
    plt.show()
```

```
##Convert window to numpy array. Return the unique values (those values were plotted).
```

```
def toNumpyArray(window):
```

```
    np_array = np.unique(np.asarray(window), axis = 0)
```

```
    return np_array
```

Takens Embedding:

```
import numpy as np # linear algebra
```

```
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import os
import math #math fun
import matplotlib.pyplot as plt #plotting
from mpl_toolkits.mplot3d import Axes3D #3d plots
from sklearn.neighbors import NearestNeighbors

def takensEmbedding (data, delay, dimension):
    "This function returns the Takens embedding of data with delay into dimension"
    if delay*dimension > len(data):
        raise NameError('Delay times dimension exceed length of data!')
    embeddedData = np.array([data[0:len(data)-delay*dimension]])
    for i in range(1, dimension):
        embeddedData = np.append(embeddedData, [data[i*delay:len(data) - delay*(d
    return embeddedData;
```

Informação Mútua (τ):

```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import os
import math #math fun
import matplotlib.pyplot as plt #plotting
from mpl_toolkits.mplot3d import Axes3D #3d plots
from sklearn.neighbors import NearestNeighbors

def mutualInformation(data, delay, nBins):
    "This function calculates the mutual information given the delay"
    I = 0;
    xmax = max(data);
```

```
xmin = min(data);
delayData = data[delay:len(data)];
shortData = data[0:len(data)-delay];
sizeBin = abs(xmax - xmin) / nBins;
#the use of dictionaries makes the process a bit faster
probInBin = {};
conditionBin = {};
conditionDelayBin = {};
for h in range(0,nBins):
    if h not in probInBin:
        conditionBin.update({h : (shortData >= (xmin + h*sizeBin)) & (shortData <
            probInBin.update({h : len(shortData[conditionBin[h]]) / len(shortData)}});
    for k in range(0,nBins):
        if k not in probInBin:
            conditionBin.update({k : (shortData >= (xmin + k*sizeBin)) & (shortData
                probInBin.update({k : len(shortData[conditionBin[k]]) / len(shortData)
            if k not in conditionDelayBin:
                conditionDelayBin.update({k : (delayData >= (xmin + k*sizeBin)) & (del
                    Phk = len(shortData[conditionBin[h] & conditionDelayBin[k]]) / len(shortDa
            if Phk != 0 and probInBin[h] != 0 and probInBin[k] != 0:
                I -= Phk * math.log( Phk / (probInBin[h] * probInBin[k]));
return I;
```

```
np_array_lrac = np.unique(np.asarray(serie_lrac), axis = 0)
datDelayInformation = []
for i in range(1,3):
    datDelayInformation = np.append(datDelayInformation,[mutualInformation(np_array_lrac
plt.plot(range(1,3),datDelayInformation);
plt.xlabel('delay - lrac');
plt.ylabel('mutual information - lrac');
```

Método dos Falsos Vizinhos Próximos (*d*):

```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import os
import math #math fun
import matplotlib.pyplot as plt #plotting
from mpl_toolkits.mplot3d import Axes3D #3d plots
from sklearn.neighbors import NearestNeighbors

def false_nearest_neighbours(data,delay,embeddingDimension):
    "Calculates the number of false nearest neighbours of embedding dimension"
    embeddedData = takensEmbedding(data,delay,embeddingDimension);
    #the first nearest neighbour is the data point itself, so we choose the second
    nbrs = NearestNeighbors(n_neighbors=2, algorithm='auto').fit(embeddedData.transpose())
    distances, indices = nbrs.kneighbors(embeddedData.transpose())
    #two data points are nearest neighbours if their distance is smaller than the
    epsilon = np.std(distances.flatten())
    nFalseNN = 0
    for i in range(0, len(data)-delay*(embeddingDimension+1)):
        if (0 < distances[i,1]) and (distances[i,1] < epsilon) and (abs(data[i+
            nFalseNN += 1;
    return nFalseNN

nFNN = []
for i in range(1,11):
    nFNN.append(false_nearest_neighbours(serie_lrac,1,i) / len(serie_lrac))
plt.plot(range(1,11),nFNN);
plt.xlabel('embedding dimension - lrac');
plt.ylabel('Fraction of fNN - lrac');
```

