

UNIVERSIDADE FEDERAL DE OURO PRETO  
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS  
DEPARTAMENTO DE COMPUTAÇÃO

VINÍCIUS COSSIO DE OLIVEIRA

**PIPELINE AUTOMATIZADO PARA GERAÇÃO E ANÁLISE DE  
GRAFOS DE CICLONES TROPICAIS A PARTIR DE DADOS ERA5**

Ouro Preto  
2026

VINÍCIUS COSSIO DE OLIVEIRA

**PIPELINE AUTOMATIZADO PARA GERAÇÃO E ANÁLISE DE GRAFOS DE  
CICLONES TROPICAIS A PARTIR DE DADOS ERA5**

Monografia apresentada ao Curso de Ciência da Computação da Universidade Federal de Ouro Preto como parte dos requisitos necessários para a obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Dr. Vander Luis de Souza Freitas

Ouro Preto,  
2026



## FOLHA DE APROVAÇÃO

**Vinícius Cossio de Oliveira**

### **Pipeline Automatizado para Geração e Análise de Grafos de Ciclones Tropicais a partir do ERA5**

Monografia apresentada ao Curso de Ciência da Computação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Ciência da Computação

Aprovada em 25 de Fevereiro de 2026

#### Membros da banca

Vander Luis de Souza Freitas (Orientador) - Doutor - Universidade Federal de Ouro Preto  
Guilherme Augusto Lopes Silva (Examinador) - Mestre - Universidade Federal de Ouro Preto  
Aurelienne Aparecida Souza Jorge (Examinadora) - Mestre - Instituto Nacional de Pesquisas Espaciais

Vander Luis de Souza Freitas, Orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 04/03/2026



Documento assinado eletronicamente por **Vander Luis de Souza Freitas, PROFESSOR DE MAGISTERIO SUPERIOR**, em 04/03/2026, às 14:01, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [http://sei.ufop.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **1062461** e o código CRC **FF668F89**.

# Resumo

Estima-se que houve mais de 300 mil vítimas fatais com a passagem do ciclone Bholá em Bangladesh e na Índia em 1970, sendo considerado o mais devastador da história. O ciclone Katrina, ocorrido nos Estados Unidos em 2005, gerou um prejuízo na casa de centenas de bilhões de dólares. A frequência da ocorrência dos ciclones e sua magnitude acendem um alerta às autoridades, que buscam formas de mitigar e antecipar tais eventos. Esta monografia apresenta a construção e aplicação de um pipeline para geração e análise de redes climáticas a partir de dados de pressão média ao nível do mar (MSLP) da reanálise ERA5, produzida pelo Centro Europeu de Previsões Meteorológicas a Médio Prazo (ECMWF). A metodologia integra a aquisição automatizada dos dados, o cálculo de anomalias, a aplicação de máscara terra-mar e a construção de matrizes de correlação utilizando o coeficiente tau de Kendall, a partir das quais são geradas redes funcionais que representam interdependências atmosféricas entre pontos da grade espacial. Sobre essas redes, são calculadas métricas topológicas como grau do nó, distância geográfica média e coeficiente de agrupamento, posteriormente submetidas à correção de efeitos de borda com base em redes aleatórias espacialmente embutidas. O pipeline permite a geração automatizada de redes e mapas das métricas para diferentes ciclones e regiões oceânicas, garantindo consistência metodológica e reprodutibilidade das análises. Como estudo de caso, foi realizada uma análise detalhada do ciclone Gaja, além da aplicação do método a outros eventos, evidenciando padrões recorrentes caracterizados por redução do grau e da distância geográfica média e aumento do coeficiente de agrupamento ao longo das trajetórias dos sistemas. Os resultados obtidos são consistentes com a literatura e demonstram o potencial do uso de redes complexas como ferramenta para a análise sistemática de ciclones tropicais e da organização espacial das interações atmosféricas associadas a eventos extremos.

**Palavras-chave:** Eventos climáticos extremos. ERA5. Pressão média ao nível do mar. Redes complexas. Redes climáticas.

# Abstract

It is estimated that there were more than 300,000 fatalities when Cyclone Bhola struck Bangladesh and India in 1970, considered the most devastating cyclone in history. Hurricane Katrina, which struck the United States in 2005, caused hundreds of billions of dollars in damage. The frequency and magnitude of cyclones have alerted authorities, who are seeking ways to mitigate and anticipate such events. This monograph presents the construction and application of a pipeline for generating and analyzing climate networks based on mean sea level pressure (MSLP) data from the ERA5 reanalysis produced by the ECMWF. The methodology integrates automated data acquisition, anomaly calculation, land-sea masking, and the construction of correlation matrices using Kendall's tau coefficient, from which functional networks representing atmospheric interdependencies between points in the spatial grid are generated. Topological metrics such as node degree, average geographic distance, and clustering coefficient are calculated on these networks and subsequently corrected for edge effects based on spatially embedded random networks. The pipeline allows for the automated generation of networks and maps of metrics for different cyclones and ocean regions, ensuring methodological consistency and reproducibility of the analyses. As a case study, a detailed analysis of Cyclone Gaja was performed, in addition to applying the method to other events, highlighting recurring patterns characterized by a reduction in the degree and average geographical distance and an increase in the clustering coefficient along the trajectories of the systems. The results obtained are consistent with the literature and demonstrate the potential of using complex networks as a tool for the systematic analysis of tropical cyclones and the spatial organization of atmospheric interactions associated with extreme events.

**Keywords:** Extreme weather events. ERA5. Mean sea level pressure. Complex networks. Climate networks.

# Lista de Figuras

Figura 2.1 – Exemplo de grafo não direcionado e não ponderado. . . . .	7
Figura 3.1 – Baía de Bengala . . . . .	14
Figura 3.2 – Estrutura da requisição automatizada utilizada na obtenção dos dados do ERA5 via API . . . . .	15
Figura 3.3 – Requisição da máscara terra-mar do ERA5 via API . . . . .	17
Figura 3.4 – Código do pipeline para processamento dos dados . . . . .	18
Figura 3.5 – Dicionário de Regiões . . . . .	19
Figura 3.6 – Dicionário de Ciclones . . . . .	20
Figura 4.1 – Comparação entre grau <b>(a)</b> e <b>(d)</b> , distância geográfica média <b>(b)</b> e <b>(e)</b> e coeficiente de agrupamento <b>(c)</b> e <b>(f)</b> antes e durante a passagem do ciclone Gaja. As figuras em uma mesma coluna apresentam a mesma escala de cores. As figuras <b>a</b> , <b>b</b> e <b>c</b> foram construídas em um período anterior à passagem do ciclone Gaja (29 de outubro de 2018 a 7 de novembro de 2018). As figuras <b>d</b> , <b>e</b> e <b>f</b> foram construídas sobre um período durante a passagem do ciclone Gaja (10 a 19 de novembro de 2018). . . . .	34
Figura 4.2 – Comparação entre grau <b>(a)</b> e <b>(d)</b> , distância geográfica média <b>(b)</b> e <b>(e)</b> e coeficiente de agrupamento <b>(c)</b> e <b>(f)</b> antes e durante a passagem do Ciclone Megh. As figuras em uma mesma coluna apresentam a mesma escala de cores. As figuras <b>a</b> , <b>b</b> e <b>c</b> foram construídas em um período anterior à passagem do Ciclone Megh (22 a 31 de outubro de 2015). As figuras <b>d</b> , <b>e</b> e <b>f</b> foram construídas sobre um período durante a passagem do Ciclone Megh (3 a 12 de novembro de 2015). . . . .	36
Figura 4.3 – Comparação entre grau <b>(a)</b> e <b>(d)</b> , distância geográfica média <b>(b)</b> e <b>(e)</b> e coeficiente de agrupamento <b>(c)</b> e <b>(f)</b> antes e durante a passagem do Ciclone Vardah. As figuras em uma mesma coluna apresentam a mesma escala de cores. As figuras <b>a</b> , <b>b</b> e <b>c</b> foram construídas em um período anterior à passagem do Ciclone Vardah (27 de novembro de 2016 a 6 de dezembro de 2016). As figuras <b>d</b> , <b>e</b> e <b>f</b> foram construídas sobre um período durante a passagem do ciclone Vardah (9 a 18 de dezembro de 2016). . . . .	36
Figura 4.4 – Comparação entre grau <b>(a)</b> e <b>(d)</b> , distância geográfica média <b>(b)</b> e <b>(e)</b> e coeficiente de agrupamento <b>(c)</b> e <b>(f)</b> antes e durante a passagem do Ciclone Luban. As figuras em uma mesma coluna apresentam a mesma escala de cores. As figuras <b>a</b> , <b>b</b> e <b>c</b> foram construídas em um período anterior à passagem do Ciclone Luban (20 a 29 de setembro de 2018). As figuras <b>d</b> , <b>e</b> e <b>f</b> foram construídas sobre um período durante a passagem do Ciclone Luban (2 a 11 de outubro de 2018). . . . .	37

Figura 4.5 – Comparação entre grau **(a)** e **(d)**, distância geográfica média **(b)** e **(e)** e coeficiente de agrupamento **(c)** e **(f)** antes e durante a passagem do Ciclone Irma. As figuras em uma mesma coluna apresentam a mesma escala de cores. As figuras **a**, **b** e **c** foram construídas em um período anterior à passagem do Ciclone Irma (20 a 29 de agosto de 2017). As figuras **d**, **e** e **f** foram construídas sobre um período durante a passagem do Ciclone Irma (1 a 10 de setembro de 2017). . . . . 37

# Lista de Tabelas

Tabela 2.1 – Exemplo de séries temporais de anomalias . . . . .	10
Tabela 2.2 – Tabela de Comparação de pares de tempo . . . . .	10



# Lista de Abreviaturas e Siglas

API	Application Programming Interface
ECMWF	Centro Europeu de Previsões Meteorológicas a Médio Prazo
MSLP	Mean Sea Level Pressure
SERN	Spatially Embedded Random Networks

# Lista de Símbolos

$\phi$	Latitude geográfica
$\lambda$	Longitude geográfica
$\tau$	Coefficiente de correlação de Kendall
$\Delta$	Diferença ou incremento

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Organização da Monografia	3
<b>2</b>	<b>Revisão Bibliográfica</b>	<b>4</b>
2.1	Trabalhos Relacionados	4
2.2	Fundamentação Teórica	6
2.2.1	Grafos e Métricas	6
2.2.2	Redes funcionais	7
2.3	Métricas	8
2.3.1	Correlação $\tau$ de Kendall	9
2.3.2	Fórmula de Haversine	11
2.3.3	Correção de efeitos de borda	11
<b>3</b>	<b>Metodologia</b>	<b>13</b>
3.1	Base de Dados	13
3.1.1	Aquisição dos dados via API	13
3.2	Visão Geral do Pipeline Computacional	16
3.2.1	Estrutura de Parametrização	18
3.2.2	Pré-processamento dos Dados	20
3.2.2.1	Cálculo da Climatologia Média	21
3.2.2.2	Cálculo das Anomalias	21
3.2.2.3	Aplicação da Máscara Terra-Mar	22
3.2.3	Construção das Séries Temporais e Janelas de Análise	22
3.2.3.1	Definição das Janelas Temporais	23
3.2.3.2	Organização Espacial dos Pontos de Grade	23
3.2.4	Construção das Redes Climáticas	23
3.2.4.1	Cálculo da Correlação de Kendall	24
3.2.4.2	Critérios de Significância Estatística	25
3.2.4.3	Construção da Matriz de Adjacência	25
3.2.5	Cálculo das Métricas de Redes Complexas	26
3.2.6	Correção de Efeitos de Borda	27
3.2.7	Geração das Visualizações	29
3.2.8	Ferramentas utilizadas	29
3.2.9	Reprodutibilidade e disponibilização pública do código	31
<b>4</b>	<b>Resultados</b>	<b>32</b>
4.1	Estudo de caso: ciclone tropical Gaja	32
4.2	Aplicação a outros ciclones tropicais	34
4.3	Validação do <i>Pipeline</i>	38

4.4	Contribuições do trabalho . . . . .	39
<b>5</b>	<b>Considerações finais . . . . .</b>	<b>40</b>
	<b>Referências . . . . .</b>	<b>42</b>

# 1 Introdução

Os ciclones tropicais estão entre os eventos climáticos extremos mais devastadores e poderosos conhecidos (KNUTSON et al., 2020). Estes eventos são sistemas atmosféricos de baixa pressão que se formam sobre águas oceânicas quentes (LIN; KLOTZBACH, 2022) em regiões específicas dos oceanos tropicais e subtropicais, onde as condições atmosféricas e oceânicas são favoráveis à sua formação. Essas áreas estão geralmente localizadas entre os trópicos de Câncer (23,5 °N) e Capricórnio (23,5 °S), mas excluem uma porção da faixa equatorial (aproximadamente entre 5 °N e 5 °S). Sua formação depende de condições específicas, como temperaturas superficiais do mar superiores a 26,5 °C, alta umidade na troposfera média e baixa estabilidade vertical. A baixa pressão central, resultante da ascensão do ar quente e úmido no núcleo do sistema, é o motor que impulsiona os ventos intensos, criando um gradiente de pressão que define a estrutura e a dinâmica do ciclone (LIN; KLOTZBACH, 2022).

O avanço acelerado nas inovações computacionais tem proporcionado soluções cada vez mais eficientes para problemas do cotidiano. A meteorologia é uma das áreas beneficiadas por esses avanços, especialmente no estudo de eventos climáticos extremos (GUPTA et al., 2021; HERSBACH et al., 2020). Nos últimos anos, diversos estudos propuseram modelos para representar esses eventos climáticos, como o trabalho de Boers et al. (2019), em que são exploradas correlações em redes para prever instabilidades climáticas em larga escala. Nesse contexto, uma rede complexa pode ser definida como um grafo que representa um sistema complexo, no qual os nós e as arestas capturam interações entre componentes interdependentes, como ocorre no caso das redes climáticas (PÓSFAI; BARABÁSI, 2016). Tsonis e Roebber (2004) descobriram que certas redes climáticas possuem propriedades de redes de pequeno mundo, que possuem alto coeficiente de agrupamento e médias baixas de comprimentos dos caminhos (WATTS; STROGATZ, 1998).

Além disso, Gupta et al. (2021) utilizaram dados de pressão média ao nível do mar (MSLP) do ERA5 (HERSBACH et al., 2020) para representar ciclones tropicais por meio de redes climáticas evolutivas, um avanço significativo tanto para a meteorologia quanto para a computação. O estudo demonstrou que métricas de redes complexas, como grau do nó, distância geográfica média e coeficiente de agrupamento, conseguem identificar assinaturas espaciais de ciclones, revelando padrões de conectividade que correspondem às suas trajetórias. Os autores escolheram o ERA5 devido a sua alta resolução espacial e temporal, permitindo capturar a variabilidade de curto prazo, característica desses eventos, além de sua capacidade de fornecer dados consistentes sobre o mar, onde os ciclones se formam.

Ao evidenciar alterações estruturais nas redes climáticas durante a ocorrência de eventos climáticos extremos, trabalhos como o de Gupta et al. (2021) demonstraram que a análise desses

eventos por meio de redes complexas é uma abordagem promissora para a compreensão da dinâmica atmosférica associada a fenômenos extremos, como ciclones tropicais.

No referido trabalho, não são disponibilizadas informações detalhadas sobre o processo de obtenção dos dados utilizados, tampouco sobre o código empregado para o processamento das séries temporais, construção das redes, cálculo das métricas ou geração dos gráficos apresentados. A ausência dessas informações e de artefatos computacionais públicos dificulta a reprodução dos experimentos, a validação independente dos resultados e a aplicação da abordagem a outras regiões ou eventos climáticos.

Nesse contexto, esta monografia busca preencher essa lacuna ao propor e implementar um *pipeline* computacional completo, automatizado e reproduzível para a obtenção, processamento e análise de dados climáticos provenientes da reanálise ERA5. Diferentemente de abordagens manuais ou estáticas, a solução desenvolvida permite a geração dinâmica dos conjuntos de dados a partir da simples parametrização da região de interesse, viabilizando a obtenção de dados de pressão média ao nível do mar com resolução temporal de três horas, ao longo de todo o período de 1979 a 2018.

Além disso, todo o processo desde a requisição dos dados via API, passando pelo cálculo de climatologia média, anomalias, aplicação de máscara terra-mar, construção das janelas temporais, cálculo das correlações e métricas de redes, até a geração das visualizações é realizado de forma automatizada e documentada. Dessa forma, a pesquisa contribui para o avanço da reprodutibilidade computacional na área de meteorologia computacional e redes climáticas, permitindo que outros pesquisadores reproduzam, validem e estendam os experimentos.

Sendo assim, o objetivo geral desta monografia é desenvolver e documentar um pipeline computacional para a geração e análise de redes climáticas associadas a eventos extremos, com ênfase em ciclones tropicais a partir de dados de MSLP da reanálise ERA5. O pipeline incorpora a obtenção dos dados e etapas de pré-processamento das séries temporais, incluindo cálculo de climatologia, anomalias e aplicação de máscara terra-mar, bem como procedimentos de correção de efeitos de borda (RHEINWALT et al., 2012) nas redes geradas, visando garantir consistência física e robustez nas métricas obtidas.

A motivação central reside no fato de que redes climáticas permitem investigar padrões estruturais associados à ocorrência de eventos extremos, contribuindo para a compreensão dos cenários atmosféricos observados durante tais eventos (BOERS et al., 2019; JORGE; COSTA; SANTOS, 2020). Além disso, os grafos produzidos podem futuramente ser utilizados em tarefas de classificação automática baseadas em Redes Neurais de Grafos (Graph Neural Networks – GNNs) (WU et al., 2020), possibilitando o desenvolvimento de ferramentas computacionais voltadas ao monitoramento e ao suporte à análise de desastres naturais, conforme discutido em estudos recentes de redes climáticas aplicadas a ciclones tropicais (GUPTA et al., 2021).

Os objetivos específicos da monografia são os seguintes:

- Reproduzir e validar a metodologia proposta por [Gupta et al. \(2021\)](#) para a construção de redes climáticas associadas à ocorrência de ciclones tropicais a partir de dados do ERA5;
- Verificar a presença de padrões estruturais nas redes climáticas que possam ser associados à ocorrência de eventos extremos;
- Demonstrar a reprodutibilidade do *pipeline* proposto, permitindo sua aplicação em diferentes regiões geográficas e eventos atmosféricos.

## 1.1 Organização da Monografia

O Capítulo 2 apresenta a revisão bibliográfica, referencial teórico e trabalhos relacionados. O Capítulo 3 apresenta os passos metodológicos para o desenvolvimento da monografia. O Capítulo 4 discute e analisa os resultados obtidos. No Capítulo 5 são apresentadas as considerações finais.

## 2 Revisão Bibliográfica

Neste Capítulo, serão apresentados os trabalhos que fundamentam a base teórica desta monografia, realizando uma análise mais aprofundada sobre os trabalhos relacionados e apresentando a fundamentação teórica.

### 2.1 Trabalhos Relacionados

Em [Tsonis e Roebber \(2004\)](#), os autores exploram a ideia de que o sistema climático pode ser representado como uma rede complexa, introduzindo o conceito de redes climáticas e demonstrando que essas redes exibem propriedades típicas de redes de pequeno mundo ([WATTS; STROGATZ, 1998](#)). Nessa abordagem, os nós representam diferentes regiões geográficas, enquanto as conexões são definidas com base na correlação entre séries temporais de anomalias climáticas. Redes de pequeno mundo são caracterizadas por combinarem alto grau de agrupamento local com a presença de conexões de longo alcance, favorecendo tanto a estabilidade quanto a propagação eficiente de informações no sistema climático.

Partindo desses conceitos, [Donges et al. \(2009\)](#) introduzem uma abordagem amplamente adotada para a construção de redes climáticas, oferecendo uma base robusta para o estudo de interdependências no sistema climático global. Nesse trabalho, cada nó representa uma posição geográfica associada a uma grade espacial de dados climáticos, enquanto as conexões são estabelecidas a partir da similaridade entre séries temporais de anomalias climáticas. Para medir essa similaridade, os autores utilizam a informação mútua, métrica capaz de capturar tanto relações lineares quanto não lineares entre as séries temporais, superando limitações de medidas tradicionais como a correlação de Pearson.

Mais recentemente, essa metodologia tem sido aplicada a eventos climáticos extremos específicos. [Gupta et al. \(2021\)](#) utilizaram redes complexas para detecção de ciclones tropicais, utilizando os dados de pressão média a nível do mar da reanálise ERA5 ([HERSBACH et al., 2020](#)) e gerando as métricas de grau do nó, distância geográfica média e coeficiente de agrupamento. A partir destas análises, os autores chegaram à conclusão de que os ciclones tropicais alteram significativamente a estrutura da rede climática, aumentando a conectividade entre diferentes pontos do oceano durante sua passagem. Em particular, o coeficiente de agrupamento mostrou-se uma métrica eficaz para identificar a localização do ciclone, pois os nós na região central do evento apresentaram os valores mais altos.

A correlação de Pearson é sensível a outliers e pressupõe linearidade, o que pode ser insuficiente para descrever a complexidade das dinâmicas climáticas ([WINTER; GOSLING; POTTER, 2016](#)). As correlações de Spearman e Kendall, por sua vez, apresentam maior robustez



por se basearem na ordenação dos dados e em relações monotônicas. Assim, a escolha da métrica de similaridade deve considerar características como a escala temporal e a duração das séries analisadas.

No contexto desta monografia, que utiliza a correlação  $\tau$  de Kendall para a detecção de ciclones tropicais em séries temporais curtas, a opção por essa métrica segue a metodologia proposta por [Gupta et al. \(2021\)](#). Essa escolha se justifica pela curta duração dos eventos analisados e pelo uso de janelas temporais de aproximadamente 10 dias, enquanto [Donges et al. \(2009\)](#) trabalham predominantemente com séries temporais mensais em escalas de décadas, cenário mais adequado à aplicação da informação mútua.

Um aspecto fundamental da metodologia de [Donges et al. \(2009\)](#) é a ênfase na significância estatística. Para evitar correlações espúrias ou influências de ruído estatístico, os autores aplicam testes para determinar se as interdependências observadas são genuínas, mantendo apenas as conexões que passam nesses critérios. Essa filtragem é essencial para garantir que a rede represente relações climaticamente relevantes.

A construção da rede climática segue, portanto, um processo sistemático: inicialmente calcula-se a medida de similaridade entre todos os pares de nós e, em seguida, define-se um limiar a partir do qual as conexões são consideradas significativas. Esse limiar não é arbitrário, sendo ajustado para selecionar apenas uma fração das interdependências mais intensas e controlar a densidade de arestas da rede. Embora [Donges et al. \(2009\)](#) adotem percentuais bastante reduzidos (como 0.5%), nesta monografia segue-se o critério de seleção dos 5% mais significativos, conforme proposto por [Gupta et al. \(2021\)](#).

Apesar desses avanços metodológicos, a construção de redes climáticas regionais apresenta desafios adicionais, especialmente relacionados aos chamados efeitos de borda. Em redes espaciais limitadas geograficamente, nós próximos às fronteiras tendem a apresentar menor conectividade não por razões físicas, mas por restrições geométricas impostas pela área de estudo. Situação semelhante ocorre nas regiões costeiras quando áreas de superfície terrestre são desconsideradas no processamento, conforme a metodologia adotada por [Gupta et al. \(2021\)](#). Nesses casos, a ausência de nós sobre o continente reduz artificialmente o número de conexões possíveis ao longo da costa, introduzindo vieses nas métricas de rede.

Para mitigar esse problema, [Rheinwalt et al. \(2012\)](#) propõem a utilização de redes aleatórias espacialmente embutidas (Spatially Embedded Random Networks – SERN), que preservam características espaciais da rede original enquanto permitem avaliar o impacto da posição geográfica dos nós sobre métricas como grau e coeficiente de agrupamento.

No contexto desta monografia, adota-se a correção de efeitos de borda seguindo a metodologia proposta por [Rheinwalt et al. \(2012\)](#). Essa etapa é fundamental para garantir que as variações observadas nas métricas de rede durante a passagem de ciclones tropicais estejam associadas a processos físicos reais, e não a limitações geométricas da área analisada.

## 2.2 Fundamentação Teórica

Esta Seção apresenta os principais conceitos teóricos e técnicas utilizadas no desenvolvimento desta monografia, fornecendo a base necessária para a compreensão da construção e análise de redes climáticas. Inicialmente, são introduzidos os conceitos fundamentais de grafos e métricas associadas, seguidos da definição de redes funcionais e das principais medidas empregadas na caracterização das redes geradas. Por fim, são apresentadas as ferramentas estatísticas e geométricas utilizadas no processamento e análise dos dados.

### 2.2.1 Grafos e Métricas

Um grafo é uma estrutura matemática usada para modelar relações entre objetos. Ele é composto por dois conjuntos principais: nós e arestas. Formalmente, um grafo  $G$  é definido como:

$$G = (V, E), \quad (2.1)$$

onde  $V$  é o conjunto de nós e  $E$  o conjunto de arestas (CHARTRAND; ZHANG, 2013), que indicam conexões ou interações entre os nós. Os grafos podem ser classificados de acordo com suas propriedades. Os grafos podem ser direcionados ou não direcionados. Em grafos não direcionados, as conexões são simétricas e, portanto:

$$(i, j) \in E \text{ implica } (j, i) \in E, \text{ sendo } i \text{ e } j \text{ nós do grafo.} \quad (2.2)$$

Já em grafos direcionados, as arestas possuem direção, representando relações assimétricas.

Os grafos podem também ser classificados como ponderados ou não ponderados. Grafos não ponderados representam apenas a presença ou ausência de conexões entre nós, enquanto grafos ponderados associam um peso a cada aresta, refletindo a intensidade da relação entre os elementos conectados. No contexto de redes climáticas, por exemplo, os grafos são frequentemente não direcionados e não ponderados, com arestas representando interdependências significativas baseadas em correlações estatísticas.

Uma forma comum de representar um grafo é por meio da matriz de adjacência  $A$ , uma matriz quadrada  $N \times N$ , onde  $N$  é o número de nós. Para um grafo não direcionado e não ponderado, ela é definida como:

$$A_{ij} = \begin{cases} 1, & \text{se existe uma aresta entre } i \text{ e } j \\ 0, & \text{se não existe uma aresta entre } i \text{ e } j. \end{cases} \quad (2.3)$$

A Figura 2.1 apresenta um exemplo de grafo não direcionado e não ponderado, acompanhado de sua respectiva matriz de adjacência.

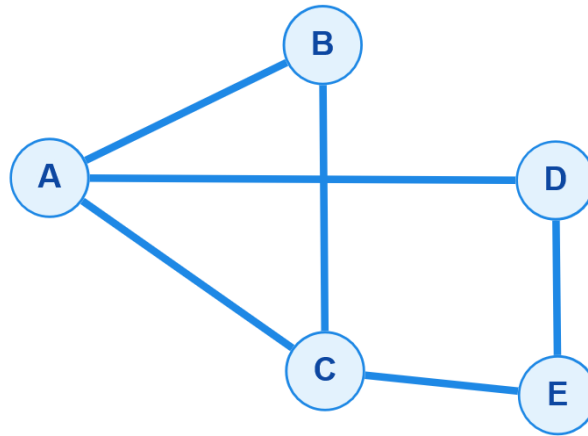


Figura 2.1 – Exemplo de grafo não direcionado e não ponderado.

$$\begin{matrix} & \begin{matrix} A & B & C & D & E \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix} & \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} \end{matrix} \quad (2.4)$$

Nesta matriz, o valor 1 indica a presença de uma aresta entre os nós correspondentes (por exemplo, A está conectado a B, C e D), enquanto 0 representa a ausência de conexão, refletindo a estrutura do grafo.

Em aplicações climáticas, os nós representam posições geográficas da grade espacial, enquanto as arestas indicam interdependências estatísticas entre as variáveis atmosféricas observadas nesses pontos.

### 2.2.2 Redes funcionais

Redes funcionais constituem um tipo específico de grafo amplamente utilizado no estudo de sistemas complexos, como o sistema climático, para modelar interdependências entre entidades a partir de relações estatísticas ou dinâmicas (BOCCALETTI et al., 2006). Diferentemente das redes estruturais, que representam conexões físicas ou espaciais diretas, as redes funcionais são construídas com base na similaridade de comportamento entre os nós, frequentemente derivada de séries temporais (NEWMAN, 2010).

Nesse tipo de abordagem, cada nó representa uma unidade do sistema, como uma posição geográfica, enquanto as arestas são estabelecidas quando há evidência estatística de interdependência entre as séries temporais associadas a esses nós. Essa estratégia permite identificar relações dinâmicas e padrões de propagação que não seriam evidentes em análises baseadas apenas na proximidade espacial (TSONIS; ROEBBER, 2004).

A utilização de redes funcionais é particularmente adequada para a análise de eventos climáticos extremos, como ciclones tropicais, pois esses fenômenos induzem padrões temporários de forte interdependência atmosférica concentrados espacialmente. Assim, variações em métricas de rede, como o grau do nó, podem revelar a organização e a propagação desses sistemas ao longo do tempo, contribuindo para a compreensão de sua dinâmica e impacto no sistema climático. Uma vez definida a estrutura da rede funcional, torna-se necessário caracterizá-la por meio de métricas topológicas.

## 2.3 Métricas

A partir da matriz de adjacência  $A_{ij}$ , obtém-se o grau de um nó  $i$ , denotado por  $k_i$ . Essa métrica indica o número de conexões de um nó na rede. Para um grafo não direcionado e não ponderado, o grau é dado por:

$$k_i = \sum_{j=1}^n A_{ij}, \quad (2.5)$$

onde  $n$  é o número total de nós da rede.

No contexto desta monografia, valores elevados de  $k_i$  indicam regiões com forte interdependência estatística entre séries temporais atmosféricas, podendo evidenciar áreas de maior relevância durante a ocorrência de eventos extremos. Por outro lado, regiões com baixo valor de  $k_i$  apresentam menor conectividade, podendo indicar áreas com menor sincronização temporal das variáveis atmosféricas ou menor influência do fenômeno analisado.

A distância geográfica média é outra métrica analisada nesta monografia. A partir da matriz de adjacência  $A_{ij}$ , calculam-se as distâncias geográficas  $\mathcal{L}_{ij}$  entre pares de nós utilizando a fórmula de Haversine, que estima a menor distância entre dois pontos na superfície terrestre assumida como esférica (ROBUSTO, 1957).

Após o cálculo de  $\mathcal{L}_{ij}$ , é possível obter a distância espacial média  $\mathcal{L}_i$  associada a cada nó  $i$ , considerando apenas os nós conectados na rede funcional:

$$\mathcal{L}_i = \frac{1}{k_i} \sum_{j=1}^n \mathcal{L}_{ij} A_{ij}, \quad (2.6)$$

onde  $j$  percorre os nós conectados a  $i$ , isto é, aqueles para os quais  $A_{ij} = 1$ .

Essa métrica permite avaliar o alcance espacial médio das conexões de cada nó na rede. Valores elevados de  $\mathcal{L}_i$  indicam interdependências de longa distância (MALIK et al., 2012), associadas a padrões atmosféricos de grande escala, enquanto valores menores sugerem interações mais localizadas. No contexto de ciclones tropicais, essa medida auxilia na identificação da extensão espacial das regiões influenciadas pelo sistema, refletindo a propagação das interdependências atmosféricas ao longo do domínio analisado.

O coeficiente de agrupamento é uma métrica que quantifica o grau de coesão local em uma rede, isto é, a tendência de formação de triângulos entre nós conectados (WATTS; STROGATZ, 1998). O coeficiente de agrupamento local  $C_i$  mede a probabilidade de que dois vizinhos de um nó  $i$  também estejam conectados entre si, refletindo o nível de interconectividade no entorno desse nó. Em um grafo não direcionado e não ponderado, o coeficiente de agrupamento local é definido por:

$$C_i = \frac{\sum_{j,h=1}^n A_{ij}A_{ih}A_{jh}}{k_i(k_i - 1)}, \quad (2.7)$$

onde  $k_i$  representa o grau do nó  $i$ . Valores elevados de  $C_i$  indicam que os vizinhos de  $i$  tendem a formar estruturas densamente conectadas, enquanto valores baixos sugerem interações mais dispersas.

O coeficiente de agrupamento global  $C$ , por sua vez, mede a densidade média de triângulos em toda a rede. Seu cálculo é dado por:

$$C = \frac{\text{número de triângulos na rede} \times 3}{\text{número total de caminhos de dois passos}}, \quad (2.8)$$

onde os caminhos de dois passos correspondem às chamadas tríades conectadas (*triplets*). Em termos da matriz de adjacência, essa expressão pode ser reescrita como:

$$C = \frac{\sum_{i,j,k} A_{ij}A_{ki}A_{jk}}{\sum_i k_i(k_i - 1)}. \quad (2.9)$$

Valores elevados de  $C$  indicam a presença de regiões com forte interdependência local entre os nós, caracterizadas por padrões atmosféricos espacialmente coerentes (GUPTA et al., 2021). Em redes climáticas, essa métrica auxilia na identificação de áreas onde as variáveis atmosféricas apresentam comportamento sincronizado em escala regional.

### 2.3.1 Correlação $\tau$ de Kendall

A correlação  $\tau$  de Kendall (KENDALL, 1938) é uma medida estatística utilizada para quantificar o grau de associação monotônica entre duas variáveis, com base na concordância entre a ordenação de seus valores ao longo do tempo. Diferentemente da correlação de Pearson, essa métrica não assume linearidade entre as variáveis e é menos sensível a outliers (GOSWAMI et al., 2017), sendo adequada para a análise de séries climáticas.

Na construção das redes funcionais, são considerados todos os pares de nós da grade espacial. Para cada par  $(i, j)$ , calcula-se a correlação  $\tau$  entre as respectivas séries temporais de anomalias. A medida é baseada na diferença entre o número de pares concordantes e discordantes ao longo da série temporal, sendo definida por:

$$\tau = \frac{C - D}{\frac{1}{2} \times n \times (n - 1)}, \quad (2.10)$$

onde  $C$  representa o número de pares concordantes,  $D$  o número de pares discordantes e  $n$  o número de observações temporais da série analisada.

Para melhor compreensão, apresenta-se na Tabela 2.1 um exemplo fictício considerando duas séries temporais de anomalias, associadas aos pontos  $P$  e  $Q$ , com cinco instantes de tempo.

Tabela 2.1 – Exemplo de séries temporais de anomalias

Tempo $t$	Anomalia em $P$	Anomalia em $Q$
$t_1$	0.2	0.1
$t_2$	0.4	0.5
$t_3$	0.1	0.3
$t_4$	0.3	0.2
$t_5$	0.5	0.6

A partir das séries da Tabela 2.1, comparam-se todos os pares de observações  $(t_i, t_j)$ , presentes na Tabela 2.2:

- Se  $P_i > P_j$  e  $Q_i > Q_j$ , ou  $P_i < P_j$  e  $Q_i < Q_j$ , o par é concordante.
- Se  $P_i > P_j$  e  $Q_i < Q_j$ , ou  $P_i < P_j$  e  $Q_i > Q_j$ , o par é discordante.

Tabela 2.2 – Tabela de Comparação de pares de tempo

Par $(t_i, t_j)$	$P_i$ vs $P_j$	$Q_i$ vs $Q_j$	tipo
$(t_1, t_2)$	$0.2 < 0.4$	$0.1 < 0.5$	C
$(t_1, t_3)$	$0.2 > 0.1$	$0.1 < 0.3$	D
$(t_1, t_4)$	$0.2 < 0.3$	$0.1 < 0.2$	C
$(t_1, t_5)$	$0.2 < 0.5$	$0.1 < 0.6$	C
$(t_2, t_3)$	$0.4 > 0.1$	$0.5 > 0.3$	C
$(t_2, t_4)$	$0.4 > 0.3$	$0.5 > 0.2$	C
$(t_2, t_5)$	$0.4 < 0.5$	$0.5 < 0.6$	C
$(t_3, t_4)$	$0.1 < 0.3$	$0.3 > 0.2$	D
$(t_3, t_5)$	$0.1 < 0.5$	$0.3 < 0.6$	C
$(t_4, t_5)$	$0.3 < 0.5$	$0.2 < 0.6$	C

Realizando as comparações entre todos os pares de instantes de tempo, obtêm-se os valores apresentados na Tabela 2.2, resultando em:

- Número de pares concordantes ( $C$ ) = 8,
- Número de pares discordantes ( $D$ ) = 2.

Aplicando a definição da correlação de Kendall:

$$\tau = \frac{8 - 2}{10} = 0.6. \quad (2.11)$$

No contexto desta monografia, esse procedimento é aplicado a todos os pares de nós da rede, gerando uma matriz de correlação  $C_{ij}$ , na qual cada elemento representa o valor de  $\tau$  entre

as séries temporais dos nós  $i$  e  $j$ . Essa matriz é simétrica, possui valores no intervalo  $[-1, 1]$  e apresenta valores iguais a 1 na diagonal principal.

Posteriormente, a matriz de correlação é utilizada como base para a construção da matriz de adjacência da rede funcional, por meio da aplicação de critérios de significância estatística e seleção dos valores mais elevados de correlação.

### 2.3.2 Fórmula de Haversine

A fórmula de Haversine, introduzida por [Robusto \(1957\)](#), é utilizada no cálculo de distâncias geográficas entre dois pontos sobre a superfície de uma esfera, como a Terra, considerando sua curvatura. No contexto desta monografia, ela é utilizada para determinar a distância geográfica média  $\mathcal{L}_i$  entre nós conectados na rede climática. A fórmula de Haversine calcula a distância  $d$  entre dois pontos definidos por suas coordenadas de latitude ( $\phi$ ) e longitude ( $\lambda$ ) da seguinte maneira:

$$a = \sin^2\left(\frac{\Delta\phi}{2}\right) + \cos(\phi_1) \times \cos(\phi_2) \times \sin^2\left(\frac{\Delta\lambda}{2}\right), \quad (2.12)$$

$$d = 2R \times \arcsin(\sqrt{a}), \quad (2.13)$$

ou então, de forma que facilite uma implementação computacional:

$$d = 2R \times \arctan 2(\sqrt{a}, \sqrt{1-a}), \quad (2.14)$$

onde  $\phi_1$  e  $\phi_2$  são as latitudes dos pontos 1 e 2, respectivamente.  $\lambda_1$  e  $\lambda_2$  são as longitudes dos pontos 1 e 2, respectivamente. Esses valores estão em radianos.  $\Delta\phi$  é a diferença entre  $\phi_1$  e  $\phi_2$ , enquanto  $\Delta\lambda$  é a diferença entre  $\lambda_1$  e  $\lambda_2$ .  $R$  é o raio da Terra, de aproximadamente  $6371 \text{ km}$  e  $\arctan 2$  é a função arco tangente de dois argumentos, que preserva o quadrante correto.

### 2.3.3 Correção de efeitos de borda

Na construção de redes climáticas a partir de domínios espaciais limitados, pode ocorrer a presença de efeitos de borda, decorrentes da delimitação geográfica da área de estudo. Esses efeitos surgem porque nós localizados nas extremidades do domínio possuem menor número potencial de conexões, uma vez que não há dados disponíveis fora da região considerada. Como consequência, métricas topológicas como grau, coeficiente de agrupamento e distância geográfica média podem ser subestimadas nessas áreas, introduzindo vieses na interpretação dos resultados.

Esse problema é particularmente relevante em redes climáticas regionais, nas quais a exclusão de regiões fora do domínio analisado, como áreas continentais desconsideradas no processamento ou porções oceânicas não incluídas no recorte espacial, altera a estrutura de conectividade da rede e pode gerar zonas artificialmente menos conectadas, especialmente ao longo das bordas do domínio e regiões costeiras. Nesses casos, a baixa conectividade pode refletir limitações geométricas e de disponibilidade de dados, e não necessariamente características físicas do sistema climático ([GUPTA et al., 2021](#)).

A presença de fronteiras artificiais pode afetar significativamente métricas topológicas, levando a interpretações equivocadas sobre a organização e a dinâmica do sistema analisado. Em particular, [Rheinwalt et al. \(2012\)](#) discute que redes espacialmente embutidas e definidas em domínios finitos estão sujeitas a vieses estruturais decorrentes da ausência de conexões além das fronteiras, o que impacta medidas como grau e coeficiente de agrupamento. Os autores propõem abordagens baseadas em redes aleatórias espacialmente embutidas para estimar e compensar tais efeitos, tornando as métricas mais representativas da dinâmica real do sistema.

No contexto desta monografia, a consideração dos efeitos de borda é fundamental para a análise das redes funcionais associadas a ciclones tropicais. A metodologia adotada segue a abordagem apresentada na literatura, incorporando a interpretação das métricas de rede à luz dessas limitações espaciais e considerando que regiões próximas às fronteiras do domínio podem apresentar conectividade reduzida não apenas por razões físicas, mas também por restrições geométricas e pela exclusão de áreas no processamento dos dados. Dessa forma, busca-se reduzir vieses e tornar a análise das interdependências atmosféricas mais consistente com a dinâmica espacial do fenômeno estudado.



## 3 Metodologia

Neste Capítulo será abordada a metodologia e os instrumentos utilizados para a obtenção dos resultados desta monografia a partir da base de dados meteorológicos ERA5.

### 3.1 Base de Dados

A base de dados utilizada neste trabalho é a reanálise atmosférica ERA5<sup>1</sup>, desenvolvida pelo ECMWF no âmbito do programa *Copernicus Climate Change Service*. O ERA5 combina observações meteorológicas provenientes de diferentes fontes com modelos numéricos de circulação atmosférica por meio de técnicas de assimilação de dados, produzindo estimativas consistentes e espacialmente contínuas de variáveis climáticas ao longo do tempo (HERSBACH et al., 2020).

A escolha do ERA5 deve-se à sua elevada resolução temporal e espacial, ampla cobertura global e consistência física, características que o tornam adequado para estudos de variabilidade atmosférica e análise de eventos extremos. Aqui, foi utilizada a variável MSLP, amplamente empregada na identificação e caracterização de sistemas atmosféricos de grande escala, incluindo ciclones tropicais.

O período analisado compreende os anos de 1979 a 2018, permitindo a construção de uma base climatológica de longo prazo e a observação de diferentes eventos ciclônicos ao longo do tempo. Embora o ERA5 disponibilize dados em resolução horária, as séries temporais foram organizadas em intervalos de três horas, em conformidade com a metodologia adotada em estudos anteriores e com o objetivo de reduzir redundâncias temporais mantendo a variabilidade relevante do sistema atmosférico.

A resolução espacial adotada foi de  $0,75^\circ \times 0,75^\circ$ , permitindo a comparação direta dos resultados e a reprodução das condições experimentais descritas no artigo de Gupta et al. (2021).

A região de estudo é a Baía de Bengala (ver Figura 3.1), localizada no Oceano Índico Norte, área frequentemente associada à formação e à atuação de ciclones tropicais. A escolha dessa região deve-se tanto à sua relevância climatológica quanto à disponibilidade de estudos prévios que possibilitam a validação dos resultados obtidos neste trabalho.

#### 3.1.1 Aquisição dos dados via API

A obtenção dos dados meteorológicos foi realizada por meio da API oficial da *Copernicus Climate Data Store*<sup>2</sup>, utilizando a biblioteca *cdsapi* na linguagem Python. A adoção dessa abordagem, em detrimento da interface gráfica baseada em formulários, está associada à necessidade

<sup>1</sup> <<https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels>>

<sup>2</sup> Os datasets da Copernicus Climate Data Store estão disponíveis em <<https://cds.climate.copernicus.eu/datasets>>

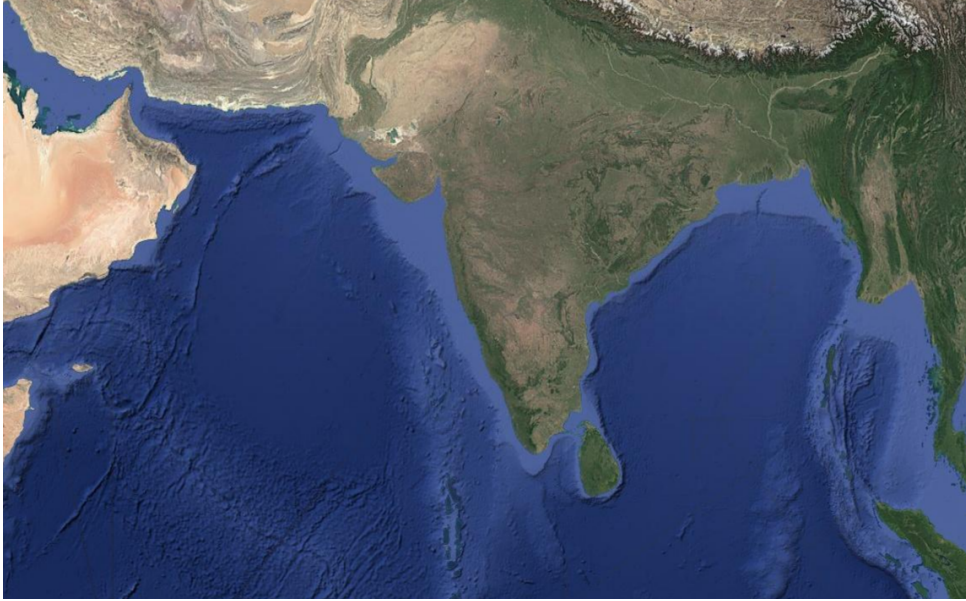


Figura 3.1 – Baía de Bengala

de automatização, reprodutibilidade e controle preciso dos parâmetros de requisição, aspectos fundamentais em estudos computacionais de larga escala e na construção de *pipelines* científicos replicáveis.

Considerando o elevado volume de dados associado à resolução temporal adotada e ao longo período analisado (1979 a 2018), a realização de requisições únicas mostrou-se inviável, resultando em falhas decorrentes de limitações operacionais do serviço. Para contornar essa restrição, as requisições foram segmentadas em intervalos temporais menores e executadas de forma automatizada e sequencial por meio de um *script*, garantindo maior estabilidade no processo de obtenção dos dados. Embora o *dataset* disponibilize informações em resolução horária, os dados foram posteriormente organizados em intervalos de três horas durante o processamento.

O processo de aquisição foi estruturado de forma parametrizável, permitindo que a região de interesse seja definida diretamente no código por meio da seleção de delimitadores geográficos previamente armazenados. Essa estratégia possibilita a obtenção automática dos dados de pressão média ao nível do mar para diferentes regiões e períodos temporais, sem a necessidade de modificações estruturais no *pipeline*. Como resultado, reduz-se a intervenção manual, minimiza-se o risco de erros operacionais e favorece-se a reprodução do experimento por outros pesquisadores.

As Figuras 3.2 e 3.3 ilustram a estrutura dos códigos responsáveis pela requisição dos dados, evidenciando a parametrização do processo e a definição explícita das variáveis, do período temporal e da região geográfica analisada. Devido ao elevado volume de dados necessário para a reprodução dos experimentos e às limitações do sistema quanto ao tamanho máximo permitido por requisição, o processo de download do MSLP foi segmentado por dia, por meio de um laço de repetição que percorre os dias possíveis de cada mês ao longo do período de estudo.

```

1  import cdsapi
2  from dictionary_regions import REGIONS
3
4  REGION = "North_Atlantic_Ocean"
5
6  if REGION not in REGIONS:
7      print(f"Região '{REGION}' não encontrada no dicionário de regiões.")
8      print("Verifique a região selecionada e tente novamente.")
9      exit()
10
11  region = REGIONS[REGION]
12
13  client = cdsapi.Client(
14      url="https://cds.climate.copernicus.eu/api",
15      key="*****",
16      verify=True,
17  )
18
19  dataset = "reanalysis-era5-single-levels"
20  days = [str(d).zfill(2) for d in range(1, 32)]
21
22  for day in days:
23      request = {
24          "product_type": ["reanalysis"],
25          "variable": ["mean_sea_level_pressure"],
26          "year": [str(y) for y in range(1979, 2019)],
27          "month": [str(m) for m in range(1, 13) if not (m == 2 and day == "29")],
28          "day": [day],
29          "time": [
30              "00:00", "03:00", "06:00", "09:00",
31              "12:00", "15:00", "18:00", "21:00"
32          ],
33          "data_format": "netcdf_legacy",
34          "download_format": "unarchived",
35          "area": region["area"],
36          "grid": "0.75/0.75"
37      }
38
39      output = f"Dataset/{region["folder"]}/mslp/mslp_dia_{day}.nc"
40      client.retrieve(dataset, request, output)

```

Figura 3.2 – Estrutura da requisição automatizada utilizada na obtenção dos dados do ERA5 via API

Foi considerada a possibilidade de paralelização das requisições com o objetivo de reduzir o tempo total de obtenção dos dados. No entanto, o serviço da *Copernicus Climate Data Store* opera com base em um mecanismo de fila, no qual apenas uma requisição por usuário é processada por vez. Dessa forma, a estratégia baseada em requisições sequenciais mostrou-se adequada e consistente para o contexto.

Como resultado desse procedimento, foram gerados arquivos distintos correspondentes a cada dia do mês ao longo do período de 1979 a 2018, contendo dados com resolução temporal de três horas. O dia 29 de fevereiro não foi considerado na análise por ocorrer apenas em anos bissextos, sendo essa decisão incorporada explicitamente à lógica de extração, de modo a garantir consistência temporal ao longo de todo o período analisado.

Os principais parâmetros definidos nas requisições estão sintetizados a seguir:

- **Product type:** tipo de produto solicitado ao ERA5, definido como *reanalysis*, correspon-

dente a dados obtidos por assimilação de observações e modelagem numérica;

- **Variable:** variável meteorológica selecionada, correspondente à pressão média ao nível do mar (*mean sea level pressure* - MSLP), utilizada na construção das redes climáticas;
- **Year:** conjunto de anos considerado na requisição, abrangendo o período de 1979 a 2018;
- **Month:** meses do ano incluídos na análise;
- **Day:** segmentação das requisições por dia do mês ao longo do período analisado;
- **Time:** horários selecionados ao longo do dia, com intervalo temporal de três horas;
- **Geographical area:** delimitação espacial da região de interesse por meio de coordenadas geográficas (*bounding box*);
- **Grid:** resolução espacial definida como  $0,75^\circ \times 0,75^\circ$ , em conformidade com a literatura;
- **Data format:** formato de saída dos dados, definido como *NetCDF*.

Para a análise de ciclones tropicais, é fundamental considerar prioritariamente informações associadas à superfície oceânica, uma vez que a dinâmica desses sistemas está diretamente relacionada às condições atmosféricas sobre o mar. Na reanálise ERA5, valores de pressão média ao nível do mar sobre áreas continentais são obtidos por extrapolação a partir da pressão de superfície, podendo introduzir inconsistências ao serem comparados com valores observados sobre o oceano.

Dessa forma, os pontos correspondentes à superfície terrestre foram desconsiderados por meio da aplicação de uma máscara terra-mar (*land-sea mask*), obtida a partir do próprio ERA5 via API, garantindo compatibilidade espacial e consistência metodológica com os dados de pressão utilizados.

A extração da máscara foi igualmente realizada de forma automatizada e parametrizada, considerando a mesma delimitação geográfica adotada para a região de estudo, como ilustrado na Figura 3.3. A partir dessa máscara, os valores de pressão associados à superfície continental foram desconsiderados nas etapas subsequentes do processamento, assegurando que a análise das redes climáticas permanecesse restrita às áreas oceânicas relevantes para o estudo dos ciclones tropicais.

## 3.2 Visão Geral do Pipeline Computacional

O *pipeline* computacional desenvolvido foi estruturado de forma modular, permitindo que todas as etapas do processamento e da análise sejam executadas a partir de um único ponto de entrada. A Figura 3.4 apresenta o trecho principal do código responsável pela orquestração das etapas do pipeline.

```

1  import cdsapi
2  from dictionary_regions import REGIONS
3
4  REGION = "BENGAL"
5  region = REGIONS[REGION]
6
7  client = cdsapi.Client(
8      url="https://cds.climate.copernicus.eu/api",
9      key="*****",
10     verify=True,
11 )
12
13 dataset = "reanalysis-era5-single-levels"
14 request = {
15     "product_type": ["reanalysis"],
16     "variable": ["land_sea_mask"],
17     "year": ["2018"],
18     "month": ["11"],
19     "day": ["10"],
20     "time": ["00:00"],
21     "data_format": "netcdf_legacy",
22     "download_format": "unarchived",
23     "area": region["area"],
24     "grid": "0.75/0.75"
25 }
26
27 output = f"Dataset/{region["folder"]}/land_sea/land_sea_mask.nc"
28 client.retrieve(
29     dataset,
30     request,
31     output
32 )

```

Figura 3.3 – Requisição da máscara terra-mar do ERA5 via API

A execução do pipeline é controlada por duas variáveis principais: a região geográfica de interesse e o ciclone selecionado. A partir dessas variáveis, todas as etapas subsequentes são realizadas automaticamente, sem a necessidade de intervenções manuais adicionais. As etapas do pipeline são resumidas a seguir:

1. **Aquisição dos dados do ERA5 via API:** Nessa etapa, o usuário deve selecionar os parâmetros da região desejada, permitindo a obtenção automatizada dos dados de pressão média ao nível do mar.
2. **Pré-processamento dos dados:** Nessa etapa é realizado o cálculo da climatologia média, das anomalias e a aplicação da máscara terra-mar.
3. **Construção das redes:** Criação das janelas temporais, cálculo da correlação de Kendall, aplicação do filtro de significância e seleção dos 5% maiores valores para construção da matriz de adjacência.

4. **Cálculo das métricas:** Cálculo do grau do nó, distância geográfica média e coeficiente de agrupamento.
5. **Correção de efeitos de borda:** Geração de 1000 SERN, cálculo das métricas esperadas e aplicação da correção.
6. **Geração das visualizações:** Geração dos mapas de grau do nó, distância geográfica média e coeficiente de agrupamento.

```

import calculate_mean_climatology as cli
import calculate_anomaly as ano
import apply_land_sea_mask as lsm
import create_sliding_windows as sw
import calculate_kendall as kendall
import calculate_degree as degree
import calculate_mean_geographical_distance as mean_dist
import calculate_clustering_coefficient as clust
import boundary_effects_correction as corr
import plot as plt

from dictionary import CYCLONES
from dictionary_regions import REGIONS

CYCLONE = "Irma"

def main():
    if CYCLONE not in CYCLONES: ...

    REGION = CYCLONES[CYCLONE]["region"]

    if REGION not in REGIONS: ...

    cli.calculate_mean_climatology(REGION)
    ano.calculate_anomaly(REGION, CYCLONE)
    lsm.apply_land_sea_mask(REGION, CYCLONE)
    sw.create_sliding_windows(REGION, CYCLONE)
    kendall.calculate_kendall(REGION, CYCLONE)
    degree.calculate_degree(REGION, CYCLONE)
    mean_dist.calculate_mean_distance(REGION, CYCLONE)
    clust.calculate_clustering(REGION, CYCLONE)
    corr.boundary_correction(REGION, CYCLONE)
    plt.plot(REGION, CYCLONE)

if __name__ == "__main__":
    main()

```

Figura 3.4 – Código do pipeline para processamento dos dados

### 3.2.1 Estrutura de Parametrização

Com o objetivo de tornar o pipeline computacional flexível, reutilizável e reproduzível, foi adotada uma estrutura de parametrização baseada em dicionários. Essa abordagem permite que diferentes regiões geográficas e eventos de ciclone sejam analisados a partir da simples modificação de variáveis de controle, sem a necessidade de alterações estruturais no código.

```

1  REGIONS = {
2      "Bengal_Bay": {
3          "area": [34.5, 49.5, 4.5, 100],
4          "folder": "Bengal_Bay",
5      },
6      "Gulf_of_Mexico": {
7          "area": [31, 262, 18, 280],
8          "folder": "Gulf_of_Mexico",
9      },
10     "North_Atlantic_Ocean": {
11         "area": [42, -100, 10, -42],
12         "folder": "North_Atlantic_Ocean",
13     }
14 }

```

Figura 3.5 – Dicionário de Regiões

A parametrização centraliza as informações relevantes para cada experimento, reduzindo o risco de erros operacionais e facilitando a extensão do pipeline para novos cenários de estudo. As subseções a seguir descrevem os dicionários utilizados para a definição das regiões de interesse e dos ciclones analisados.

As regiões geográficas analisadas são definidas por meio de um dicionário, no qual cada entrada associa um identificador textual a um conjunto de parâmetros espaciais. Esses parâmetros incluem a delimitação geográfica da região, expressa por meio de coordenadas que definem uma caixa delimitadora (*bounding box*), bem como informações auxiliares relacionadas à organização dos arquivos gerados.

Essa estrutura permite que a seleção da região de estudo seja realizada a partir da simples atribuição de uma variável no código principal do pipeline. A partir dessa definição, todas as etapas subsequentes, incluindo a extração dos dados via API, o pré-processamento e a construção das redes climáticas, utilizam automaticamente os limites espaciais correspondentes, garantindo consistência ao longo de todo o fluxo de processamento.

A adoção de um dicionário de regiões viabiliza ainda a reutilização do pipeline para diferentes áreas geográficas, possibilitando análises comparativas ou a extensão do estudo para outras regiões de interesse sem a necessidade de modificações adicionais no código.

De forma análoga, os eventos de ciclone considerados na análise são definidos por meio de um dicionário específico. Cada ciclone é associado a informações temporais relevantes, incluindo obrigatoriamente o intervalo correspondente ao período de atuação do sistema (*durante*) e, quando disponível, um intervalo adicional referente ao período anterior ao evento (*antes*). Esses intervalos são utilizados nas etapas de cálculo das anomalias e na construção das janelas temporais empregadas na análise das redes climáticas.

A definição do período anterior ao ciclone não é obrigatória no *pipeline*. Quando essa informação não está presente, as métricas de rede são calculadas e as visualizações são geradas apenas para o intervalo correspondente à atuação do sistema atmosférico. Essa flexibilidade



```

1  CYCLONES = {
2      "Gaja": {
3          "antes": ("2018-10-29 00:00", "2018-11-07 21:00"),
4          "durante": ("2018-11-10 00:00", "2018-11-19 21:00"),
5          "folder": "Gaja",
6          "region": "Bengal_Bay"
7      },
8      "Luban": {
9          "antes": ("2018-09-02 00:00", "2018-09-11 21:00"),
10         "durante": ("2018-10-02 00:00", "2018-10-11 21:00"),
11         "folder": "Luban",
12         "region": "Bengal_Bay"
13     },
14     "Vardah": {
15         "antes": ("2016-11-26 00:00", "2016-12-05 21:00"),
16         "durante": ("2016-12-09 00:00", "2016-12-18 21:00"),
17         "folder": "Vardah",
18         "region": "Bengal_Bay"
19     },
20     "Megh": {
21         "durante": ("2015-11-03 00:00", "2015-11-12 21:00"),
22         "folder": "Megh",
23         "region": "Bengal_Bay"
24     },
25     "Irma": {
26         "durante": ("2017-09-01 00:00", "2017-09-10 21:00"),
27         "folder": "Irma",
28         "region": "North_Atlantic_Ocean"
29     }
30 }
31 }

```

Figura 3.6 – Dicionário de Ciclones

permite que eventos com diferentes níveis de informação temporal sejam analisados sem a necessidade de adaptações estruturais no código.

Ademais, diferentes ciclones são processados a partir da mesma infraestrutura computacional, bastando alterar a variável associada no código principal. O uso de um dicionário centralizado assegura que todas as informações temporais e espaciais associadas ao ciclone sejam utilizadas de forma consistente ao longo de todo o *pipeline*.

A combinação dos dicionários de regiões e de ciclones constitui, portanto, um mecanismo de parametrização que separa claramente os aspectos experimentais das implementações computacionais, reforçando a modularidade, a clareza e a reprodutibilidade do *pipeline* desenvolvido, além de permitir sua aplicação tanto a estudos comparativos (antes vs. durante) quanto a análises focadas exclusivamente na atuação dos ciclones.

### 3.2.2 Pré-processamento dos Dados

O pré-processamento dos dados constitui uma etapa fundamental do *pipeline*, sendo responsável por preparar as séries temporais de MSLP para a construção das redes climáticas. Essa etapa tem como objetivo remover padrões sazonais de longo prazo, destacar variações relevantes associadas aos eventos analisados e garantir a coerência física dos dados utilizados nas análises subsequentes.

A etapa de pré-processamento inclui o cálculo da climatologia média, a obtenção das anomalias de MSLP e a aplicação da máscara terra-mar, conforme detalhado a seguir.



### 3.2.2.1 Cálculo da Climatologia Média

O cálculo da climatologia média constitui a primeira etapa do pré-processamento dos dados e tem como objetivo caracterizar o comportamento médio esperado da MSLP para cada ponto da grade espacial, considerando o período de referência adotado.

A climatologia foi calculada individualmente para cada ponto de grade da região de estudo a partir de toda a série temporal disponível no período de 1979 a 2018 (GUPTA et al., 2021), com resolução temporal de três horas. O procedimento adotado baseia-se na construção de uma climatologia diária do calendário, na qual os dados são agrupados de acordo com o dia do ano.

Para cada ponto de grade e para cada dia do calendário (1 a 365), foram considerados todos os valores de MSLP correspondentes àquele dia ao longo dos anos analisados, considerando as oito observações diárias associadas à resolução temporal de três horas. Em seguida, foi calculada a média desses valores, resultando em um campo médio representativo para cada dia do ano.

Como resultado, obtém-se um conjunto de 365 campos espaciais de MSLP, cada um representando o estado climatológico médio esperado da atmosfera para um dia específico do ano. Essa abordagem permite capturar o ciclo anual da pressão atmosférica e fornece uma referência adequada para o cálculo das anomalias, descrito na subseção seguinte.

Cabe destacar que o cálculo da climatologia média é realizado antes da aplicação da máscara terra-mar, uma vez que a climatologia representa o comportamento médio do campo de pressão independentemente da classificação oceânica ou continental.

### 3.2.2.2 Cálculo das Anomalias

A partir da climatologia média, são calculadas as anomalias de pressão média ao nível do mar. As anomalias representam o desvio da variável em relação ao seu comportamento médio esperado, permitindo destacar variações e padrões atmosféricos associados à ocorrência de ciclones tropicais.

Para cada ponto da grade espacial e para cada instante de tempo, a anomalia de MSLP foi obtida pela diferença entre o valor observado e a climatologia média correspondente. Esse procedimento remove o componente sazonal dominante da série temporal, permitindo que as flutuações associadas a processos atmosféricos de menor escala temporal sejam analisadas de forma mais adequada.

As séries temporais de anomalias resultantes constituem a base para a aplicação das janelas temporais de análise e para o cálculo das medidas de similaridade estatística entre os nós da rede, etapas que serão descritas nas seções subsequentes.

### 3.2.2.3 Aplicação da Máscara Terra-Mar

Através da requisição via *API*, os dados de pressão média ao nível do mar do ERA5 são obtidos diretamente na resolução espacial desejada de  $0,75^\circ \times 0,75^\circ$ . No entanto, ainda é necessário realizar ajustes adicionais para garantir a coerência física da análise, especialmente no que se refere à distinção entre superfícies oceânicas e continentais.

A máscara terra-mar disponibilizada pelo ERA5 fornece valores contínuos no intervalo  $[0, 1]$ , nos quais valores próximos de 0 indicam superfície oceânica e valores próximos de 1 indicam superfície terrestre. Devido à resolução espacial da grade e à presença de regiões costeiras ou parcialmente continentais, esses valores não são estritamente binários, podendo introduzir ambiguidades na identificação dos pontos oceânicos.

Para mitigar esse problema, a máscara terra-mar foi binarizada adotando-se um limiar de 0.5, de modo que pontos com valores inferiores a esse limiar foram classificados como oceânicos, enquanto pontos com valores iguais ou superiores foram classificados como continentais. Essa abordagem é amplamente adotada na literatura e permite uma separação clara entre as duas superfícies.

A aplicação da máscara foi realizada após o cálculo das anomalias de MSLP. Inicialmente, a máscara terra-mar foi interpolada para a mesma grade espacial das séries temporais de anomalias, garantindo compatibilidade entre os conjuntos de dados. Em seguida, os valores de anomalia associados a pontos continentais foram tornados nulos (*NaN*), assegurando que apenas regiões oceânicas fossem consideradas nas etapas subsequentes da construção das redes climáticas.

Esse procedimento evita a introdução de ruídos artificiais decorrentes da extrapolação da pressão ao nível do mar sobre o continente, além de reduzir o número efetivo de nós da rede, contribuindo para maior eficiência computacional e maior coerência física na análise dos ciclones tropicais.

### 3.2.3 Construção das Séries Temporais e Janelas de Análise

Após o pré-processamento dos dados, foram construídas as séries temporais de anomalias de MSLP para cada ponto da grade espacial. Cada ponto oceânico válido, definido pela aplicação da máscara terra-mar, é tratado como um nó da rede climática, associado a uma série temporal unidimensional de anomalias de MSLP.

A partir dessas séries temporais completas, foram definidas janelas temporais de análise com o objetivo de capturar a evolução dinâmica dos padrões atmosféricos associados aos ciclones tropicais, mantendo um equilíbrio entre resolução temporal e estabilidade estatística das medidas de correlação.

### 3.2.3.1 Definição das Janelas Temporais

As janelas temporais foram definidas com base na metodologia proposta por [Gupta et al. \(2021\)](#), considerando um intervalo total de 10 dias centrado no período de interesse do ciclone. Esse intervalo é suficientemente longo para capturar a organização espacial da pressão ao nível do mar associada ao sistema, sem diluir excessivamente os sinais dinâmicos relevantes.

Os dados de MSLP possuem resolução temporal original de uma hora, porém, conforme adotado neste trabalho, foram organizados em intervalos de três horas durante o processamento. Dessa forma, cada janela temporal contém 80 instantes no tempo, correspondentes a oito observações diárias ao longo de dez dias consecutivos. Para cada ponto da grade espacial, a série temporal de anomalias dentro da janela definida é utilizada como entrada para o cálculo das medidas de similaridade estatística entre os nós da rede. Essa abordagem garante consistência temporal entre todas as séries analisadas e viabiliza a construção das matrizes de correlação necessárias para a definição das conexões da rede climática.

### 3.2.3.2 Organização Espacial dos Pontos de Grade

A organização espacial dos pontos de grade é uma etapa fundamental para a construção das redes climáticas, pois define a correspondência entre a estrutura espacial do campo de pressão e os nós da rede. Após a aplicação da máscara terra-mar, cada ponto oceânico válido da grade espacial, definido por uma combinação de latitude e longitude, é associado a um nó da rede climática.

Considerando uma grade espacial regular, o número total de nós da rede corresponde ao número de pontos oceânicos remanescentes após a exclusão dos pontos continentais. Essa abordagem reduz a dimensionalidade do problema e assegura que apenas regiões fisicamente relevantes para a análise de ciclones tropicais sejam consideradas.

Para cada nó, é associada uma série temporal de anomalias de MSLP correspondente ao intervalo temporal definido pela janela de análise. A indexação dos nós é realizada de forma consistente ao longo de todo o pipeline, garantindo que a posição geográfica de cada nó possa ser recuperada posteriormente para o cálculo de métricas espaciais, como a distância geográfica média, e para a geração das visualizações.

Essa correspondência direta entre os pontos da grade espacial e os nós da rede permite que a estrutura topológica das redes climáticas preserve a organização geográfica subjacente do sistema atmosférico, facilitando a interpretação física dos padrões identificados nas etapas subsequentes da análise.

## 3.2.4 Construção das Redes Climáticas

Nesta etapa, os dados climáticos pré-processados são utilizados para a construção das redes climáticas associadas aos eventos de ciclones tropicais analisados. A rede climática é

representada como um grafo não direcionado, no qual cada nó corresponde a um ponto da grade espacial (latitude, longitude) localizado sobre a superfície oceânica, e as arestas representam relações estatísticas significativas entre as séries temporais de anomalias de MSLP desses pontos.

A construção da rede baseia-se na análise de similaridade entre as séries temporais associadas aos nós, calculadas dentro de janelas temporais específicas. A partir dessas medidas de similaridade, são aplicados critérios de significância estatística e limiares de seleção, de modo a definir quais pares de nós estarão efetivamente conectados na rede.

O processo de construção das redes climáticas é composto pelas seguintes etapas principais: (i) cálculo da medida de similaridade estatística entre os pares de nós; (ii) aplicação de critérios de significância estatística para filtragem das relações espúrias; e (iii) definição da matriz de adjacência que representa a topologia final da rede. Cada uma dessas etapas é detalhada nas subseções a seguir.

### 3.2.4.1 Cálculo da Correlação de Kendall

A construção das redes climáticas baseia-se na quantificação da similaridade estatística entre as séries temporais de anomalias de MSLP associadas aos diferentes nós da rede. Para esse fim, foi adotada a correlação  $\tau$  de Kendall (KENDALL, 1938), uma medida não paramétrica que avalia o grau de concordância entre duas séries temporais a partir da ordenação de seus valores. Formalmente, o coeficiente  $\tau$  é definido por

$$\tau = \frac{C - D}{\frac{1}{2} \times n \times (n - 1)},$$

onde  $C$  é o número de pares concordantes,  $D$  o número de pares discordantes e  $n$  o número de observações temporais na série (detalhes da derivação encontram-se na Subseção 2.3.1).

A escolha da correlação de Kendall é particularmente adequada ao contexto deste trabalho, uma vez que as séries temporais analisadas possuem curta duração, resultantes da aplicação de janelas temporais de aproximadamente 10 dias, com resolução temporal de três horas. Nesse cenário, medidas clássicas como a correlação de Pearson podem apresentar desempenho inferior, por assumirem linearidade e maior sensibilidade a valores extremos. A correlação de Kendall, por outro lado, apresenta maior robustez frente a não linearidades e ruídos estatísticos (GOSWAMI et al., 2017).

Para cada par de nós  $(i, j)$  da rede, a correlação  $\tau_{ij}$  é calculada a partir das respectivas séries temporais de anomalias de MSLP dentro de uma mesma janela temporal. Esse procedimento resulta em uma matriz de correlação simétrica, na qual cada elemento representa o grau de associação estatística entre dois pontos da grade espacial.

Os valores de correlação obtidos constituem a base para as etapas subsequentes de filtragem estatística e definição das conexões da rede.

### 3.2.4.2 Critérios de Significância Estatística

Após o cálculo da correlação  $\tau$  de Kendall entre todas as séries temporais de anomalias associadas aos pares de nós da rede, foi aplicada uma etapa de filtragem baseada em significância estatística. O objetivo dessa etapa é reduzir a influência de correlações espúrias decorrentes de ruído estatístico, garantindo que apenas interdependências estatisticamente relevantes sejam consideradas na construção da rede climática.

Para cada valor de correlação calculado, foi avaliada sua significância estatística adotando-se um nível de confiança de 95% ( $p$ -valor inferior a 0.05). Correlações que não atingiram esse critério foram consideradas estatisticamente insignificantes e tiveram seus valores descartados, sendo tratados como ausência de conexão entre os respectivos nós.

Esse procedimento é fundamental em redes climáticas, uma vez que o grande número de pares de nós pode levar à identificação de correlações fracas ou aleatórias que não refletem interações físicas reais no sistema climático. A eliminação dessas correlações reduz a densidade artificial da rede e melhora a interpretabilidade das estruturas resultantes, conforme discutido por [Donges et al. \(2009\)](#).

A aplicação de critérios rigorosos de significância estatística assegura que as conexões mantidas na rede representem relações fisicamente plausíveis e estatisticamente robustas, constituindo uma base confiável para a etapa subsequente de construção da matriz de adjacência.

### 3.2.4.3 Construção da Matriz de Adjacência

A construção da rede climática é realizada a partir da definição da matriz de adjacência  $A_{ij}$ , que representa a existência ou não de conexões entre os nós da rede. Cada nó corresponde a um ponto da grade espacial (latitude, longitude), e uma conexão entre dois nós indica a presença de uma interdependência estatisticamente significativa entre suas séries temporais de anomalias de MSLP.

Inicialmente, obtém-se a matriz de correlação  $\tau_{ij}$  de Kendall, calculada para todos os pares de nós da rede. Após a aplicação do critério de significância estatística ( $p < 0.05$ ), apenas os valores de correlação considerados estatisticamente relevantes são mantidos, enquanto os demais são descartados.

Em seguida, com o objetivo de controlar a densidade da rede e destacar apenas as interdependências mais fortes, foi adotado um critério adicional baseado em percentil. Somente os 5% maiores valores de correlação estatisticamente significativos foram utilizados para definir as conexões da rede, conforme metodologia proposta por [Gupta et al. \(2021\)](#). Esse limiar corresponde ao percentil de 95% da distribuição dos valores de correlação válidos.

A partir desses critérios, a matriz de adjacência  $A_{ij}$  é definida da seguinte forma:

$$A_{ij} = \begin{cases} 1, & \text{se } \tau_{ij} \text{ é estatisticamente significativo e pertence aos 5\% maiores valores} \\ 0, & \text{caso contrário.} \end{cases} \quad (3.1)$$

A matriz de adjacência resultante é binária, não direcionada e simétrica, refletindo o caráter bidirecional da correlação utilizada. Essa representação permite a aplicação direta das métricas de redes complexas, como grau do nó, distância geográfica média e coeficiente de agrupamento, analisadas nas seções subsequentes.

A adoção combinada de testes de significância estatística e limiar baseado em percentil garante um equilíbrio entre robustez estatística e sensibilidade física, evitando redes excessivamente densas ou dominadas por ruídos, ao mesmo tempo em que preserva as estruturas relevantes associadas à dinâmica dos ciclones tropicais.

### 3.2.5 Cálculo das Métricas de Redes Complexas

Uma vez construída a matriz de adjacência  $A_{ij}$ , a rede climática encontra-se completamente definida, permitindo a extração de métricas que caracterizam sua estrutura topológica e espacial. Utiliza-se três métricas clássicas de redes complexas amplamente utilizadas na literatura de redes climáticas: o grau do nó, a distância geográfica média e o coeficiente de agrupamento (suas definições formais estão apresentadas no Capítulo 2).

O grau do nó representa o número de conexões estabelecidas entre um ponto da rede climática e os demais pontos do domínio. Cada nó corresponde a um ponto da grade espacial e as conexões são definidas a partir de correlações estatisticamente significativas entre as séries temporais de MSLP. O grau foi calculado para cada nó das redes geradas em cada janela temporal analisada e posteriormente representado espacialmente. Valores elevados indicam regiões com maior número de interações estatísticas com o restante do domínio, enquanto valores reduzidos sugerem conectividade mais limitada e espacialmente concentrada. Na análise dos ciclones tropicais, variações espaciais dessa métrica são utilizadas como indicativo de alterações na organização das interações atmosféricas ao longo da trajetória dos sistemas.

A distância geográfica média quantifica o alcance espacial das conexões estabelecidas por cada nó da rede climática. Enquanto o grau indica o número de conexões, essa métrica descreve a separação espacial média entre os nós conectados, permitindo avaliar se as interações ocorrem predominantemente em escala local ou em longas distâncias. As distâncias geográficas são calculadas a partir das coordenadas de latitude e longitude dos pontos da grade, considerando a separação espacial sobre a superfície terrestre. A métrica foi obtida para cada nó e mapeada espacialmente nas diferentes janelas temporais analisadas. Valores elevados indicam conexões de maior alcance espacial, enquanto valores reduzidos sugerem interações predominantemente locais.

O coeficiente de agrupamento mede o grau de interconectividade local da rede, indicando a tendência de os vizinhos de um nó também estarem conectados entre si. Essa métrica foi utilizada para identificar regiões onde as conexões formam estruturas locais mais densas nas redes climáticas construídas. O coeficiente foi calculado para cada nó das redes obtidas nas diferentes janelas temporais e representado espacialmente. Valores próximos de zero indicam baixa interconectividade entre os vizinhos, enquanto valores próximos de um indicam a formação de estruturas locais densamente conectadas. Na análise das redes climáticas associadas aos ciclones tropicais, essa métrica é empregada como indicador da organização local das interações atmosféricas. Alterações espaciais no coeficiente de agrupamento podem evidenciar regiões onde a variabilidade da pressão ao nível do mar apresenta comportamento mais coerente e concentrado ao longo da trajetória dos sistemas ciclônicos.

### 3.2.6 Correção de Efeitos de Borda

Em redes espaciais limitadas a uma região específica, como é o caso das redes climáticas regionais utilizadas aqui, medidas topológicas podem ser significativamente influenciadas pela presença de fronteiras artificiais. Esse fenômeno, conhecido como efeito de borda, ocorre porque conexões potencialmente existentes entre nós localizados dentro da região de estudo e nós externos ao domínio não são observadas, alterando a estrutura da rede e os valores das métricas calculadas.

Em redes climáticas regionais, esse problema é particularmente relevante, uma vez que a análise é necessariamente restrita a uma área geográfica específica do globo. Nessas condições, nós localizados próximos às fronteiras tendem a apresentar valores artificialmente reduzidos de conectividade e alterações em métricas espaciais, não necessariamente associadas a processos físicos reais, mas sim à limitação do domínio analisado.

Além do limite geográfico da região de estudo, outro fator que contribui para a ocorrência de efeitos de borda no nosso contexto é a consideração exclusiva de pontos oceânicos na construção das redes climáticas. A aplicação da máscara terra-mar remove os pontos continentais da análise, fazendo com que regiões costeiras também passem a se comportar como áreas de fronteira topológica. Nessas áreas, conexões que poderiam existir com pontos sobre o continente deixam de ser consideradas, o que pode alterar artificialmente os valores das métricas calculadas.

Para mitigar essas limitações, foi aplicada uma correção de efeitos de borda baseada na abordagem de redes aleatórias espacialmente embutidas (*Spatially Embedded Random Networks* – SERN) (RHEINWALT et al., 2012). Essa estratégia permite estimar o viés introduzido pelo embutimento espacial da rede por meio da comparação entre a rede observada e um conjunto de redes aleatórias que preservam a distribuição espacial dos nós e a dependência das conexões em relação à distância geográfica.

Inicialmente, a posição geográfica dos nós é mantida fixa, considerando suas coordenadas de latitude e longitude. Em seguida, é estimada a probabilidade de conexão entre pares de nós

como função da distância geográfica a partir da rede original. Com base nessa relação, são geradas 1000 redes aleatórias espacialmente embutidas (*surrogates*) (GUPTA et al., 2021), que preservam tanto a distribuição espacial dos nós quanto a probabilidade de ligação dependente da distância. Um teste de convergência foi realizado variando o número de *surrogates* entre 100 e 2000. Verificou-se que a partir de aproximadamente 400 redes as métricas corrigidas estabilizam com variação muito pequena. Optou-se por gerar 1000 redes a fim de manter fidelidade à metodologia adotada por Gupta et al. (2021).

Para cada uma dessas redes substitutas, são recalculadas as métricas de interesse (grau do nó, distância geográfica média e coeficiente de agrupamento) resultando em um conjunto de valores esperados para cada nó considerando apenas o efeito do embutimento espacial da rede. A média dessas métricas no conjunto de redes substitutas é então utilizada como referência para estimar o viés introduzido pelas fronteiras geográficas e pela própria configuração espacial dos nós.

Então, as métricas da rede climática original são ajustadas em relação ao valor esperado devido aos efeitos de borda por meio da razão

$$M_{\text{corr}} = \frac{M_{\text{obs}}}{\langle M_{\text{SERN}} \rangle},$$

onde  $M_{\text{obs}}$  é o valor da métrica (grau, distância geográfica média ou coeficiente de agrupamento) calculado na rede real e  $\langle M_{\text{SERN}} \rangle$  é a média da mesma métrica obtida no conjunto de 1000 SERN. Essa normalização produz uma métrica adimensional que representa o valor relativo ao esperado pela incorporação espacial da rede, removendo o viés geométrico introduzido pelas fronteiras artificiais da região de estudo. A formulação segue exatamente a abordagem proposta por Rheinwalt et al. (2012) e adotada por Gupta et al. (2021). Como resultado, obtêm-se campos espaciais corrigidos que reduzem a influência artificial das fronteiras regionais e das zonas costeiras, preservando predominantemente os padrões associados à dinâmica física do sistema atmosférico.

Esse procedimento é aplicado individualmente às métricas de grau, coeficiente de agrupamento e distância geográfica média, tanto para as janelas temporais anteriores quanto para aquelas correspondentes à atuação dos ciclones. Como resultado, obtêm-se campos espaciais corrigidos que reduzem a influência artificial das fronteiras regionais e das zonas costeiras, preservando predominantemente os padrões associados à dinâmica física do sistema atmosférico.

A incorporação dessa etapa ao pipeline computacional aumenta a robustez das análises e a confiabilidade das interpretações físicas das métricas de rede, assegurando que as variações observadas estejam majoritariamente associadas à presença e à evolução dos ciclones tropicais, e não a artefatos decorrentes da delimitação espacial da região ou da exclusão dos pontos continentais.



### 3.2.7 Geração das Visualizações

A etapa final do pipeline computacional consiste na geração das visualizações espaciais das métricas de redes complexas, permitindo a análise qualitativa e comparativa dos padrões associados aos ciclones tropicais. As visualizações são produzidas automaticamente a partir das métricas previamente calculadas e corrigidas para efeitos de borda, garantindo consistência entre as diferentes execuções do experimento.

As métricas analisadas são inicialmente armazenadas em formato vetorial, correspondendo aos nós oceânicos válidos definidos após a aplicação da máscara terra-mar. Para a geração dos mapas, esses vetores são reestruturados para o formato bidimensional da grade espacial original, restabelecendo a correspondência entre cada valor da métrica e sua posição geográfica (latitude e longitude).

Para ciclones que possuem intervalos temporais definidos tanto para o período anterior quanto para o período de atuação do sistema, as visualizações são organizadas em painéis comparativos, nos quais cada métrica é representada em duas linhas correspondentes às janelas temporais “antes” e “durante”. Quando apenas o período de atuação do ciclone está disponível, os mapas são apresentados em uma única linha. Essa organização facilita a análise das alterações estruturais da rede ao longo da evolução do sistema atmosférico.

As escalas de cores são definidas de forma consistente entre os mapas comparáveis, permitindo a avaliação direta das variações espaciais das métricas entre diferentes períodos. Para cada métrica, os limites mínimo e máximo da escala são determinados a partir do conjunto de valores analisados, evitando distorções visuais decorrentes de normalizações independentes. As regiões continentais são representadas como áreas nulas com uma cor cinza, destacando o domínio oceânico efetivamente considerado na construção das redes climáticas.

As figuras são geradas em alta resolução e armazenadas automaticamente em diretórios organizados por região e ciclone, assegurando rastreabilidade e reprodutibilidade dos resultados. Esse procedimento permite a rápida inspeção visual das estruturas emergentes, bem como sua utilização direta nas etapas de análise e discussão dos resultados.

A integração dessa etapa ao pipeline computacional garante que todo o processo seja executado de forma automatizada e padronizada, reduzindo intervenções manuais e aumentando a consistência metodológica do estudo.

### 3.2.8 Ferramentas utilizadas

O desenvolvimento do pipeline computacional, bem como as etapas de processamento dos dados, construção das redes climáticas e geração das visualizações, foram realizados na linguagem *Python*. A escolha dessa linguagem deve-se à sua ampla utilização em computação científica e à disponibilidade de bibliotecas especializadas para manipulação de dados climáticos, análise estatística, modelagem de redes e visualização científica.

As principais bibliotecas empregadas neste trabalho são descritas a seguir. Para garantir a reprodutibilidade completa dos resultados, são informadas as versões exatas utilizadas no ambiente de desenvolvimento e execução do pipeline.

- **cdsapi** (versão 0.7.7)<sup>3</sup>: utilizada para a obtenção automatizada dos dados da reanálise ERA5 por meio da API oficial da *Copernicus Climate Data Store*;
- **xarray** (versão 2025.12.0)<sup>4</sup>: empregada na leitura, organização e manipulação de arquivos no formato *NetCDF*, permitindo o tratamento eficiente de dados multidimensionais;
- **numpy** (versão 2.4.0)<sup>5</sup>: utilizada em operações numéricas, manipulação de matrizes e cálculos estatísticos necessários à construção das redes climáticas;
- **pandas** (versão 2.3.3)<sup>6</sup>: utilizada na manipulação de séries temporais e no gerenciamento de estruturas tabulares auxiliares ao processamento dos dados;
- **scipy** (versão 1.16.3)<sup>7</sup>: empregada na implementação de métodos estatísticos, em especial no cálculo da correlação  $\tau$  de Kendall utilizada para quantificar a similaridade entre séries temporais de pressão ao nível do mar;
- **networkx** (versão 3.6.1)<sup>8</sup>: utilizada para a construção, representação e análise das redes complexas a partir das matrizes de adjacência obtidas;
- **matplotlib** (versão 3.10.8)<sup>9</sup>: utilizada na geração das visualizações espaciais das métricas de rede e na produção das figuras empregadas na análise dos resultados;
- **joblib** (versão 1.5.3)<sup>10</sup>: empregada na paralelização de tarefas computacionalmente intensivas, como a geração de redes substitutas utilizadas na correção de efeitos de borda;
- **tqdm** (versão 4.67.1)<sup>11</sup>: utilizada para monitoramento do progresso das rotinas de processamento e geração de redes substitutas;
- **sern**: implementação obtida do repositório de Rheinwalt et al. (2012)<sup>12</sup>, utilizada para a geração de SERN, empregadas na etapa de correção de efeitos de borda;
- Módulos da biblioteca padrão do Python:

<sup>3</sup> <<https://cds.climate.copernicus.eu/how-to-api>>

<sup>4</sup> <<https://docs.xarray.dev/en/v2025.12.0/getting-started-guide/installing.html>>

<sup>5</sup> <<https://numpy.org/doc/stable/release/2.4.0-notes.html>>

<sup>6</sup> <<https://pandas.pydata.org/docs/whatsnew/v2.3.3.html>>

<sup>7</sup> <<https://scipy.org/news/#scipy-1163-released>>

<sup>8</sup> <<https://pypi.org/project/networkx/3.6.1/>>

<sup>9</sup> <[https://matplotlib.org/stable/users/release\\_notes#version-3-10](https://matplotlib.org/stable/users/release_notes#version-3-10)>

<sup>10</sup> <<https://pypi.org/project/joblib/1.5.3/>>

<sup>11</sup> <<https://tqdm.github.io/releases/#v4671-2024-11-27>>

<sup>12</sup> Mais informações sobre a implementação de SERN podem ser encontradas no repositório <<https://github.com/Rheinwalt/spatial-effects-networks>>.

- **pickle**: utilizada para armazenamento e recuperação das estruturas de dados intermediárias e das métricas calculadas ao longo do pipeline;
- **itertools**: empregada na geração de combinações entre pares de nós durante a construção das matrizes de correlação;
- **glob** e **os**: utilizadas para gerenciamento de arquivos e diretórios no fluxo de processamento;
- **datetime**: utilizada no tratamento e organização das informações temporais associadas aos dados e aos eventos de ciclones analisados;
- **math**: utilizada em operações matemáticas auxiliares, incluindo cálculos de distância geográfica.

O ambiente de execução utilizou a versão **Python 3.13.11**<sup>13</sup>. Recomenda-se a utilização de um ambiente com essas versões exatas para reproduzir os resultados sem incompatibilidades.

O uso integrado dessas bibliotecas possibilitou a implementação de um pipeline automatizado, modular e reprodutível, abrangendo desde a obtenção dos dados meteorológicos até a construção das redes climáticas, aplicação de correções metodológicas e geração das visualizações finais.

### 3.2.9 Reprodutibilidade e disponibilização pública do código

Com o objetivo de garantir a reprodutibilidade dos experimentos e a transparência metodológica, todo o pipeline computacional desenvolvido neste trabalho está disponibilizado publicamente em um repositório *GitHub*<sup>14</sup>. O repositório inclui:

- códigos para aquisição dos dados ERA5 via API;
- rotinas de pré-processamento e cálculo das anomalias;
- implementação da construção das redes climáticas;
- scripts para cálculo das métricas e correção de efeitos de borda;
- rotinas de geração das visualizações;
- dados intermediários necessários à reprodução das análises.

A disponibilização pública desses recursos permite que outros pesquisadores reproduzam integralmente os resultados apresentados, além de possibilitar a extensão do pipeline para novos eventos e regiões de estudo.

<sup>13</sup> <<https://www.python.org/downloads/release/python-31311/>>

<sup>14</sup> <<https://github.com/Cossio00/functional-networks-for-tropical-cyclones.git>>

## 4 Resultados

Este Capítulo apresenta os resultados obtidos a partir da execução do *pipeline* computacional desenvolvido para a geração e análise de redes climáticas associadas a ciclones tropicais. A aplicação do sistema permitiu a obtenção automatizada dos dados da reanálise ERA5, o processamento das séries temporais e a geração das métricas de rede e visualizações utilizadas na análise dos eventos.

A execução do *pipeline* resultou na construção de redes funcionais e mapas das métricas topológicas para diferentes ciclones tropicais e regiões oceânicas, possibilitando a observação da organização espacial das interdependências atmosféricas durante a ocorrência desses eventos. Em particular, foi conduzido um estudo de caso detalhado para o ciclone Gaja, incluindo a análise comparativa entre períodos anteriores e durante a ocorrência do sistema.

### 4.1 Estudo de caso: ciclone tropical Gaja

Como forma de validar a aplicação do *pipeline* e analisar o comportamento das métricas de rede em um evento real, foi conduzido um estudo de caso com o ciclone tropical Gaja, seguindo a mesma configuração temporal e espacial adotada por [Gupta et al. \(2021\)](#). A escolha desse evento deve-se à disponibilidade de referências na literatura e à possibilidade de comparação direta com resultados previamente reportados. A execução do pipeline para este evento utilizou os seguintes parâmetros:

- ***product\_type***: "reanalysis";
- ***variable***: "mean\_sea\_level\_pressure";
- ***year***: [str(y) for y in range(1979, 2019)] (todos os anos de 1979 a 2018);
- ***month***: [str(m) for m in range(1, 13) if not (m == 2 and day == "29")] (todos os meses exceto fevereiro quando o dia é 29);
- ***day***: [str(d).zfill(2) for d in range(1, 32)] (todos os dias de um mês, de 1 a 31);
- ***time***: ["00:00", "03:00", "06:00", "09:00", "12:00", "15:00", "18:00", "21:00"];
- ***data\_format***: "netcdf\_legacy";
- ***download\_format***: "unarchived";
- ***area***: region["area"] (obtida do dicionário de regiões. Neste caso, [34.5, 49.5, 4.5, 100]);
- ***grid***: "0.75/0.75".

A obtenção dos dados brutos é realizada por região geográfica e não por ciclone específico. Dessa forma, todos os ciclones que ocorrem na mesma região podem compartilhar o mesmo conjunto de arquivos baixados, desde que suas datas de ocorrência estejam contidas no intervalo temporal definido para o download (no caso, 1979 a 2018).

Para analisar um ciclone localizado em outra região, basta alterar o valor da variável REGION no script de aquisição de dados (ver Figura 3.2). Já no estágio de pré-processamento (cálculo de anomalias, seleção de janelas temporais e construção das redes), é necessário informar o ciclone de interesse por meio da variável CYCLONE (ver Figura 3.4).

É fundamental que os dicionários de regiões e ciclones estejam corretamente preenchidos, contendo, respectivamente, as coordenadas de delimitação geográfica de cada região e as informações temporais e regionais precisas de cada evento ciclônico analisado.

Para o ciclone Gaja, foram gerados mapas das métricas de rede considerando janelas temporais de 10 dias. Diferentemente dos demais eventos analisados nesta monografia, para esse caso específico foram avaliados dois cenários: um período anterior à ocorrência do ciclone e o período correspondente à sua atuação sobre a região. Essa abordagem permite observar mudanças estruturais associadas à presença do sistema atmosférico.

A Figura 4.1 apresenta, em uma única visualização, os mapas das três métricas analisadas antes e durante a ocorrência do ciclone Gaja. A primeira linha corresponde ao período anterior ao evento, enquanto a segunda representa o período durante sua atuação. Os eixos  $x$  e  $y$  representam, respectivamente, a longitude e a latitude geográficas dos pontos da grade espacial. A trajetória do ciclone é representado por círculos pretos

Antes da passagem do ciclone, as métricas apresentam distribuição espacial relativamente homogênea, sem estruturas marcantes associadas a sistemas atmosféricos organizados. Durante o evento, entretanto, surgem padrões espaciais bem definidos ao longo da trajetória do sistema, indicando modificações na topologia da rede climática.

De acordo com Gupta et al. (2021), ciclones tropicais tendem a induzir uma reorganização espacial da rede caracterizada por redução do grau e da distância geográfica média nos nós localizados ao longo da trajetória do sistema e aumento do coeficiente de agrupamento nessas mesmas regiões. Esse comportamento está associado à natureza localizada do fenômeno, que promove conexões mais curtas e concentradas espacialmente, formando estruturas coesas relativamente isoladas do restante da rede.

Os resultados obtidos para o ciclone Gaja reproduzem esse padrão (ver Figura 4.1). Observa-se que o grau do nó tende a assumir valores menores ao longo da trajetória do ciclone, enquanto regiões adjacentes apresentam conectividade relativamente mais elevada. De modo semelhante, a distância geográfica média apresenta valores reduzidos nas áreas diretamente afetadas, refletindo a predominância de interações espaciais de menor alcance durante o evento.

Em contraste, o coeficiente de agrupamento apresenta aumento significativo nas regiões

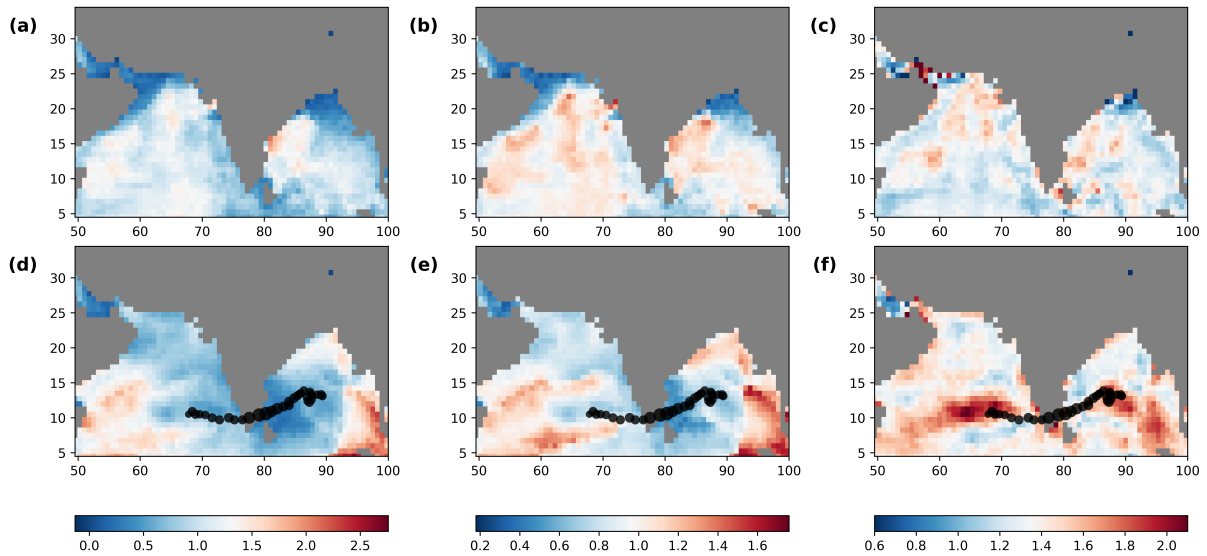


Figura 4.1 – Comparação entre grau (a) e (d), distância geográfica média (b) e (e) e coeficiente de agrupamento (c) e (f) antes e durante a passagem do ciclone Gaja. As figuras em uma mesma coluna apresentam a mesma escala de cores. As figuras a, b e c foram construídas em um período anterior à passagem do ciclone Gaja (29 de outubro de 2018 a 7 de novembro de 2018). As figuras d, e e f foram construídas sobre um período durante a passagem do ciclone Gaja (10 a 19 de novembro de 2018).

próximas à trajetória do ciclone, especialmente nas áreas associadas ao núcleo do sistema. Esse comportamento indica a formação de estruturas locais mais coesas, nas quais os nós apresentam forte interconectividade entre si, configurando agrupamentos espaciais contínuos ao longo da trilha do ciclone.

Esse conjunto de padrões evidencia que a presença do ciclone não implica aumento uniforme da conectividade da rede, mas sim uma redistribuição estrutural das relações entre os nós. A rede passa a apresentar regiões com conexões mais curtas e densamente agrupadas ao longo da trajetória do sistema, ao mesmo tempo em que áreas adjacentes mantêm ou ampliam conexões de maior alcance espacial.

De modo geral, os resultados obtidos para o ciclone Gaja são consistentes com o comportamento reportado por Gupta et al. (2021), no qual ciclones tropicais produzem regiões de baixo grau e baixa distância geográfica média ao longo de suas trajetórias, acompanhadas por elevados valores de coeficiente de agrupamento. Esse padrão reflete a formação de agrupamentos locais relativamente isolados e a reorganização espacial da rede climática durante o evento.

## 4.2 Aplicação a outros ciclones tropicais

Após a análise detalhada do ciclone Gaja, o *pipeline* foi aplicado a outros eventos com o objetivo de avaliar sua capacidade de generalização e verificar se os padrões estruturais observados nas métricas de rede se mantêm em diferentes sistemas atmosféricos.

Foram analisados os ciclones Luban, Megh e Vardah, localizados na mesma bacia oceânica do Gaja, além do furacão Irma, pertencente ao Atlântico Norte. Para todos os casos, foram consideradas janelas temporais de 10 dias centradas no período de atuação do ciclone, seguindo a mesma configuração metodológica adotada anteriormente.

Para cada evento, foram gerados mapas das métricas de grau, distância geográfica média e coeficiente de agrupamento durante a ocorrência do ciclone. Diferentemente do estudo conduzido para o Gaja, não foi realizada comparação explícita com períodos anteriores ao evento, priorizando-se a análise das estruturas de rede formadas durante a atuação dos sistemas.

Os resultados obtidos indicam a recorrência de padrões espaciais semelhantes entre os diferentes eventos analisados. De forma consistente, as regiões ao longo da trajetória dos ciclones apresentam redução nos valores de grau e distância geográfica média, acompanhadas por aumentos no coeficiente de agrupamento local. Esses padrões refletem a formação de estruturas mais coesas e espacialmente confinadas durante a atuação dos sistemas atmosféricos.

As Figuras 4.2, 4.3 e 4.4 ilustram esses comportamentos para ciclones localizados na mesma bacia oceânica do Gaja, evidenciando a repetição das assinaturas topológicas nas métricas de rede. Observa-se que as trilhas dos ciclones tornam-se visualmente identificáveis nos mapas, especialmente por meio da combinação entre baixos valores de grau e distância geográfica média e altos valores de coeficiente de agrupamento.

O caso do furacão Irma, apresentado na Figura 4.5, reforça ainda mais a robustez da abordagem proposta. Mesmo pertencendo a uma bacia oceânica distinta, os padrões observados nas métricas mantêm-se consistentes, com organização espacial semelhante àquela identificada nos demais eventos. Esse resultado indica que o comportamento topológico das redes climáticas associadas a ciclones tropicais não depende exclusivamente da região analisada, mas está relacionado à própria dinâmica atmosférica desses sistemas.

De modo geral, a aplicação do *pipeline* a diferentes ciclones e domínios espaciais demonstra sua capacidade de adaptação e evidencia a presença de padrões recorrentes nas métricas de rede durante a ocorrência desses eventos, reforçando seu potencial como ferramenta para análise sistemática de ciclones tropicais.

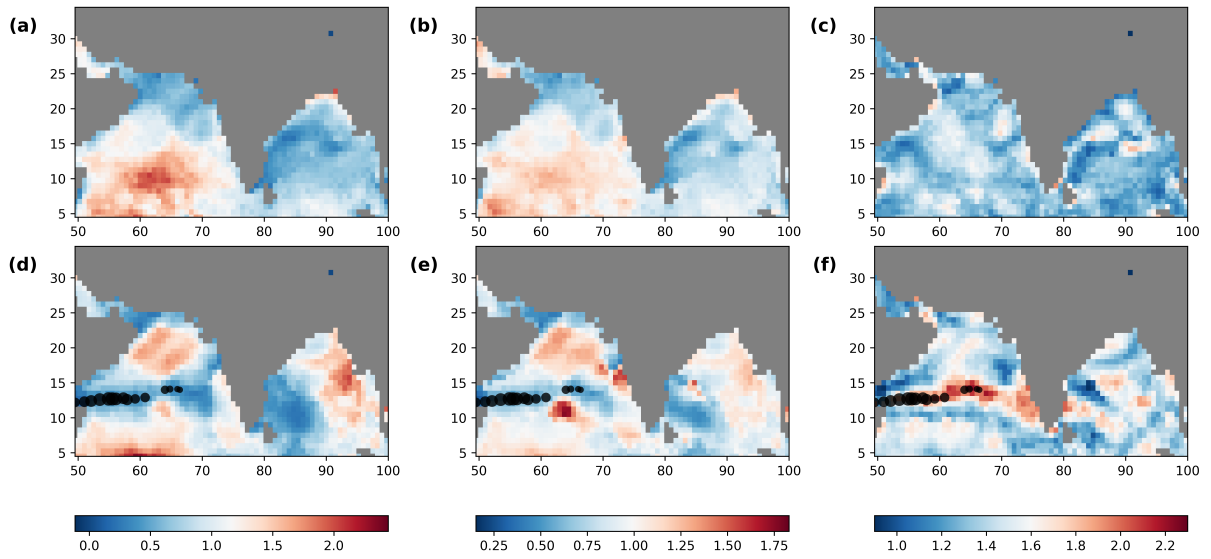


Figura 4.2 – Comparação entre grau (a) e (d), distância geográfica média (b) e (e) e coeficiente de agrupamento (c) e (f) antes e durante a passagem do Ciclone Megh. As figuras em uma mesma coluna apresentam a mesma escala de cores. As figuras a, b e c foram construídas em um período anterior à passagem do Ciclone Megh (22 a 31 de outubro de 2015). As figuras d, e e f foram construídas sobre um período durante a passagem do Ciclone Megh (3 a 12 de novembro de 2015).

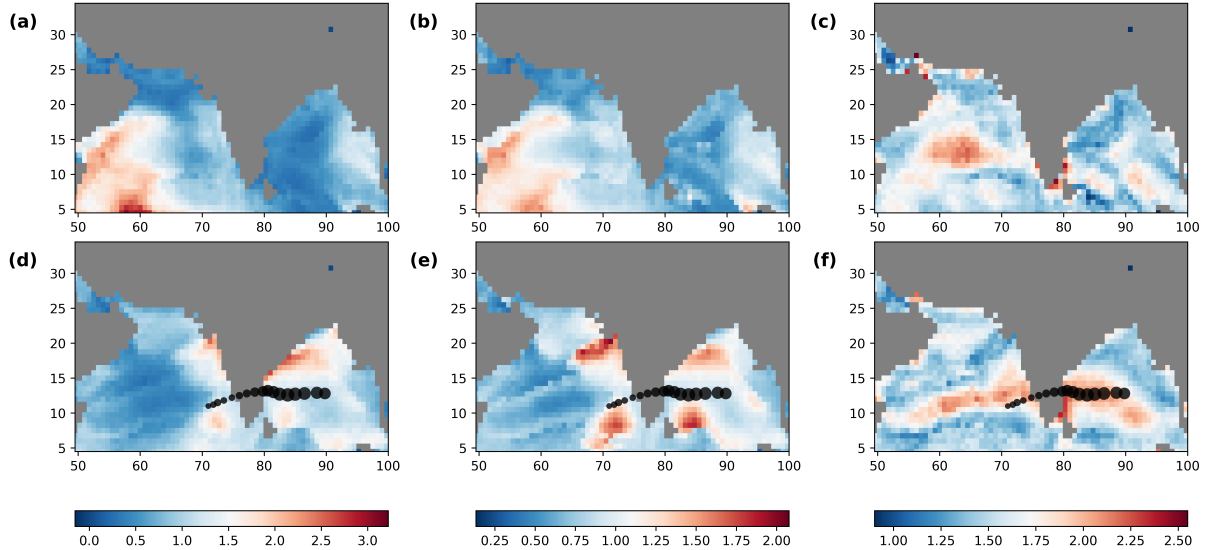


Figura 4.3 – Comparação entre grau (a) e (d), distância geográfica média (b) e (e) e coeficiente de agrupamento (c) e (f) antes e durante a passagem do Ciclone Vardah. As figuras em uma mesma coluna apresentam a mesma escala de cores. As figuras a, b e c foram construídas em um período anterior à passagem do Ciclone Vardah (27 de novembro de 2016 a 6 de dezembro de 2016). As figuras d, e e f foram construídas sobre um período durante a passagem do ciclone Vardah (9 a 18 de dezembro de 2016).



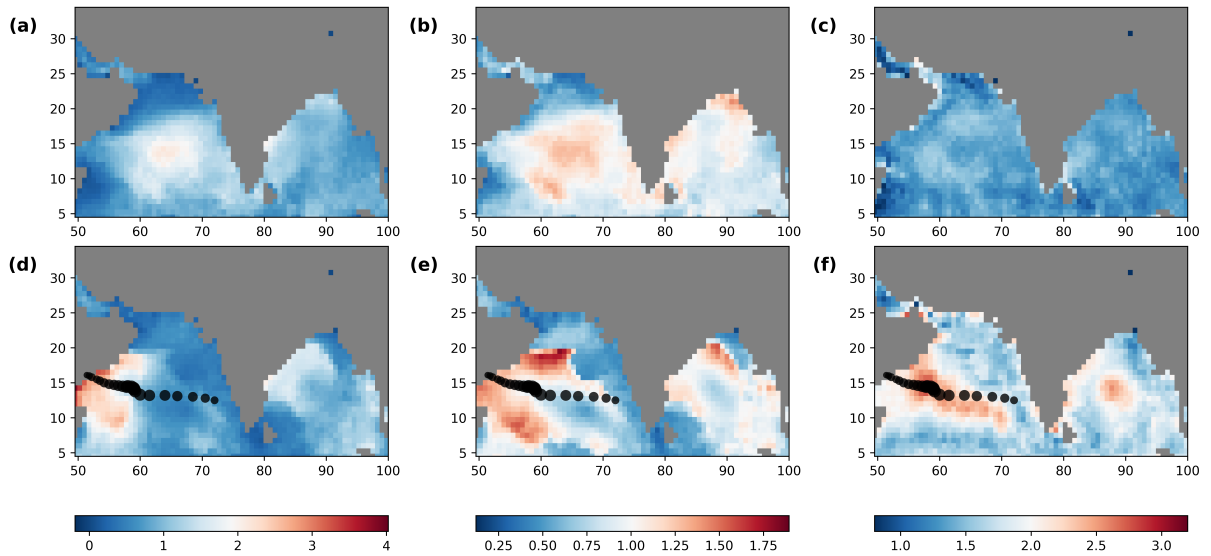


Figura 4.4 – Comparação entre grau (a) e (d), distância geográfica média (b) e (e) e coeficiente de agrupamento (c) e (f) antes e durante a passagem do Ciclone Luban. As figuras em uma mesma coluna apresentam a mesma escala de cores. As figuras a, b e c foram construídas em um período anterior à passagem do Ciclone Luban (20 a 29 de setembro de 2018). As figuras d, e e f foram construídas sobre um período durante a passagem do Ciclone Luban (2 a 11 de outubro de 2018).

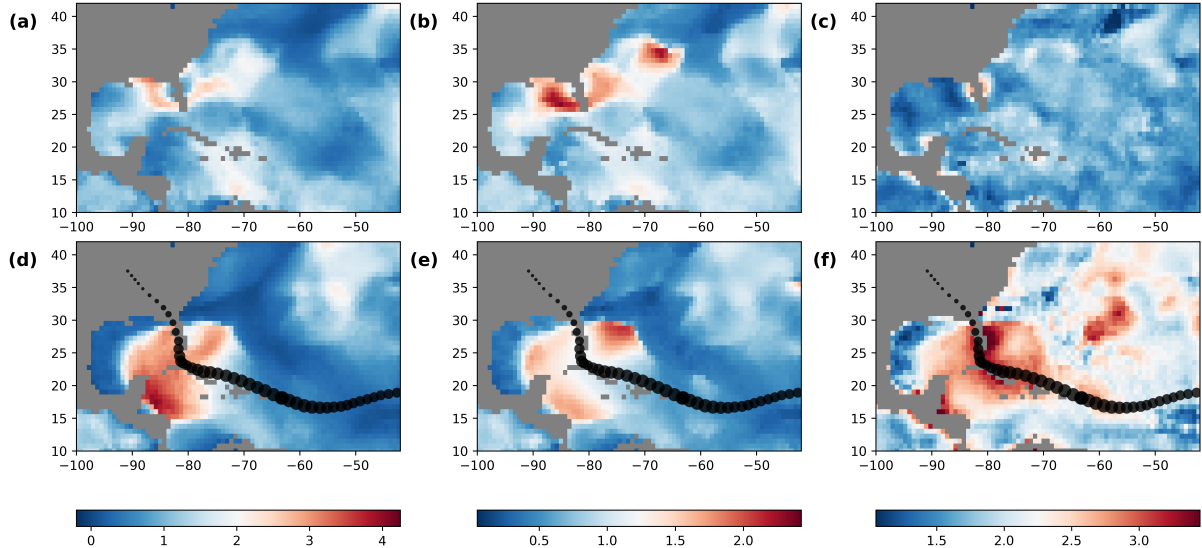


Figura 4.5 – Comparação entre grau (a) e (d), distância geográfica média (b) e (e) e coeficiente de agrupamento (c) e (f) antes e durante a passagem do Ciclone Irma. As figuras em uma mesma coluna apresentam a mesma escala de cores. As figuras a, b e c foram construídas em um período anterior à passagem do Ciclone Irma (20 a 29 de agosto de 2017). As figuras d, e e f foram construídas sobre um período durante a passagem do Ciclone Irma (1 a 10 de setembro de 2017).

### 4.3 Validação do *Pipeline*

Os resultados obtidos neste trabalho foram comparados com aqueles reportados por [Gupta et al. \(2021\)](#), que constituem a principal referência teórica e metodológica desta pesquisa. O objetivo dessa comparação é verificar a consistência dos padrões observados nas redes climáticas geradas pelo *pipeline* desenvolvido, bem como avaliar sua capacidade de reproduzir comportamentos já descritos na literatura.

De forma geral, as métricas obtidas neste estudo apresentaram comportamento qualitativamente semelhante ao reportado por [Gupta et al. \(2021\)](#) durante a ocorrência de ciclones tropicais. Em particular, observou-se a redução dos valores de grau e de distância geográfica média ao longo das trajetórias dos ciclones, acompanhada por aumentos no coeficiente de agrupamento local nessas regiões. Esse padrão reflete a formação de estruturas de conectividade espacialmente confinadas, nas quais os nós associados ao sistema atmosférico estabelecem interações predominantemente locais.

Esse comportamento está em concordância com a interpretação proposta por [Gupta et al. \(2021\)](#), segundo a qual ciclones tropicais constituem sistemas altamente organizados e espacialmente concentrados, capazes de reconfigurar temporariamente a topologia das redes climáticas. A diminuição do grau e da distância geográfica média indica a predominância de conexões de curta escala espacial, enquanto o aumento do coeficiente de agrupamento sugere a formação de grupos coesos de nós ao longo da trilha do ciclone.

Além da análise do ciclone Gaja, a aplicação do *pipeline* a outros eventos, como os Ciclones Luban, Megh, Vardah e Irma, revelou padrões consistentes com aqueles descritos na literatura, mesmo em diferentes bacias oceânicas. Esse resultado reforça a hipótese de que as assinaturas topológicas observadas nas métricas de rede estão associadas à dinâmica intrínseca dos ciclones tropicais e não a características específicas de uma única região.

Embora os resultados obtidos sejam majoritariamente consistentes com os de [Gupta et al. \(2021\)](#), pequenas diferenças espaciais podem ser observadas na distribuição das métricas, possivelmente decorrentes de fatores como a aplicação da máscara terra-mar, procedimentos de correção de efeitos de borda e características próprias do conjunto ERA5, que consiste em uma reanálise atmosférica baseada em assimilação de dados e reprocessamentos periódicos. Tais variações são esperadas em estudos baseados em dados climáticos e não comprometem a interpretação geral dos padrões observados ([HERSBACH et al., 2020](#)).

Nesse contexto, os resultados apresentados fornecem evidências de que o *pipeline* desenvolvido é capaz de ser aplicado na análise de redes climáticas associadas a ciclones tropicais.

## 4.4 Contribuições do trabalho

Os resultados apresentados ao longo deste capítulo evidenciam contribuições relevantes deste trabalho no contexto da análise de redes climáticas associadas a ciclones tropicais, tanto do ponto de vista metodológico quanto computacional.

Em primeiro lugar, destaca-se a implementação de um *pipeline* automatizado capaz de integrar todas as etapas necessárias à construção e análise de redes climáticas a partir de dados de reanálise. Diferentemente de abordagens baseadas em procedimentos manuais ou scripts isolados, a solução desenvolvida permite a execução contínua do fluxo completo de processamento, desde a obtenção dos dados do ERA5 até a geração das métricas e visualizações, a partir da simples parametrização da região e do evento de interesse.

Os experimentos conduzidos demonstraram que o *pipeline* é capaz de reproduzir padrões estruturais consistentes com aqueles descritos na literatura para ciclones tropicais, tanto no estudo detalhado do ciclone Gaja quanto na aplicação a outros eventos. A recorrência de assinaturas topológicas semelhantes em diferentes casos indica que a abordagem implementada preserva os principais aspectos físicos e estruturais das redes climáticas associadas a esses sistemas atmosféricos.

Além disso, a aplicação do método a ciclones pertencentes a diferentes bacias oceânicas, como o caso do furacão Irma no Atlântico Norte, evidencia a capacidade de generalização do *pipeline*. Esse resultado sugere que a metodologia não está restrita a uma região específica, podendo ser empregada em distintos contextos climáticos e geográficos sem necessidade de adaptações estruturais no código.

Outro aspecto relevante é a incorporação explícita de etapas de pré-processamento e correção metodológica, como o cálculo de climatologia, geração de anomalias, aplicação de máscara terra-mar e procedimentos de correção de efeitos de borda. A integração desses elementos contribui para a consistência física das redes geradas e para a confiabilidade das métricas obtidas, aproximando o processo computacional das práticas adotadas em estudos científicos da área.

Do ponto de vista da reprodutibilidade, o trabalho oferece uma estrutura organizada e documentada que permite a replicação dos experimentos e sua extensão para novos cenários. A definição das regiões e dos eventos por meio de dicionários, aliada à execução automatizada das etapas do *pipeline*, reduz a dependência de intervenções manuais e facilita sua utilização por outros pesquisadores.

Por fim, os grafos e métricas gerados constituem uma base potencial para investigações futuras envolvendo técnicas de aprendizado de máquina aplicadas a dados estruturados, em especial modelos baseados em Redes Neurais de Grafos (Graph Neural Networks). Nesse contexto, o *pipeline* desenvolvido não apenas viabiliza a análise de eventos extremos, mas também estabelece uma infraestrutura computacional capaz de sustentar estudos mais avançados voltados à detecção, classificação e monitoramento automatizado de fenômenos atmosféricos.

## 5 Considerações finais

Este trabalho teve como objetivo principal desenvolver e aplicar um *pipeline* computacional para a construção e análise de redes climáticas associadas a ciclones tropicais, a partir de dados de pressão ao nível do mar provenientes da reanálise ERA5. A proposta integrou, de forma automatizada, as etapas de aquisição de dados, pré-processamento, construção das redes funcionais, cálculo de métricas topológicas, correção de efeitos de borda e geração de visualizações espaciais, permitindo a análise sistemática da organização das interdependências atmosféricas durante a ocorrência desses eventos extremos.

Os resultados obtidos demonstraram que o *pipeline* é capaz de gerar redes climáticas consistentes e reproduzir padrões estruturais descritos na literatura para ciclones tropicais. No estudo de caso do ciclone Gaja, observou-se a reorganização da topologia da rede ao longo da trajetória do sistema, caracterizada pela redução do grau e da distância geográfica média e pelo aumento do coeficiente de agrupamento nas regiões diretamente afetadas. Esse comportamento está associado à formação de estruturas locais mais coesas e espacialmente confinadas, refletindo a natureza organizada e de escala relativamente limitada desses sistemas atmosféricos.

A aplicação do método a outros ciclones, em diferentes regiões oceânicas, evidenciou a recorrência dessas assinaturas topológicas, indicando que os padrões observados não se restringem a um evento específico, mas estão associados à dinâmica intrínseca dos ciclones tropicais. Esse resultado reforça a robustez da abordagem adotada e demonstra o potencial das redes climáticas como ferramenta para a análise estrutural de fenômenos atmosféricos extremos.

Do ponto de vista metodológico, uma contribuição central deste trabalho é a implementação de um fluxo computacional integrado e reproduzível, capaz de automatizar a geração de dados, redes e métricas a partir da simples parametrização da região e do evento de interesse. A incorporação de etapas como cálculo de anomalias, aplicação de máscara terra-mar e correção de efeitos de borda contribui para a consistência física das redes construídas e para a confiabilidade das análises realizadas.

Apesar dos resultados promissores, algumas limitações devem ser consideradas. A análise foi baseada exclusivamente em dados de pressão ao nível do mar e em redes construídas a partir de correlações estatísticas, o que implica que outras variáveis atmosféricas relevantes, como vento e temperatura, não foram diretamente incorporadas. Além disso, a definição das redes depende de escolhas metodológicas específicas, como o limiar de significância estatística e o tamanho das janelas temporais, que podem influenciar a estrutura resultante.

Outra limitação está associada ao próprio escopo regional das redes analisadas. Embora a correção de efeitos de borda tenha sido aplicada, a delimitação espacial e a consideração exclusiva de pontos oceânicos podem introduzir restrições na representação completa das interações

atmosféricas em escala global.

Como perspectivas futuras, destacam-se diversas possibilidades de extensão deste trabalho. Entre elas, incluem-se a incorporação de múltiplas variáveis climáticas na construção das redes, a análise de séries temporais mais longas e a investigação de outros tipos de eventos extremos, como frentes frias e sistemas convectivos de grande escala. Adicionalmente, os grafos gerados neste estudo constituem uma base promissora para a aplicação de técnicas de aprendizado de máquina, especialmente modelos baseados em Redes Neurais de Grafos, com potencial para a detecção automática e a previsão de padrões associados a ciclones tropicais.

Por fim, ressalta-se que o *pipeline* desenvolvido estabelece uma infraestrutura computacional capaz de sustentar investigações futuras no campo das redes climáticas, contribuindo para a integração entre métodos de ciência de dados e a análise de fenômenos atmosféricos. Dessa forma, este trabalho não apenas reproduz resultados previamente descritos na literatura, mas também amplia as possibilidades de aplicação de redes complexas no estudo sistemático de eventos climáticos extremos.

# Referências

- BOCCALETTI, S.; LATORA, V.; MORENO, Y.; CHAVEZ, M.; HWANG, D.-U. Complex networks: Structure and dynamics. *Physics Reports*, v. 424, n. 4, p. 175–308, 2006.
- BOERS, N.; GOSWAMI, B.; RHEINWALT, A.; BOOKHAGEN, B.; HOSKINS, B.; KURTHS, J. Complex networks reveal global pattern of extreme-rainfall teleconnections. *Nature*, v. 566, n. 7744, p. 373–377, 2019. ISSN 14764687.
- CHARTRAND, G.; ZHANG, P. *A first course in graph theory*. [S.l.]: Courier Corporation, 2013.
- DONGES, J. F.; ZOU, Y.; MARWAN, N.; KURTHS, J. The backbone of the climate network. *Europhysics Letters*, v. 87, n. 4, p. 48007, 2009.
- GOSWAMI, B.; SCHULTZ, P.; HEINZE, B.; MARWAN, N.; BODIRSKY, B. L.; LOTZE-CAMPEN, H.; KURTHS, J. Inferring interdependencies from short time series. 2017. Disponível em: <<https://api.semanticscholar.org/CorpusID:158478916>>.
- GUPTA, S.; BOERS, N.; PAPPENBERGER, F.; KURTHS, J. Complex network approach for detecting tropical cyclones. *Climate Dynamics*, v. 57, p. 3355–3364, 2021.
- HERSBACH, H.; BELL, B.; BERRISFORD, P.; HIRAHARA, S.; HORÁNYI, A.; MUÑOZ-SABATER, J.; NICOLAS, J.; PEUBEY, C.; RADU, R.; SCHEPERS, D. et al. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, v. 146, p. 1999–2049, 2020.
- JORGE, A. A.; COSTA, I. C.; SANTOS, L. B. Geographical complex networks applied to describe meteorological data. *GEOINFO*, v. 21, p. 258–263, 2020.
- KENDALL, M. G. A new measure of rank correlation. *Biometrika*, v. 30, p. 81–93, 1938. Disponível em: <<https://doi.org/10.1093/biomet/30.1-2.81>>.
- KNUTSON, T.; CAMARGO, S. J.; CHAN, J. C.; EMANUEL, K.; HO, C.-H.; KOSSIN, J.; MOHAPATRA, M.; SATOH, M.; SUGI, M.; WALSH, K. et al. Tropical cyclones and climate change assessment: Part ii: Projected response to anthropogenic warming. *Bulletin of the American Meteorological Society*, American Meteorological Society, v. 101, n. 3, p. E303–E322, 2020.
- LIN, T. Q. J.; KLOTZBACH, P. Tropical cyclones. *Atmosphere-Ocean*, v. 60, p. 360–398, 2022.
- MALIK, N.; BOOKHAGEN, B.; MARWAN, N.; KURTHS, J. Analysis of spatial and temporal extreme monsoonal rainfall over south asia using complex networks. *Climate dynamics*, Springer, v. 39, n. 3, p. 971–987, 2012.
- NEWMAN, M. *Networks: An Introduction*. Oxford University Press, 2010. ISBN 9780199206650. Disponível em: <<https://doi.org/10.1093/acprof:oso/9780199206650.001.0001>>.
- PÓSFAL, M.; BARABÁSI, A.-L. *Network science*. [S.l.]: Cambridge University Press Cambridge, UK., 2016. v. 3.

RHEINWALT, A.; MARWAN, N.; KURTHS, J.; WERNER, P.; GERSTENGARBE, F.-W. Boundary effects in network measures of spatially embedded networks. *Europhysics Letters*, EDP Sciences, IOP Publishing and Società Italiana di Fisica, v. 100, n. 2, p. 28002, oct 2012. Disponível em: <<https://doi.org/10.1209/0295-5075/100/28002>>.

ROBUSTO, C. The cosine-haversine formula. *American Mathematical Monthly*, v. 64, p. 38, 1957.

TSONIS, A.; ROEBBER, P. The architecture of the climate network. *Physica A: Statistical Mechanics and its Applications*, v. 333, p. 497–504, 2004.

WATTS, D. J.; STROGATZ, S. H. Collective dynamics of ‘small-world’ networks. *Nature*, v. 393, p. 440–442, 1998. Disponível em: <<https://api.semanticscholar.org/CorpusID:3034643>>.

WINTER, J. C. D.; GOSLING, S. D.; POTTER, J. Comparing the pearson and spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychological methods*, American Psychological Association, v. 21, n. 3, p. 273, 2016.

WU, Z.; PAN, S.; CHEN, F.; LONG, G.; ZHANG, C.; YU, P. S. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, IEEE, v. 32, n. 1, p. 4–24, 2020.